How Good Are Clinical Severity and Outcome Measures for Psoriasis?: Quantitative Evaluation in a Systematic Review

Phyllis I. Spuls[1,4], Lidian L.A. Lecluse[1], Marie-Louise N.F. Poulsen[1], Jan D. Bos[1], Robert S. Stern[2] and Tamar Nijsten[3]

A large number of clinical measures of psoriasis are used in clinical trials and daily practice. These measures lack uniformity and validation. However, valid outcome and severity measures for psoriasis are a prerequisite for fully informative clinical research and evidence-based medicine. The purpose of this study was to identify all clinical measures of psoriasis severity and outcome in use and to evaluate the quality of these measures using clinimetric criteria; we identified 53 separate clinical measures, which were regrouped into 11 measures for quality analysis. No measure could be scored on all items used in the clinimetric analysis. The Lattice System Physician's Global Assessment and Physician's Global Assessment were most highly noted. We conclude that none of the psoriasis measures is adequately validated. The Psoriasis Area and Severity Index is the most commonly used clinical measure in research, but it has substantial limitations such as low response distribution, no consensus on interpretability, and low responsiveness in mild disease. Nevertheless, because of its widespread use the Psoriasis Area and Severity Index permits some degree of comparison of results among clinical trials. Overall, no best instrument was identified, and different situations may call for different measures.

Journal of Investigative Dermatology (2010) **130,** 933–943;
doi:10.1038/jid.2009.391; published online 31 December 2009

[1]*Department of Dermatology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands;* [2]*Department of Dermatology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA;* [3]*Department of Dermatology, Erasmus Medical Center, Rotterdam, The Netherlands and* [4]*Dutch Cochrane Centre, Amsterdam, The Netherlands*

Correspondence: Phyllis Spuls, Department of Dermatology, A0–227 (main center where the work was done), Academic Medical Center, PO Box 22700, 1100 DD Amsterdam, The Netherlands. E-mail: PH.I.SPULS@AMC.UVA.NL

Abbreviations: BSA, body surface area; HRQoL, health-related quality of life; LS-PGA, Lattice System Physician's Global Assessment; PASI, Psoriasis Area and Severity Index; PASS, Psoriasis Assessment Severity Score; PEASI, Psoriasis Exact Area and Severity Index; PGA, Physicians Global Assessment; PLASI, Psoriasis Log-Based Area and Severity Index; SAPASI, self-administered PASI; SPASI, simplified PASI

Received 16 July 2009; revised 25 October 2009; accepted 27 October 2009; published online 31 December 2009

INTRODUCTION

To measure disease severity and effectiveness of therapies, good clinical psoriasis measures are a necessity. For psoriasis, the range of severity and outcome measures is extensive and they lack uniformity and validation.

Psoriasis is a common disease with a high burden of disease and substantial impact on quality of life (Krueger *et al.*, 2001). As there are no biomarkers available to assess disease severity, clinical measures are used in clinical trials and daily practice to measure severity and treatment response. These measures are also used to categorize disease severity and to allocate resources for instance considering reimbursement criteria (NICE, 2006; www.labag.nl/info.htm).

A systematic review of clinical psoriasis measures in randomized-controlled trials up to 2000, identified no less than 44 different clinical measures used for psoriasis (Naldi *et al.*, 2003). Besides poor standardization of outcomes in clinical trials, there is a lack of validation of the used measures. Another review showed that no available measure fulfilled all requirements of a validated instrument for disease assessment (Ashcroft *et al.*, 1999). Subsequent reviews concluded that there is no ''best''' outcome or severity measure for use in clinical trials (Weisman *et al.*, 2003; Feldman and Krueger, 2005).

The focus of severity measures is the discriminant ability between severity levels of psoriasis, whereas outcome measures should be able to detect changes as a result of treatment. Both types of measures must have the properties of validity and reliability. The ideal measure is clearly defined with maximum objectivity, universally applicable, is easy to use, flexible and has clinical significance (Bigby and Gadenne, 1996; Jemec and Wulf, 1997). The availability of such a measure seems utopian and therefore it is necessary to know the quality and validity of the existing clinical measures. This information is important because consequential decisions are based on the scores of these measures. Signaling imperfections of the measures may lead to improved outcome and severity measures and as a result in better decision making in patient care.

The aim of this systematic research was to update the list of all clinical psoriasis measures and to evaluate their quality in a clinimetric way.

RESULTS

Data synthesis

The two searches, designed to identify severity measures gave 807 hits (search 1) and 366 hits (search 2). Overall they

yielded 53 different clinical psoriasis measures, which were often similar and could be regrouped into 11 measures for analysis (Table 1).

The search for articles regarding quality criteria of the clinical psoriasis measures identified 6,815 articles of which 42 articles were included for data extraction based on the inclusion and exclusion criteria.

Evaluation of the instruments

Table 2 gives an overview of the analyzed clinical psoriasis measures. Table 3 presents the results of the clinimetric evaluation. The available clinimetric data of each instrument are described below:

Body surface area (BSA)

Many assessments of psoriasis severity incorporate an estimation of involved BSA. Most commonly used to estimate the BSA is the ''rule of nines'' (Ramsay and Lawrence, 1991). It is defined as 9% coverage for the head and neck, each arm, anterior and posterior leg as well as the four trunk quadrants respectively, leaving 1% for the genitalia. The BSA can also be estimated by the number of patients' hand areas affected, assuming that one ''handprint'' reflects approximately 1% of BSA (Ramsay and Lawrence, 1991; Long *et al.*, 1992; Savolainen *et al.*, 1997, 1998; Kreft *et al.*, 2006; Thomas and Finlay, 2007). Validity was only tested with the PGA, which correlated strongly (Langley and Ellis, 2004). In contrast to the varying inter-rater reliability, the intra-rater reliability for area estimation was described as excellent (Marks *et al.*, 1989; Ramsay and Lawrence, 1991; Yune *et al.*, 2003).

The BSA was least improved after 2 weeks treatment compared with other psoriasis parameters, which is negative for responsiveness (Ormerod *et al.*, 1997). Several scales are in use, ordinal and continuous (Long *et al.*, 1992; Feldman *et al.*, 1996; Harari *et al.*, 2000; Gottlieb *et al.*, 2003; Jacobson and Kimball, 2004). The clinical assessed BSA did not differ statistically from objectively assessed BSA, for instance with computer-based image analysis in two studies (Kanthraj *et al.*, 1997; Ormerod *et al.*, 1997). However, several other studies showed that patients and clinicians significantly overestimate the affected area, especially in mild cases and when untrained (Ramsay and Lawrence, 1991; Tiling-Grosse and Rees, 1993; Savolainen *et al.*, 1997, 1998; Yune *et al.*, 2003; Kreft *et al.*, 2006). The correlation coefficient between BSA assessed by dermatologists and patients ranged from $r = 0.38$–0.82 (Feldman *et al.*, 1996, 2005; Fleischer, Jr *et al.*, 1999; Szepietowski *et al.*, 2001; Sampogna *et al.*, 2003).

Physical signs: erythema, scaling and induration

Plaque characteristics erythema, scaling and induration have been widely used to evaluate psoriasis severity and remission. Validity for the sum of signs was only tested with the PGA, correlation ran from $r = 0.3$ to $r = 0.6$ (Langley and Ellis, 2004). Reliability and other criteria were not tested. Many different scales were in use ranging from 2 to 8 points (Serup and Agner, 1990; Lahti *et al.*, 1993; Feldman *et al.*,

Table 1. All retrieved clinical outcome measures re-grouped in 11 main clinical severity and outcome measures for analysis

1. Psoriasis Area and Severity Index (PASI)
   - Psoriasis Area and Severity Index (PASI)
   - Extend Score of the Salford Psoriasis Index (SPI)

2. Body surface area (BSA)
   - Body surface area (BSA)
   - Total Body Surface Involvement
   - Area Index (AI)
   - Rule of Nines

3. Physician's/Psoriasis Global Assessment (PGA)

   Static assessment:
   - Physician's Global Assessment (PGA)
   - Psoriasis Global Assessment (PGA)
   - Investigator's Global Assessment of Plaque Severity
   - Investigator's Global Severity Assessment of Psoriasis
   - Investigator's Global Assessment of Disease Severity
   - Physicians' Overall Assessment of the Extent of Psoriatic Involvement
   - Investigator's Global Assessment of Overall Severity
   - Overall Lesion Severity Scale (OLS)
   - Physician Static Global Assessment

   Dynamic assessment:
   - Investigator's Overall Response Assessment
   - Investigator's Assessment of Improvement
   - Physicians' Assessment of Clinical Response
   - Investigator's Global Assessment of Improvement form baseline
   - Clinical Response to Treatment
   - Physician's Global Assessment of Response to Treatment
   - Overall Global Improvement of Psoriatic Lesions
   - Physician's Gross Assessment of Clinical Response
   - Global Improvement Score
   - Dynamic Global Assessment

4. Patient's Global Assessment (PaGA)

   Static assessment:
   - Patient's Global Assessment (PaGA)
   - Patient's Global Assessment of Plaque Severity
   - Subject's Global Severity Assessment of Psoriasis
   - Patient's Global Psoriasis Assessment (PGPA)

Table 1 continued on the following page

Table 1. Continued

Dynamic assessment:
• Patient's Overall Response Assessment
• Patient's Global Assessment of Improvement
• Patient's Assessment of Treatment Effect
• Patient's Assessment of Clinical Response
• Patient's Global Response to Treatment
5. Sum scores physical signs
• Psoriasis Severity Index/Scale
• Target lesion assessment/Score
• Local Psoriasis Severity Index of Target Lesions
• Target Area Score
• Target Plaque Severity Score
• Dermatological Sum Score
• Plaque Severity Score
• Total Sign Score (TSS)
• Plaque Modified Psoriasis Activity and Severity Index (PSI)
• Total Severity Score
• Plaque Severity Index
• Severity Index
• Psoriasis Grading Scale
6. Lattice-System Physician's Global Assessment (LS-PGA)
7. Psoriasis Assessment Severity Score (PASS)
8. Simplified PASI (SPASI)
9. Psoriasis Exact Area and Severity Index (PEASI)
10. Psoriasis Long-based Area and Severity Index (PLASI)
11. Self-Administered PASI (SAPASI)

Initially retrieved clinical outcome measures, n=53.

Table 2. Overview of the 11 main clinical psoriasis measures that were analyzed

Instrument	Description
BSA	Estimation of involved body surface area, several scores are used
Signs	Evaluation of the plaque characteristics erythema, scaling, and induration. Erythema and scaling are easily influenced by external factors
PASI	The affected area and lesion characteristics are entered in a formula that results in a score from 0 to 72. The PASI is most often used in clinical trials
PGA	The PGA is a 5, 6, or 7-point ordinal rating ranging from ''clear'' to ''very severe psoriasis''
PaGA	The PaGA is an ordinal rating ranging from ''clear'' to ''very severe psoriasis'' assessed by the patient
SAPASI	The SAPASI is a structured PASI-like instrument designed for patient self-assessments of severity
PASS	The affected area and plaque characteristics are entered in a formula that results in a score from 0 to 140. Infiltration is given more weight than erythema and scaling.
LS-PGA	The LS-PGA integrates ranges of involved BSA and the overall plaque morphology in which infiltration is given more weight
SPASI	The SPASI equals the sum of the average redness, thickness, and scaling of all the psoriasis lesions multiplied by the percentage of body surface area involved
PEASI	The PLASI is derived from the PASI but uses actual BSA percentages instead of an area score
PLASI	The PLASI is derived from the PASI but uses six BSA groupings with finer partitioning for smaller extents of BSA

1996; Ormerod *et al.*, 1997; Harari *et al.*, 2000; Gottlieb *et al.*, 2003). Erythema assessment is likely to be affected by several factors including viewing geometry, ambient illumination, tanning of the surrounding skin, edema and the experience and visual acuity of the observer (Lahti *et al.*, 1993). Varying correlations were found between clinical assessment of erythema versus objective erythema apparatus (r=0.3–0.79) (Serup and Agner, 1990; Lahti *et al.*, 1993; Ormerod *et al.*, 1997). The correlation coefficient between erythema assessed by dermatologists and patients ranged was moderate (r=0.37–0.4) (Feldman *et al.*, 1996; Fleischer, Jr *et al.*, 1999; Sampogna *et al.*, 2003).

Patients rate degree of scaling as a strong indicator of disease severity. However, application of moisturizers rapidly affects scaling, making it an unstable parameter. Clinical measurement of scaling correlated moderately to ultrasound entry echo (r=0.53) (Ormerod *et al.*, 1997). Patient versus dermatologist scaling assessment did not correlate (Fleischer, Jr *et al.*, 1999). Induration or thickness of psoriatic lesions is probably the most specific parameter of psoriasis activity (Gottlieb *et al.*, 2003; Langley and Ellis, 2004). Elevation correlated moderately to ultrasound thickness (r=0.58) (Ormerod *et al.*, 1997). Patient versus dermatologist induration assessment ranged from a weak correlation (r=0.24) to no correlation at all (Feldman *et al.*, 1996; Fleischer, Jr *et al.*, 1999).

The Psoriasis Area and Severity Index (PASI)

In 1978, the PASI was developed to assess the effects of retinoids in psoriasis (Fredriksson and Pettersson, 1979). The affected area and lesion characteristics are entered in a formula that results in a score from 0 to 72. PASI has been criticized for being resource intensive, complex, lacking sensitivity, low in accuracy and having a non-linear scale (Harari *et al.*, 2000; Berth-Jones *et al.*, 2006). Nevertheless, the PASI is often used as the standard measurement in the validation of new measures and correlated well in most cases with other physician-based assessments, but not with HRQoL and symptoms measurements. Reliability was only once correctly calculated (Berth-Jones *et al.*, 2006). Response distribution is low, because practically only half of the scale is used (Fredriksson and Pettersson, 1978; Harari *et al.*, 2000; Jacobson and Kimball, 2004; Langley and Ellis, 2004;

Table 3. The grading of the clinimetric properties of the clinical psoriasis severity measures

	Construct validity					Test-retest reliability	Inter-rater reliability	Respon-siveness	Response distribution	Interpretability		Easy to administer	Unifor-mity	Number of studies included	References used	Considerations for choosing measures
	Content validity	Criterion validity	Clinical	Symptom	HRQoL					Categori-zation	MIC					
Components of measures																
BSA	E	—	A	—	—	—	D	E	—	E	E	—	E	18	Feldman *et al.*, 1996; Fleischer, Jr *et al.*, 1999; Gottlieb *et al.*, 2003; Harari *et al.*, 2000; Jacobson and Kimball, 2004; Kanthraj *et al.*, 1997; Kreft *et al.*, 2006; Langley and Ellis, 2004; Long *et al.*, 1992; Marks *et al.*, 1989; Ormerod *et al.*, 1997; Ramsay and Lawrence, 1991; Sampogna *et al.*, 2003; Savolainen *et al.*, 1998; Savolainen *et al.*, 1997; Szepietowski *et al.*, 2001; Tiling-Grosse and Rees, 1993; Yune *et al.*, 2003	- Fit for static measurement (clinical construct) - Unfit for measuring response (responsiveness)
Erythema																
Scaling	E	—	B–C	—	—	—	—	—	—	E	E	—	E	10	Feldman *et al.*, 1996; Fleischer, Jr *et al.*, 1999; Gottlieb *et al.*, 2003; Harari *et al.*, 2000; Lahti *et al.*, 1993; Langley and Ellis, 2004; Ormerod *et al.*, 1997; Sampogna *et al.*, 2003; Serup and Agner, 1990; Szepietowski *et al.*, 2001	- Fit for evaluating single lesions - Unfit for overall score (construct validity)
Induration																
Clinical severity measures																
PASI	C	X	A–D	D	C–E	A	A	E	E	E	E	—	A	28	Ashcroft *et al.*, 1999; Berth-Jones *et al.*, 2006; Carlin *et al.*, 2004; Ellis, 2007; Feldman, 2004; Feldman *et al.*, 2005; Feldman *et al.*, 1996; Feldman and Krueger, 2005; Fleischer, Jr *et al.*, 1994; Fredriksson *et al.*, 1983; Fredriksson and Pettersson, 1978; Gottlieb *et al.*, 2003; Harari *et al.*, 2000; Jacobson and Kimball, 2004; Jemec and Wulf, 1997; Katz, 2005;	- Fit for comparing data (widespread use) - Unfit for assessing mild disease (responsiveness)

Table 3. Continued

	Construct validity					Test-retest reliability	Inter-rater reliability	Respon-siveness	Response distribution	Interpretability		Easy to administer	Unifor-mity	Number of studies included	References used	Considerations for choosing measures
	Content validity	Criterion validity	Clinical	Symptom	HRQoL					Categori-zation	MIC					
															Kirby *et al.*, 2000; Kirby *et al.*, 2001; Langley and Ellis, 2004; Louden *et al.*, 2004; Papp and Henninger, 2005; Rapaport, 2005; Sampogna *et al.*, 2004; Sampogna *et al.*, 2003; Schmitt and Wozel, 2005; Szepietowski *et al.*, 2001; van de Kerkhof, 1992; Vardy *et al.*, 1993	
PGA	E	A	A–C	—	—	A	B	—	A	A	E	—	E	5	Berth-Jones *et al.*, 2006; Ellis, 2007; Gottlieb *et al.*, 2003; Langley and Ellis, 2004; Weisman *et al.*, 2003	- Fit for assessment of severity of lesions - Unfit for overall assessment (content validity)
PaGA	—	—	—	—	—	—	—	—	—	—	—	—	—	—		- No data available
SAPASI	C	B–D	B–D	D	D–E	D	D	A	—	A	E	—	A	10	Feldman *et al.*, 2005; Feldman *et al.*, 1996; Fleischer, Jr *et al.*, 1996; Fleischer, Jr *et al.*, 1994; Kirby *et al.*, 2000; Kirby *et al.*, 2001; Sampogna *et al.*, 2004; Sampogna *et al.*, 2003; Szepietowski *et al.*, 2001	- Fit for surveys
PASS	A	—	—	—	—	—	—	A	E	E	E	—	A	1	Harari *et al.*, 2000	- Fit for static assessment and response - No data available for comparing scores
LS-PGA	A	A	A	—	—	A	A	—	A	A	E	—	A	3	Berth-Jones *et al.*, 2006; Ellis, 2007; Langley and Ellis, 2004	- Fit for static assessment - Unfit for assessment of response, responsiveness unknown
SPASI	C	A	—	—	—	—	—	—	—	E	E	—	A	1	Louden *et al.*, 2004	- Too little data available
PEASI	C	A	—	—	—	—	—	A	E	E	E	—	A	1	Jacobson and Kimball, 2004	- Too little data available
PLASI	C	A	—	—	—	—	—	A	E	E	E	—	A	1	Jacobson and Kimball, 2004	- Too little data available

—, No information found; X, PASI is used as standard and is per definition 100% correlated to itself. MIC, minimal important change. (Fredriksson *et al.*, 1983; Ramsay and Lawrence, 1991; van de Kerkhof, 1992; Tiling-Grosse and Rees, 1993; Vardy *et al.*, 1993; Fleischer, Jr *et al.*, 1999; Feldman, 2004; Papp and Henninger, 2005; Ellis, 2007).

Berth-Jones *et al.*, 2006). PASI responsiveness is weak when patients reach <10% BSA in any body area because changes in the PASI entirely depend on plaque severity score improvement and may underestimate the general degree of improvement. PASI assessment gets more reliably by experience (Langley and Ellis, 2004). There is no consensus on the interpretability (van de Kerkhof, 1992; Gottlieb *et al.*, 2003; Carlin *et al.*, 2004; Jacobson and Kimball, 2004; Langley and Ellis, 2004; Katz, 2005; Rapaport, 2005; Berth-Jones *et al.*, 2006). However, several proposals have been made, for instance by Schmitt *et al.* who translated the PASI ranges into the terms ''mild'', ''moderate'', and ''severe'' (Schmitt and Wozel, 2005).

Physician's Global Assessment (PGA)
Typically, the PGA is a 5, 6 or 7-point ordinal rating ranging from ''clear'' to ''very severe psoriasis''. The PGA can be used to show improvement by a comparison with baseline disease severity (dynamic PGA) or it can be an assessment made at one moment in time (static PGA). The PGA correlated well with other clinically assessed, symptom and HRQoL psoriasis measurements including the PASI (Gottlieb *et al.*, 2003; Langley and Ellis, 2004; Berth-Jones *et al.*, 2006). The PGA correlated more with BSA than with signs, although the extent of involvement should not be incorporated in the PGA (Langley and Ellis, 2004). Reliability was calculated to be good and experience appeared to have a neglectable effect on PGA assessment (Langley and Ellis, 2004). The scales are clear and most of the scale is used (Gottlieb *et al.*, 2003; Berth-Jones *et al.*, 2006).

Patient's Global Assessment
No studies were identified assessing clinimetric properties of the patient's global assessment.

Self-Administered PASI (SAPASI)
The SAPASI is a structured PASI-like instrument designed for patient self-assessments of severity (Fleischer, Jr *et al.*, 1994). Patients shade in affected areas on a silhouette of a body to estimate body surface area and complete visual analog scales for the extent of erythema, induration and scaling of their ''average'' lesion. The investigator uses these data and combine them into a complex score, ranging from 0–72. The SAPASI correlation with the PASI has been measured many times and appeared to be strong in most cases (Feldman *et al.*, 1996, 2005; Kirby *et al.*, 2000, 2001; Szepietowski *et al.*, 2001; Sampogna *et al.*, 2003, 2004). The SAPASI did not correlate well with HRQoL measurements (Kirby *et al.*, 2000; Sampogna *et al.*, 2004). Reliability of the SAPASI was described to be very good (Feldman *et al.*, 1996; Fleischer, Jr *et al.*, 1996). The SAPASI was reported to be responsive to changes in severity over time because it correlated well with changes in PASI ($r=0.62$), but the responsiveness of the latter is doubtful (Feldman *et al.*, 2005). Psoriasis is defined as ''in remission'' when SAPASI = 0, ''mild'' when 0 < SAPASI < 3, ''moderate'' when 3 < SAPASI < 15 and ''severe'' when SAPASI >15 (Fleischer Jr, *et al.*, 1996).

Psoriasis Assessment Severity Score (PASS)
The PASS was developed to be simpler and faster than the PASI (Harari *et al.*, 2000). Overall evaluation is divided in two stages: in the first the BSA is determined in percentage, than the general erythema, desquamation and induration are assessed on a three-point scale. Finally, the sign scores together with the total percentage BSA are combined in a complex formula, which gives an overall score between 0 and 140. In this calculation, infiltration is weighted more heavily than erythema. Almost all patients are in the lower half of the score. The PASS has not been validated. Inter-rater reliability was described to be better than the PASI (Harari *et al.*, 2000).

Lattice System Physician's Global Assessment (LS-PGA)
The LS-PGA is similar to the PGA, but takes a quantitative approach to the global assessment of disease severity by integrating ranges of involved BSA and the overall plaque morphology (Langley and Ellis, 2004). The BSA percentage involved is measured in categories of 0, 1–3, 4–9, 10–20, 21–29, 30–50 and 51–100%. The LS-PGA gives more weight to induration compared with scaling and erythema. Validity and reliability were shown to be very good. Psoriasis severity is stratified in eight categories (clear to very severe) and most of the scale was used (Berth-Jones *et al.*, 2006). Although it has been suggested that this measure is sensitive to clinical change, this is not well documented (Berth-Jones *et al.*, 2006).

Simplified PASI (SPASI)
The SPASI is mathematically derived from the PASI and is meant to be easy to use (Louden *et al.*, 2004). The SPASI equals the sum of the average redness, thickness, and scaling of all the psoriasis lesions, multiplied by an estimate of total percentage body surface areas involved. Using a simulated patient database, correlation coefficients between the SPASI and PASI exceeded 0.90. Reliability has not been tested. Although not formally studied, the SPASI seems relatively insensitive to change, especially with less extensive involvement (BSA <10%) or localized disease (Louden *et al.*, 2004).

Psoriasis Log-Based Area and Severity Index (PLASI) and Psoriasis Exact Area and Severity Index (PEASI)
Derived from the PASI, the PLASI and the PEASI are intended to provide more accurate assessment of improvement. The PLASI uses six BSA groupings (100–46, 46–21, 21–10, 10–5, 5–2 and 2–0%) with finer partitioning for smaller extents of BSA affected. This is supposed to reduce the error resulting from inaccurate estimation of BSA in patients with less extensive disease, also to increase sensitivity among patients with mild-to-moderate disease in detecting changes in psoriasis severity.

The PEASI uses actual BSA percentages instead of an area score for each body area. The PEASI and PLEASI have not been validated and are not tested for reliability. Considering responsiveness the observed percentage change was greater for both the PLEASI and the PEASI than with the PASI

(Jacobson and Kimball, 2004). Most patients score below 360 on a 1200-point scale. Both ΔPEASI and ΔPLEASI corresponded with patients' self-assessments (Jacobson and Kimball, 2004).

The frequency of use

In randomized-controlled trials published between 2000 and May 2007, the PASI was used most often (126 times). Frequently used were sum-scores of erythema, scaling and induration, and the PGA (67 and 52 times, respectively). BSA involvement and SAPASI were used occasionally (10 and 2 times, respectively). Several clinical severity measures were developed recently and had not yet been cited in psoriasis randomized-controlled trials published before May 2007.

Summary of results

Only PASS and LS-PGA scored high for ''content validity'', because they gave more weight to induration. PGA, LS-PGA, PEASI, PLASI, and SAPASI correlated well with the PASI, which was used for ''criterion validity'' in the absence of consensus on a gold standard. BSA, PASI, PGA, and LS-PGA correlated well with other clinical psoriasis measures and thus scored well on ''construct validity''. For ''test-retest reliability'' and ''inter-rater reliability'' the LS-PGA, the PGA, and the PASI scored best. Only for the SAPASI, PASS, PEASI, and PLASI, positive information was found on responsiveness. Clinical relevant categorization was found for PGA, SAPASI, and LS-PGA. The minimal important change, however, was not defined for any of the measures. BSA and signs were the only measures for which there are several scales in use.

In Table 3 the number of studies evaluating the severity measures is given. No trend could be found between the number of articles evaluating the measures and the values given to the quality criteria.

DISCUSSION

In this study, we showed that there are many different clinical measures in use for psoriasis and that their number still rises. In 2000, Naldi *et al.* had identified 44 psoriasis measures in clinical trials, which had increased to 53 clinical measures in our search. For quantitative analysis of the main psoriasis measures, the number of eligible articles per measure varied from 0 to 28. Table 3 enables researchers and clinicians involved in clinical psoriasis assessment to make an evidence-based choice for selecting an appropriate measure based on the evaluation of the appropriate clinimetric dimensions. None of the measures had been tested for all of the clinimetric items. More importantly, most of them had not been tested properly for most of the items.

Surprisingly, many of the clinical psoriasis measures that have been developed to overcome limitations of the PASI (SAPASI, PASS, SPASI, PEASI, and PLASI), could not exceed the PASI on most of the clinimetric properties. Often quality data were only available in the single article that introduced these measures.

Criteria for objectively assessing administerability were derived from Schmitt *et al.* (2007). However, none of the included articles gave an indication of the time needed for the assessment (although this may vary from person to person). All remarks on administerability in the included articles were highly subjective.

A remarkable finding in the review was the weak correlation of HRQoL measures with the PASI, which is used as an almost universal outcome measure in psoriasis trials. The majority of the correlation values ranged from 0.1 to 0.3 (Kirby *et al.*, 2000; Sampogna *et al.*, 2004). This weak correlation between clinical severity and HRQoL was also seen when patients themselves assessed the clinical severity with the SAPASI (Sampogna *et al.*, 2004). We expected clinical severity to correlate more with quality of life. Especially, as we know that reductions in physical and mental functioning because of psoriasis have found to be similar to those reported in patients with cancer, arthritis, hypertension, heart disease, diabetes, and depression (Rapp *et al.*, 1999). The discrepancy could be explained by the fact that an objectively moderate psoriasis plaque may have a great impact on HRQoL, if on a visible area of the body or around the genitals. Even so, a large mild plaque may give a high objective clinical score, but may have a low impact on HRQoL. There is also evidence that a mismatch of low PASI and significantly impaired HRQoL suggests comorbid depression. Especially patients with high HRQoL impairment despite objectively mild psoriasis should be screened for depression (Schmitt and Ford, 2007).

As severity measures and psychometric measures assess different constructs, they should be presented as separate scores and not summed in a single score. However, in assessing disease severity, HRQoL scores are complementary to the clinical severity scores. For instance, the PASI may show significant improvement in clinical disease severity and HRQoL measures can be used to confirm that these changes are clinically meaningful (Feldman and Krueger, 2005).

Issues like these can only be identified using well validated, one-dimensional measures in research and practice. For this reason, we excluded multidimensional measures such as NPF-PS, Beer Sheva Severity Score and Dermatology Index of Disease Severity from our systematic review.

However, most of the one-dimensional severity measures for psoriasis, like the PASI, are in a way composite measures. The end points of erythema, induration, scaling, and area involvement are integrated to form a single value. When using these measures, it should be kept in mind that a significant change in the score does not imply that all components necessarily trended in the same direction (Buzney and Kimball, 2008).

A limitation of the study is, that the adapted quality criteria for evaluating clinical psoriasis measures are not validated. However, no validated list for evaluating validation studies of clinical severity measures is available. The HRQoL criteria we based our criteria on are widely accepted and used before (Both *et al.*, 2007). Furthermore, the used items and applied criteria are clearly described.

Another limitation is lack of quality criteria for the included articles. The study results are influenced by the included population, the assessor and the circumstances among other

Table 4. The definitions of the clinimetric properties, grading categories, and weighted score for clinical severity outcome measures

Property	Definition	Score	Quality criteria
Validity			
1a. Content validity	The extent to which the type and number of scale items of the outcome measure adequately represent the underlying construct	A	Fulfillment of all demands: A clear description is provided of the measurement aim The target population is described All the concepts of disease severity are included, such as body surface area involved, erythema, scaling and induration, and the rating used in the clinical severity measure is elucidated Plaque elevation is given more weight, as it is assumed to be the most significant clinical sign of the disease
		C	All of the above-mentioned demands are rated positive, but plaque elevation is not given more weight
		E	One or more above-mentioned demands are not fulfilled
1b. Criterion validity	The extent to which clinical measures relate to the PASI (only original data used)		Spearman's rank correlation between clinical severity measure and PASI
		A	$r \geqslant 0.7$(very strong correlation)
		B	r=0.5–0.7 (strong correlation)
		C	r=0.3–0.49 (moderate correlation)
		D	r=0.29–0.1 (weak correlation)
		E	r=0 (no correlation)
1c. Construct validity	The extent to which scores relate to other measures. The results are shown separately for relation to other severity measures, to symptom measures, and to HRQoL measures (only original data used)		Overall Spearman's rank correlations with other outcome measures divided into (1) other clinical severity measures, (2) symptom measures and (3) quality of life measures. Overall Spearman's rank correlation for each group:
		A	r=> 0.7 (very strong correlation)
		B	r=0.5–0.7 (strong correlation)
		C	r=0.3–0.49 (moderate correlation)
		D	r=0.29–0.1 (weak correlation)
		E	r=0 (no correlation)
Reliability			
2a. Test-retest/ interpreter reliability	Examines the influence of random error by determining how consistent scores remain across multiple administrations of the instrument, and can be determined by correlating rating scores from multiple testing sessions (only original data used)		ICC or weighted kappa calculated:
		A	81–100% (substantial)
		B	61–80% (moderate)
		C	41–60% (fair)
		D	11–40% (slight)
		E	0–10%: (virtually none)
			Only correlation coefficients are calculated:
		D	81–100%
		E	0–80%
2b Inter-rater reliability	Examines the degree to which multiple observers agree on the assignment of scales (only original data used)		ICC or weighted kappa calculated:
		A	81–100% (substantial)
		B	61–80% (moderate)
		C	41–60% (fair)
		D	11–40% (slight)
		E	0–10%: (virtually none)

Table 4 continued on following page

Table 4. Continued

Property	Definition	Score	Quality criteria
			Only correlation coefficients are calculated:
		D	81–100%
		E	0–80%
3. Responsiveness	The ability of an instrument to detect changes over time. The instrument should be able to distinguish clinically important change from measurement error		As this is not tested for clinical severity measures in psoriasis, we only distinguish between:
		A	Positive information found on responsiveness
		E	Negative information found on responsiveness
4. Response distribution	Examines whether the entire range of a scale is used (only original data used)	A	Positive information found on the usage of the entire range of a scale
		E	Negative information found on the usage of the entire range of a scale
5. Interpretability	The degree to which one can assign qualitative meaning to quantitative scores. Scores should provide information about what (change in) score would be clinically meaningful		A:
		A	Clinically relevant categorization is defined
		E	Clinically relevant categorization is not defined
			B:
		A	Minimal important change is defined
		E	Minimal important change is not defined
6. Easy to administer	The degree to which an outcome measure can easily be used in clinical practice (only original data used)	A	Fulfillment of all demands:
			Fulfilling the rating is not time consuming (not exceeding 3 min)
			No extra tools (except score form) are needed
			The rating is easily understandable
		C	Fulfilling criteria above, but fulfilling the rating takes between 3 and 7 min
		E	Not fulfilling $\geqslant$1 of the above criteria
7. Uniformity	The degree to which there are variations in used scales with the same clinical severity measure	A	Only one rating is used per clinical severity measure
		E	More than one rating is used per clinical severity measure

Criteria were adjusted from Terwee *et al.* (2007) and Both *et al.* (2007).

things. We had no minimum criteria for inclusion, like the use of experienced assessors, a minimum number of patients, an applicable population or stable circumstances. Also the quantity of included studies per measurement is likely to influence the grading.

CONCLUSION

The main conclusion is that there are no adequately validated clinical measures for psoriasis. As there is no supreme measure, different measures might be ideal for different situations and we might need all of them. When choosing a measure, it is important to determine the most needed features, for example, good responsiveness or sensitivity in mild disease. It may be necessary to combine two or more scores to satisfy all needs. For example, PASI may not be particularly sensitive for mild disease, but it may be outstanding for a study in which patients have severe disease. It also provides the advantage of a large base of studies in which it has been used. Another instrument may have some characteristics that are better, but this may not outweigh the benefit of being able to compare with the existing database of studies that used PASI. For interventional studies responsiveness is important, which points to some newer measures like the PLASI and PEASI. In cross-sectional studies interpretability is important which favors the PGA, SAPASI, and LS-PGA. If someone would do a mail survey of psoriasis patients, the SAPASI is preferred because this measure is developed for patient assessment. If future authors want a reliable instrument, then the LS-PGA and PASI would be best, with the PGA a close follow-up.

Further validation studies are needed to complete the overview of the clinimetric properties of the evaluated measures. An item that is specifically important is the

definition of minimal important changes and categorizations, for better interpretation of the results of the measures. Furthermore, future research is needed that focuses on responsiveness, which is an important item especially when evaluating new therapies.

MATERIALS AND METHODS

Search strategy

To update the list of all clinical psoriasis measures a Medline search was conducted in May 2007, using the term ''psoriasis'' combined with common clinical severity measures to retrieve an initial list of existing clinical outcome measures. An additional Medline search with ''psoriasis'' and ''Randomized-Controlled Trials'' was conducted limited to the years 2000–2007.

To evaluate the found measures on the quality criteria, a search in Medline, Embase, Central and DARE was conducted, with ''psoriasis'' and an extensive validation filter created by a clinical librarian. The specific search details are available online (see Supplementary Table S1).

Selection

For the list of the clinical psoriasis measures all studies including instruments that used clinical physical examination to measure psoriasis severity were eligible. Composite, symptom and/or HRQoL measures were excluded. We excluded direct physical measures, such as ultrasound in this review. These measures are not generally practical for daily use and their clinical utility is not clear. For the quality evaluation all studies evaluating the eligible measures for validity, reliability, responsiveness, response distribution, interpretability, ease to administer or uniformity were included (Supplementary Figure S1). Two reviewers independently screened all articles on title/abstract. Disagreements were solved by discussion. The reference lists of these articles were screened for additional studies. Final selection for inclusion was based on the assessment of the full-text article.

Critical appraisal

There is no format for critical appraisal of these publication types available.

Data extraction and analyses

Two reviewers independently extracted data from the selected articles.

We extracted all different clinical psoriasis measures used from 2000 up to May 2007, which had been done by Naldi *et al.* up to 2000 (Naldi *et al.*, 2003). For evaluating the quality of the measures, we adjusted guidelines for HRQoL instruments (Both *et al.*, 2007; Terwee *et al.*, 2007). Table 4 presents the definitions of the quality criteria and the grading categories. Each quality criterion was scored for all the measures.

There is no consensus on a gold standard for psoriasis, although to PASI is sometimes considered as such (Feldman *et al.*, 1996; Ashcroft *et al.*, 1999; Fleischer, Jr *et al.*, 1999; Jacobson and Kimball, 2004; Berth-Jones *et al.*, 2006). Nevertheless, we used the PASI for criterion validity, because it is an almost universal outcome measure in clinical trials of antipsoriatic agents. No overall sum-scores are given since this would assume that the contribution of the different measurement properties to the overall quality is known and that these properties are equally important.

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGMENTS

We acknowledge Arnold Leenders, clinical librarian from Central Medical Library of the Academic Medical Center, for his help with the searches. This project was the undergraduate thesis of Louise Poulsen.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/jid

REFERENCES

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

Berth-Jones J, Grotzinger K, Rainville C *et al.* (2006) A study examining inter- and intrarater reliability of three scales for measuring severity of psoriasis: Psoriasis Area and Severity Index, Physician's Global Assessment and Lattice System Physician's Global Assessment. *Br J Dermatol* 155:707–13

Bigby M, Gadenne AS (1996) Understanding and evaluating clinical trials. *J Am Acad Dermatol* 34:555–90

Both H, Essink-Bot ML, Busschbach J *et al.* (2007) Critical review of generic and dermatology-specific health-related quality of life instruments. *J Invest Dermatol* 127:2726–39

Buzney EA, Kimball AB (2008) A critical assessment of composite and coprimary endpoints: a complex problem. *J Am Acad Dermatol* 59:890–6

Carlin CS, Feldman SR, Krueger JG *et al.* (2004) A 50% reduction in the Psoriasis Area and Severity Index (PASI 50) is a clinically significant endpoint in the assessment of psoriasis. *J Am Acad Dermatol* 50: 859–66

Ellis CN (2007) Scoring severity of psoriasis. *JEADV* 21(Suppl 1):30–69

Feldman SR (2004) A quantitative definition of severe psoriasis for use in clinical trials. *J Dermatolog Treat* 15:27–9

Feldman SR, Clark AR, Venkat AP *et al.* (2005) The Self-Administered Psoriasis Area and Severity Index provides an objective measure of psoriasis severity. *Br J Dermatol* 152:382–3

Feldman SR, Fleischer AB Jr, Reboussin DM *et al.* (1996) The self-administered psoriasis area and severity index is valid and reliable. *J Invest Dermatol* 106:183–6

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

Fleischer AB Jr, Feldman SR, Dekle CL (1999) The SAPASI is valid and responsive to psoriasis disease severity changes in a multi-center clinical trial. *J Dermatol* 26:210–5

Fleischer AB Jr, Feldman SR, Rapp SR *et al.* (1996) Disease severity measures in a population of psoriasis patients: the symptoms of psoriasis correlate with self-administered psoriasis area severity index scores. *J Invest Dermatol* 107:26–9

Fleischer AB Jr, Rapp SR, Reboussin DM *et al.* (1994) Patient measurement of psoriasis disease severity with a structured instrument. *J Invest Dermatol* 102:967–9

Fredriksson T, Lassus A, Salde L (1983) Reproducibility of clinical trials of topical glucocorticosteroids. *Int J Dermatol* 22:536–9

Fredriksson T, Pettersson U (1978) Severe psoriasis – oral therapy with a new retinoid. *Dermatologica* 157:238–44

Fredriksson T, Pettersson U (1979) Oral treatment of pustulosis palmoplantaris with a new retinoid, Ro 10–9359. *Dermatologica* 158:60–4

Gottlieb AB, Chaudhari U, Baker DG *et al.* (2003) The National Psoriasis Foundation Psoriasis Score (NPF-PS) system versus the Psoriasis Area Severity Index (PASI) and Physician's Global Assessment (PGA): a comparison. *J Drugs Dermatol* 2:260–6

Harari M, Shani J, Hristakieva E *et al.* (2000) Clinical evaluation of a more rapid and sensitive Psoriasis Assessment Severity Score (PASS), and its

comparison with the classic method of Psoriasis Area and Severity Index (PASI), before and after climatotherapy at the Dead-Sea. *Int J Dermatol* 39:913–8

Jacobson CC, Kimball AB (2004) Rethinking the psoriasis area and severity index: the impact of area should be increased. *Br J Dermatol* 151:381–7

Jemec GB, Wulf HC (1997) The applicability of clinical scoring systems: SCORAD and PASI in psoriasis and atopic dermatitis. *Acta Derm Venereol* 77:392–3

Kanthraj GR, Srinivas CR, Shenoi SD *et al.* (1997) Comparison of computer-aided design and rule of nines methods in the evaluation of the extent of body involvement in cutaneous lesions. *Arch Dermatol* 133:922–3

Katz KA (2005) Psoriasis Area and Severity Index 50 as an endpoint in psoriasis trials: an unconvincing proposal. *J Am Acad Dermatol* 53:547–51

Kirby B, Fortune DG, Bhushan M *et al.* (2000) The Salford Psoriasis Index: an holistic measure of psoriasis severity. *Br J Dermatol* 142:728–32

Kirby B, Richards HL, Woo P *et al.* (2001) Physical and psychologic measures are necessary to assess overall psoriasis severity. *J Am Acad Dermatol* 45:72–6

Kreft S, Kreft M, Resman A *et al.* (2006) Computer-aided measurement of psoriatic lesion area in a multicenter clinical trial-Comparison to physician's estimations. *J Dermatol Sci* 44:21–7

Krueger G, Koo J, Lebwohl M *et al.* (2001) The impact of psoriasis on quality of life: results of a 1998 National Psoriasis Foundation patient-membership survey. *Arch Dermatol* 137:280–4

Lahti A, Kopola H, Harila A *et al.* (1993) Assessment of skin erythema by eye, laser Doppler flowmeter, spectroradiometer, two-channel erythema meter and Minolta chroma meter. *Arch Dermatol Res* 285:278–82

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

Long CC, Finlay AY, Averill RW (1992) The rule of hand: 4 hand areas=2 FTU=1 g. *Arch Dermatol* 128:1129–30

Louden BA, Pearce DJ, Lang W *et al.* (2004) A simplified psoriasis area severity index (SPASI) for rating psoriasis severity in clinic patients. *Dermatol Online J* 10:7

Marks R, Barton SP, Shuttleworth D *et al.* (1989) Assessment of disease progress in psoriasis. *Arch Dermatol* 125:235–40

Naldi L, Svensson A, Diepgen T *et al.* (2003) Randomized clinical trials for psoriasis 1977–2000: the EDEN survey. *J Invest Dermatol* 120: 738–41

NICE (2006) New NICE guidance to help sufferers of poriasis and psoriatic arthritis. http://www.nice.org.uk/nicemedia/pdf/2006-034PsoriasisJointLaunch.pdf

Ormerod AD, Dwyer CM, Weller R *et al.* (1997) A comparison of subjective and objective measures of reduction of psoriasis with the use of ultrasound, reflectance colorimetry, computerized video image analysis, and nitric oxide production. *J Am Acad Dermatol* 37:51–7

Papp KA, Henninger E (2005) Safe psoriasis control: a new outcome measure for the composite assessment of the efficacy and safety of psoriasis treatment. *J Cutan Med Surg* 9:276–83

Ramsay B, Lawrence CM (1991) Measurement of involved surface area in patients with psoriasis. *Br J Dermatol* 124:565–70

Rapaport MJ (2005) PASI 75. *J Am Acad Dermatol* 52:183–4

Rapp SR, Feldman SR, Exum ML *et al.* (1999) Psoriasis causes as much disability as other major medical diseases. *J Am Acad Dermatol* 41:401–7

Sampogna F, Sera F, Abeni D (2004) Measures of clinical severity, quality of life, and psychological distress in patients with psoriasis: a cluster analysis. *J Invest Dermatol* 122:602–7

Sampogna F, Sera F, Mazzotti E *et al.* (2003) Performance of the self-administered Psoriasis Area and Severity Index in evaluating clinical and sociodemographic subgroups of patients with psoriasis. *Arch Dermatol* 139:353–8

Savolainen L, Kontinen J, Alatalo E *et al.* (1998) Comparison of actual psoriasis surface area and the psoriasis area and severity index by the human eye and machine vision methods in following the treatment of psoriasis. *Acta Derm Venereol* 78:466–7

Savolainen L, Kontinen J, Roning J *et al.* (1997) Application of machine vision to assess involved surface in patients with psoriasis. *Br J Dermatol* 137:395–400

Schmitt J, Ford DE (2007) Understanding the relationship between objective disease severity, psoriatic symptoms, illness-related stress, health-related quality of life and depressive symptoms in patients with psoriasis – a structural equations modeling approach. *Gen Hosp Psychiatry* 29:134–40

Schmitt J, Langan S, Williams HC (2007) What are the best outcome measurements for atopic eczema? A systematic review. *J Allergy Clin Immunol* 120:1389–98

Schmitt J, Wozel G (2005) The psoriasis area and severity index is the adequate criterion to define severity in chronic plaque-type psoriasis. *Dermatology* 210:194–9

Serup J, Agner T (1990) Colorimetric quantification of erythema—a comparison of two colorimeters (Lange Micro Color and Minolta Chroma Meter CR-200) with a clinical scoring scheme and laser-Doppler flowmetry. *Clin Exp Dermatol* 15:267–72

Szepietowski JC, Sikora M, Pacholek T *et al.* (2001) Clinical evaluation of the self-administered psoriasis area and severity index (SAPASI). *Acta Dermatovenerologica Alpina, Panonica et Adriatica* 10:79–83

Terwee CB, Bot SD, de Boer MR *et al.* (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60:34–42

Thomas CL, Finlay AY (2007) The ''handprint'' approximates to 1% of the total body surface area whereas the ''palm minus the fingers'' does not. *Br J Dermatol* 157:1080–1

Tiling-Grosse S, Rees J (1993) Assessment of area of involvement in skin disease: a study using schematic figure outlines. *Br J Dermatol* 128: 69–74

van de Kerkhof PC (1992) On the limitations of the psoriasis area and severity index (PASI). *Br J Dermatol* 126:205

Vardy DA, Guberman D, Lichtenstein DA *et al.* (1993) Assessment of severity score in patients with psoriasis. *Br J Dermatol* 129:349–50

Weisman S, Pollack CR, Gottschalk RW (2003) Psoriasis disease severity measures: comparing efficacy of treatments for severe psoriasis. *J Dermatolog Treat* 14:158–65

(2007) The dutch national evaluation of applications of drugs (LABAG). www.labag.nl/info.htm

Yune Y-M, Park SY, Oh HS *et al.* (2003) Objective assessment of involved surface area in patients with psoriasis. *Skin Res Technol* 9:339–42