Psoriasis: implications of biologics
Lecluse, L.L.A.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
How good are clinical severity outcome measures for psoriasis:
Quantitative evaluation in a systematic review

Phil Spuls\textsuperscript{ad}, LLA Lecluse\textsuperscript{a}, MNF Poulsen\textsuperscript{a}, JD Bos\textsuperscript{a}, RS Stern\textsuperscript{b}, T Nijsten\textsuperscript{c}

\textsuperscript{a}Department of Dermatology, Academic Medical Center, University of Amsterdam, the Netherlands
\textsuperscript{b}Department of Dermatology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, USA
\textsuperscript{c}Department of Dermatology, Erasmus Medical Center, Rotterdam, the Netherlands
\textsuperscript{d}Dutch Cochrane Centre, the Netherlands

*Journal of Investigative Dermatology, in press*
Abstract

Background A large number of clinical measures of psoriasis are used in clinical trials and daily practice. These measures lack uniformity and validation. Valid outcome and severity measures for psoriasis are a pre-requisite for fully informative clinical research and evidence based medicine.

Purpose To identify all clinical measures of psoriasis severity and outcome in use and to evaluate the quality of these measures using clinimetric criteria.

Results We identified 53 separate clinical measures and they 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.

Conclusions None of the psoriasis measures is adequately validated. The Psoriasis Area and Severity Index is the most commonly used clinical measure in research, but has substantial limitations including 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.
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. Since 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.

A systematic review of clinical psoriasis measures in randomized clinical trials (RCTs) up to 2000, identified no less than 44 different clinical measures used for psoriasis. Besides poor standardization of outcomes in clinical trials, there is lack of validation of the used measures. Another review showed that no available measure fulfilled all requirements of a validated instrument for disease assessment. Subsequent reviews concluded that there is no ‘best’ outcome or severity measure for use in clinical trials.

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. 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 since 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.

Methods

Search strategy

To update the list of all clinical psoriasis measures a Medline search was conducted in May 2007, using the term ‘psoriasis’ in combination with the clinical severity measures.
“PASI/ psoriasis area and severity index”, “BSA/ body surface area”, “PGA/ physician global assessment”, “patients global assessment”, “skin index” and “SAPASI/ self assessed PASI” 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 (Table 1).

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 those using ultrasound, laser Doppler velocimetry, transepidermal water loss and stereoimage optical topometry 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 (Fig. 1). Two reviewers independently primarily 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 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. For evaluating the quality of the measures, we adjusted guidelines for HRQoL instruments. Table 3 presents the definitions of the quality criteria (validity, reliability, responsiveness, response distribution, interpretability, ease to administer and uniformity) 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. Nevertheless, we used the PASI for criterion validity, since 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.

Results

Data synthesis (Table 1)
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. Many of these 53 measures were similar and could be regrouped into 11 measures for analysis (Table 2). The search for articles regarding quality criteria of the clinical psoriasis measures identified 6815 articles of which 42 articles were included for data-extraction based on the in- and exclusion criteria (Fig. 1).
Table 1. Search strategy
Figure 1 Flow-chart of the search and selection process for articles on the outcome and severity measures

Inclusion criteria title/abstract:
- Psoriasis
- Clinical severity measures based on physical examination

Exclusion criteria title/abstract:
- Other skin diseases
- RCT’s with outcome measure to compare therapeutic effect
- Other clinical severity measures
  - Psoriatic arthritis measures
  - Nail psoriasis measures
  - Scalp psoriasis measures
- Pure symptom measures for psoriasis
- Pure HRQoL measures
- Objective measures
  - Laboratory measures
  - Measures such as photographs, Doppler, computer

Inclusion criteria full text:
Clinical measures evaluated on
- Validity
- Reliability (original data)
- Responsiveness (original data)
- Response distribution (original data)
- Interpretability
- Ease to administer (original data)
- Uniformity

Exclusion criteria full text:
- Measures that take into account symptoms and/or QoL
- Russian and Chinese articles

42 articles for data-extraction
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 extend 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

Table 2. All retrieved clinical outcome measures re-grouped in 11 main clinical severity and outcome measures for analysis
4. Patient's Global Assessment (PaGA)
   o 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)
   o 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

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)

Initial retrieved clinical outcome measures n=53
Evaluation of the instruments

Table 3 presents the definitions and the grading of the quality criteria used. Table 4 gives an overview of the analyzed clinical psoriasis measures. Table 5 presents the results of the clinimetric evaluation. The available clinimetric data of each instrument are described in more detail 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’, which was originally developed for estimating the surface area of burns.\(^{16}\) 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.\(^{16-21}\) Validity was only tested with the PGA, which correlated strongly.\(^{22}\) In contrast to the varying inter-rater reliability, the intra-rater reliability for area estimation was described as excellent.\(^{16,23,24}\)

The BSA was least improved after two weeks treatment compared to other psoriasis parameters, which is negative for responsiveness.\(^{25}\) Several scales are in use, ordinal and continuous.\(^{12,15,18,26,27}\) The clinically assessed BSA did not differ statistically from objectively assessed BSA, for instance with computer-based image analysis in two studies.\(^{25,28}\) However, several other studies demonstrated that patients and clinicians significantly overestimate the affected area, especially in mild cases and when untrained.\(^{16,17,19,20,24,29}\) The correlation coefficient between BSA assessed by dermatologists and patients ranged from \(r=0.38-0.82.\)\(^{12,14,30-32}\)
<table>
<thead>
<tr>
<th>Property Validity</th>
<th>Definition</th>
<th>score</th>
<th>Quality criteria</th>
</tr>
</thead>
</table>
| 1a. Content validity | The extent to which the type and number of scale items of the outcome measure adequately represent the underlying construct | A | Fulfilment 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 like 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 | $r > 0.7$ (very strong correlation)  
B | $0.5 < r < 0.7$ (strong correlation)  
C | $0.3 < r < 0.49$ (moderate correlation)  
D | $0.29 < r < 0.3$ (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 Correlation 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 | $0.5 < r < 0.7$ (strong correlation)  
C | $0.3 < r < 0.49$ (moderate correlation)  
D | $0.29 < r < 0.3$ (weak correlation)  
E | $r < 0$ (no correlation) |
<table>
<thead>
<tr>
<th>Property</th>
<th>Definition</th>
<th>score</th>
<th>Quality criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reliability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a. test-retest / intrarater reliability</td>
<td>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)</td>
<td>A</td>
<td>ICC or weighted Kappa calculated:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• 81-100% (substantial)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• 61-80% (moderate)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• 41-60% (fair)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• 11-40% (slight)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• 0-10%: (virtually none)</td>
</tr>
<tr>
<td></td>
<td>Only correlation coefficients are calculated:</td>
<td>D</td>
<td>• 81-100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>• 0-80%</td>
</tr>
<tr>
<td>2b. inter-rater reliability</td>
<td>Examines the degree to which multiple observers agree on the assignment of scales (only original data used)</td>
<td>A</td>
<td>ICC or weighted Kappa calculated:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• 81-100% (substantial)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• 61-80% (moderate)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• 41-60% (fair)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• 11-40% (slight)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• 0-10%: (virtually none)</td>
</tr>
<tr>
<td></td>
<td>Only correlation coefficients are calculated:</td>
<td>D</td>
<td>• 81-100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>• 0-80%</td>
</tr>
<tr>
<td>3. Responsiveness</td>
<td>The ability of an instrument to detect changes over time. The instrument should be able to distinguish clinically important change from measurement error</td>
<td>A</td>
<td>As this is not tested for clinical severity measures in psoriasis, we only distinguish between:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Positive information found on responsiveness</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Negative information found on responsiveness</td>
</tr>
<tr>
<td>4. Response distribution</td>
<td>Examines whether the entire range of a scale is used (only original data used)</td>
<td>A</td>
<td>• Positive information found on the usage of the entire range of a scale</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>• Negative information found on the usage of the entire range of a scale</td>
</tr>
<tr>
<td>5. Interpretability</td>
<td>The degree to which one can assign qualitative meaning to quantitative scores. Scores should provide information about what (change in) score would be clinical meaningful.</td>
<td>A</td>
<td>A: Clinical relevant categorisation is defined</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>• Clinical relevant categorisation is not defined</td>
</tr>
<tr>
<td></td>
<td>B:</td>
<td>A</td>
<td>Minimal Important Change is defined</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>• Minimal Important Change is not defined</td>
</tr>
</tbody>
</table>
### Property Definition score Quality criteria

<table>
<thead>
<tr>
<th>Property</th>
<th>Definition</th>
<th>score</th>
<th>Quality criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Ease to administer</td>
<td>The degree to which an outcome measure can easily be used in clinical practice (only original data used)</td>
<td>A</td>
<td>• Fulfilment of all demands: o fulfilling the rating is not time consuming (not exceeding 3 minutes)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>o no extra tools (except score form) are needed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>o the rating is easy understandable fulfilling criteria above, but fulfilling the rating takes between 3 and 7 minutes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>not fulfilling ≥1 of the above criteria</td>
</tr>
<tr>
<td>7. Uniformity</td>
<td>The degree to which there are variations in used scales with the same clinical severity measure</td>
<td>A</td>
<td>• only one rating is used per clinical severity measure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>• more than one rating is used per clinical severity measure</td>
</tr>
</tbody>
</table>

**Table 3. The definitions of the clinimetric properties, grading categories and weighted score for clinical severity outcome measures.**

Criteria were adjusted from Terwee 11 and Both 10

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSA</td>
<td>Estimation of involved body surface area, several scores are used</td>
</tr>
<tr>
<td>Signs</td>
<td>Evaluation of the plaque characteristics erythema, scaling and induration. Erythema and scaling are easily influenced by external factors</td>
</tr>
<tr>
<td>PASI</td>
<td>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</td>
</tr>
<tr>
<td>PGA</td>
<td>The PGA is a 5, 6 or 7-point ordinal rating ranging from “clear” to “very severe psoriasis.”</td>
</tr>
<tr>
<td>PaGA</td>
<td>The PaGA is an ordinal rating ranging from &quot;clear&quot; to “very severe psoriasis assessed by the patient”</td>
</tr>
<tr>
<td>SAPASI</td>
<td>The SAPASI is a structured PASI-like instrument designed for patient self-assessments of severity</td>
</tr>
<tr>
<td>PASS</td>
<td>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</td>
</tr>
<tr>
<td>LS-PGA</td>
<td>The LS-PGA integrates ranges of involved BSA and the overall plaque morphology in which infiltration is given more weight</td>
</tr>
<tr>
<td>SPASI</td>
<td>The SPASI equals the sum of the average redness, thickness and scaling of all the psoriasis lesions multiplied by the % body surface area involved</td>
</tr>
<tr>
<td>PEASI</td>
<td>The PLASI is derived from the PASI but uses actual BSA percentages instead of an area scores</td>
</tr>
<tr>
<td>PLASI</td>
<td>The PLASI is derived from the PASI but uses six BSA groupings with finer partitioning for smaller extents of BSA</td>
</tr>
</tbody>
</table>

**Table 4 Overview of the 11 main clinical psoriasis measures that were analyzed**
<table>
<thead>
<tr>
<th>Components of measures</th>
<th>Content validity</th>
<th>Criterion validity</th>
<th>Clinical</th>
<th>Symptom</th>
<th>HRQoL</th>
<th>Test-retest reliability</th>
<th>Inter-rater reliability</th>
<th>Responsiveness</th>
<th>Response distribution</th>
<th>Min. imp. change</th>
<th>Easy to administer</th>
<th>Uniformity</th>
<th>Number of studies included</th>
<th>References used</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSA</td>
<td>E</td>
<td>-</td>
<td>A</td>
<td>-</td>
<td>-</td>
<td>D</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>-</td>
<td>E</td>
<td>E</td>
<td>18</td>
<td>12;14-20;22-29;31;32</td>
</tr>
<tr>
<td>Erythema</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12;14;22;25-27;31-34</td>
</tr>
<tr>
<td>Scaling</td>
<td>E</td>
<td>B-C</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>E</td>
<td>E</td>
<td>-</td>
<td>E</td>
<td>E</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Induration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clinical severity measures</th>
<th>Content validity</th>
<th>Criterion validity</th>
<th>Clinical</th>
<th>Symptom</th>
<th>HRQoL</th>
<th>Test-retest reliability</th>
<th>Inter-rater reliability</th>
<th>Responsiveness</th>
<th>Response distribution</th>
<th>Min. imp. change</th>
<th>Easy to administer</th>
<th>Uniformity</th>
<th>Number of studies included</th>
<th>References used</th>
</tr>
</thead>
<tbody>
<tr>
<td>PASI</td>
<td>C</td>
<td>X</td>
<td>A-D</td>
<td>D</td>
<td>C-E</td>
<td>A</td>
<td>A</td>
<td>E</td>
<td>E</td>
<td>-</td>
<td>E</td>
<td>A</td>
<td>28</td>
<td>5;6;9;12;13;15;22;26;27;30-32;36-45;47;51-55</td>
</tr>
<tr>
<td>PGA</td>
<td>E</td>
<td>A</td>
<td>A-C</td>
<td>-</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>5</td>
<td>7;13;22;26;52</td>
</tr>
<tr>
<td>PaGA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAPASI</td>
<td>C</td>
<td>B-D</td>
<td>B-D</td>
<td>D</td>
<td>D-E</td>
<td>D</td>
<td>D</td>
<td>A</td>
<td>A</td>
<td>E</td>
<td>A</td>
<td>A</td>
<td>10</td>
<td>12;14;30-32;42-46</td>
</tr>
<tr>
<td>PASS</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td>A</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>LS-PGA</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>-</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>E</td>
<td>A</td>
<td>A</td>
<td>3</td>
<td>13;22;52</td>
</tr>
<tr>
<td>SPASI</td>
<td>C</td>
<td>A</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>E</td>
<td>E</td>
<td>A</td>
<td>1</td>
<td>47</td>
</tr>
<tr>
<td>PEASI</td>
<td>C</td>
<td>A</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>E</td>
<td>E</td>
<td>A</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>PLASI</td>
<td>C</td>
<td>A</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>A</td>
<td>A</td>
<td>1</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 5. The grading of the clinimetric properties of the clinical psoriasis severity measures.
- no information found; X PASI is used as standard and is per definition 100% correlated to itself
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 $r=0.6$. Reliability and other criteria were not tested. Many different scales were in use ranging from 2-8 points. 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. Moderate and strong correlations were found between clinical assessment of erythema versus technical colorimetry, spectroradiometer, two-channel erythema meter, minolta chroma meter and nitric oxide production ($r=0.53-0.79$), to a lesser extent versus laser-doppler flowmetry and erythema reflectance meter ($r=0.3-0.33$). The correlation coefficient between erythema assessed by dermatologists and patients ranged was moderate ($r=0.37-0.4$). 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 ($r=0.53$). Patient versus dermatologist scaling assessment did not correlate. Induration or thickness of psoriatic lesions is probably the most specific parameter of psoriasis activity. Elevation correlated moderately to ultrasound thickness ($r=0.58$). Patient versus dermatologist induration assessment correlated weakly ($r=0.24$) to not.

The Psoriasis Area and Severity Index (PASI)

In 1978, the PASI was developed to assess the effects of retinoids in psoriasis. 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. 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. Response distribution is low, since practically only half of the scale is used. Another issue of the PASI is its responsiveness; when patients reach <10% BSA in any body area, 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. There is no consensus on the interpretability. However, several proposals have...
been done, for instance by Schmitt et al who translated the PASI ranges into the terms ‘mild’, ‘moderate’, and ‘severe’.41

**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 comparison to the 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.13;22;26 The PGA correlated more with BSA than with signs, although the extent of involvement should not be incorporated in the PGA.22 Reliability was calculated to be good and experience appeared to have a negligible effect on PGA assessment.22 The scales are clear and most of the scale is used.13;26

**Patient’s Global Assessment (PaGA)**

No studies were identified assessing clinimetric properties of the PaGA.

**Self-Administered PASI (SAPASI)**

The SAPASI is a structured PASI-like instrument designed for patient self-assessments of severity.42 Patients shade in affected areas on a silhouette of a body to estimate body surface area and complete visual analogue scales for 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.12;30-32;43-45 Interestingly, the SAPASI did not correlate well with HRQoL measurements.43;45 Reliability of the SAPASI was described to be very good.12;46 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.40 Psoriasis is defined as “in remission” when SAPASI = 0, “mild” when 0 < SAPASI < 3, “moderate” when 3 < SAPASI < 15 and “severe” when SAPASI > 15.46

**Psoriasis Assessment Severity Score (PASS)**

The PASS was developed to be simpler and faster than the PASI.27 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 0-140. 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. Interrater reliability was described to be better than with the PASI.27

**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.22 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 to 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.13 Although it has been suggested that this measure is sensitive to clinical change, this is not well documented.13

**Simplified PASI (SPASI)**

The SPASI is mathematically derived from the PASI and is meant to be easy to use.47 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.47

**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 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 scores 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. Most patients score below 360 on a 1200 point scale. Both ΔPEASI and ΔPLEASI corresponded with patients' self-assessments.

The frequency of use
In RCTs 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 only used occasionally (10 and 2 times respectively). Several new clinical severity measures were developed recently and had not yet been cited in psoriasis RCTs published before May 2007.

Summary of results
Only PASS and LS-PGA scored high for ‘content validity,’ since 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 import 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 5 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 updated the list of all clinical psoriasis severity and outcome measures. We showed that there are many different clinical measures for psoriasis in use and that their number still rises. Naldi et al. had identified 44 psoriasis measures in clinical trials up to the year 2000. In our search in 2007 this number had increased to 53 clinical measures for psoriasis.
For quantitative analysis of the main psoriasis measures, the number of eligible articles per measure varied from 0 to 28. Table 5 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. For objectively assessing ease of administration the cut-off points for time of assessment were derived from Schmitt et al.\textsuperscript{48} 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 0.1-0.3\textsuperscript{43,45} This weak correlation between clinical severity and HRQoL was also seen when patients themselves assessed the clinical severity with the SAPASI\textsuperscript{45} We expected that clinical severity would correlate more with quality of life. Especially, since we know that reductions in physical and mental functioning due to psoriasis have found to be similar to those reported in patients with cancer, arthritis, hypertension, heart disease, diabetes and depression.\textsuperscript{49} 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.\textsuperscript{50} Issues like these can be only be identified using well validated, unidimensional measures in research and practice. For this reason, we excluded multidimensional measures such as NPF-PS, BPSS (Beer Sheva Severity Score) and DIDS (Dermatology Index of Disease Severity) from our systematic review. Since 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.6

The adapted quality criteria for evaluating clinical psoriasis measures we used have not been validated, which is a limitation of the study. 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.10 Furthermore, the items and applied criteria we used 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 things. We had no minimum criteria for inclusion, like the use of experienced assessors, a minimum number of patients, an applicable population of patients and unchanging circumstances. Apart from the quality of the included studies, 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.

Since there is no supreme measure, different measures might be ideal for different situations and we might need all of them. 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 to the existing database of studies that used the 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 since it 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.
Acknowledgments
We would like to acknowledge Arnold Leenders, clinical librarian from Central Medical Library of the Academic Medical Center.

Reference List

3. NICE. new NICE guidance to help sufferers of poriasis and psoriatic arthritis. 2006.
11. Terwee CB, Bot SD, de Boer MR et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007; 60: 34-42.

15. Jacobson CC, Kimball AB. Rethinking the Psoriasis Area and Severity Index: The impact of area should be increased. *Br J Dermatol* 2004; **151**: 381-7.


21. Thomas CL, Finlay AY. The ‘handprint’ approximates to 1% of the total body surface area whereas the ‘palm minus the fingers’ does not. *Br J Dermatol* 2007; **157**: 1080-1.


27. Harari M, Shani J, Hristakieva E *et al.* 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. *International Journal of Dermatology* 2000; **39**: 913-8.


47. Louden BA, Pearce DJ, Lang W et al. A simplified psoriasis area severity index (SPASI) for rating psoriasis severity in clinic patients. *Dermatology Online Journal* 2004; **10**.


