Coloring the spots

Diagnosis, measurement instruments and treatment in vitiligo

Lommerts, J.E.

Citation for published version (APA):
CHAPTER 3.1

DEVELOPMENT AND VALIDATION OF THE VITILIGO EXTENT SCORE (VES): AN INTERNATIONAL COLLABORATIVE INITIATIVE.


Journal of Investigative Dermatology 2016 May;136(5):978-84
ABSTRACT

The clinical assessment of vitiligo involves an estimation of the affected body surface area. The most commonly used method is the "palm of hand 1% rule" as integrated in the Vitiligo Area Scoring Index. However, this method can be challenging and time consuming. In this study, we introduce a global Vitiligo Extent Score (VES). In the first part of the study, this measurement instrument was developed and subsequently optimized during a pilot scoring session. In a subsequent stage, the inter- and intrarater reliability of the instrument were tested. Live scoring showed an excellent interrater reliability for the VES (intraclass correlation VES: 0.924 vs. Vitiligo Area Scoring Index: 0.846). Subsequent scoring on pictures was comparable with the live evaluation and demonstrated an excellent intrarater reliability. A high intraclass correlation for the VES (intraclass correlation VES: 0.923 vs. Vitiligo Area Scoring Index: 0.757) was also found in an additional subgroup of patients with extensive vitiligo. Moreover, user-friendliness and timing were scored very favorably. In conclusion, this measurement instrument allows us to monitor accurately and easily the affected body surface area in a standardized way. Moreover, our results provide evidence that the VES can be proposed as a promising tool to measure the vitiligo extent in clinical trials and in daily practice.
**INTRODUCTION**

There is currently a lack of consensus in the standardization of outcome measures in vitiligo, which makes it difficult to compare the outcomes of different studies and therefore hampers evidence based recommendations.\(^1\,^2\) Previously, two systematic reviews on vitiligo outcome measurements were published in 2012\(^3\,^4\) and demonstrated that 25 different outcomes were included that were measured by 11 different instruments. A consensus\(^5\) on the core outcome sets for clinical trials in vitiligo has very recently been reached ("what" should be measured)\(^6\). Three core outcomes were deemed essential (i.e., to be measured in every trial): repigmentation, side effects, and maintenance of gained repigmentation. The next crucial step is to identify appropriate instruments that assess the core outcome domains of vitiligo. The measurement properties (e.g., validity, reliability, and responsiveness) of the available instruments were critically appraised using the Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist.\(^7\,^8\) Strong evidence was only found for a positive internal consistency of the Dermatology Life Quality Index.\(^4\)

Currently, evidence for measurement instruments for vitiligo is rather scarce. The Vitiligo European Task Force (VETF) proposed in 2007\(^9\) the Vitiligo European Task Force assessment tool (VETFa) combining analysis of extent, stage of disease, and disease progression. The Vitiligo Area Scoring Index (VASI) is another score system\(^10\) that offers a disease severity index. Recently, the VASI and VETFa were reported\(^11\) to be reliable and responsive instruments to assess the degree of depigmentation in patients with vitiligo. However, authors reported that for use in clinical practice, caution is needed when interpreting score changes in individual patients because of the relatively large smallest detectable change (SDC). The SDC concerns the minimal difference in disease extent that can be accurately measured by the instrument. Nonetheless, even the VASI, which is the most frequently cited tool, is so far not widely used by researchers\(^3\), which might point to limitations in user-friendliness.

A good and validated scoring instrument permits valid measurement of disease extent, a necessary feature that allows the results of clinical trials to be accurately interpreted and compared. It can help in improvement of management, estimation of prognosis, understanding of disease progression, and the development of therapeutic options in vitiligo. Furthermore, it would be possible to categorize disease severity and allocate resources for instance considering reimbursement criteria.
On the basis of this need, we introduce in this study a clinical scoring instrument (i.e., Vitiligo Extent Score (VES)) that fits in the outcome domain requirements. The key principles in the design of this tool were intelligibility, clarity, simplicity, logic, feasibility, and availability for use in both clinical and research settings.

**MATERIALS AND METHODS**

**Study design and ethics**

This trial was conducted on behalf of the International Vitiligo Score Working group in close cooperation with the VETF and VGICC group (the International Vitiligo Score Working Group comprises: Nanja van Geel, Reinhart Speeckaert, Janny Lommerts, Marcel Bekkenk, Albert Wolkerstorfer, Viktoria Eleftheriadou, Khaled Ezzedine, Mauro Picardo, and Alain Taïeb). The development and validation of the VES was performed by six scoring sessions as outlined below. This included a face-to-face scoring session (session 2) and a scoring session by using pictures (sessions 1, 3-6).

The results of the VES were compared with the VASI. Clinical disease extent was scored independently by the participating physicians using standardized scoring sheets for both methods. To minimize the bias of subsequent scoring by two methods (VES and VASI), a two-sequence design was randomly adopted (system: GraphPad Software, La Jolla, CA).

The study was approved by the local ethics committees and was performed according to the Declaration of Helsinki. Written informed consents were obtained from all participants. The COSMIN checklist was used as a guidance for designing and reporting our study.

**Assessors and participants**

Assessors with different levels of experience (dermatology residents, dermatologists, and vitiligo experts) were selected by the principle investigator. Vitiligo experts were defined as dermatologists with at least 3 years of clinical experience in a tertiary vitiligo center. In general, assessors received a short training on how to use both tools (VES and VASI). One day before the face-to-face scoring session (session 2) a second training session was delivered. This training session included mainly the practicalities of the study.
Participants were volunteers of all ages with clinically diagnosed vitiligo (non-segmental). Both patients on or without current treatment were included. Exclusion criteria were unclear diagnosis and patients with segmental vitiligo. In case of mixed vitiligo, only nonsegmental lesions were evaluated. Patients were recruited in a consecutive manner for sessions 1-4. For sessions 5 and 6, patients with vitiligo (non-segmental) and an estimated BSA of more than 5% were selected retrospectively from the existing patients’ database. Patients for sessions 1 and 5 were recruited from the Ghent University Hospital (Belgium), whereas patients for session 2 were recruited from the Academic Medical Centre in Amsterdam (the Netherlands). Patients’ pictures, which were included in session 2, were used in sessions 3 and 4, whereas pictures from session 5 were reused in session 6.

**Design of the model**

**Session 1**
The original concept of the VES was based on clinical pictures of 10 different body locations reflecting the degrees of involvement. Percentage of depigmentation in each picture was calculated by ImageJ analysis (e.g., 1%, 5%, 10%, 25%, 50%, and 75%). Pictures were designed to mimic the natural evolution of vitiligo (nonsegmental). The purpose was to create a template that would allow an easy comparison between the “real life” patients and the scoring sheet. Total and final VES patient’s score was the sum of all surface measurements that were calculated from a converting table.

Session 1 was intended to optimize the initial version of the VES template. Based on the experience/results of small-scale testing before conducting session 1, several changes of VES were made to allow a more precise scoring, such as division to left-right and up-down and possibility of selecting 1/2 and 1/4 of the involved area in the first two categories. Complete depigmentation could be chosen as a separate option. After the first scoring session (=session 1), the option of a “plus” or “minus” was added in the other categories, representing 1/4 more or less than the representing picture (=19-item ordinal scale). The >75% depigmentation option represents depigmentation higher than 75% but less than 100% (i.e., complete depigmentation). The final version of the VES included 19 body areas (Figure 1, Supplementary Figures S1 and S2 online). A comparison between the two versions (6-item vs. 19-item ordinal scale) showed that the detailed template resulted in an improved interrater reliability. Therefore, the detailed scale was incorporated in sessions 2-6.
Validation phase

Different scoring sessions were carried out to validate the VES by assessing its inter- and intrarater reliability. Sessions that involved evaluation of the same patients scored previously were separated by 2 weeks to avoid recall bias. If face-to-face evaluation was compared with digital pictures, the pictures were taken during the live evaluation to ensure stability of the lesions during the scoring process.

Session 2

Session 2 was conducted to measure the interrater reliability of face-to-face scoring. At the beginning of the session, each patient was randomized to a room. The assessors were assigned to examination rooms to assess the VASI and VES in a randomized order. Assessors were given 15 minutes per patient evaluation.

Sessions 3 and 4

The aim of session 3 was to investigate the interrater reliability of scoring on pictures. Moreover, the intrarater reliability of face-to-face scoring (session 2) and scoring of pictures was assessed and compared with scoring on pictures at two separate time points (sessions 3 and 4). Standardized digital clinical pictures were taken and scored by the same investigators (session 3), including a retest 2 weeks later (session 4). Again, the VASI and VES were assessed in a randomized order.

Sessions 5 and 6

To evaluate the VES in patients with extensive vitiligo (BSA > 5%), digital clinical pictures of 10 new patients were evaluated in an additional session (session 5). Reassessment on these digital pictures took place 2 weeks later (session 6). One vitiligo expert was not able to participate in this session because of practical reasons. Patients who took part in sessions 5 and 6 were analyzed separately.

Statistics/Data analysis

Statistical analyses were performed using SPSS 22.0 (SPSS Science, Chicago, IL). For data analyses, the interrater and intrarater reliability were assessed and analyzed separately for all different scoring sessions. The interrater and intrarater reliability were determined by their ICCs. The ICC was calculated in a two-way mixed model with absolute agreement and reported as single measures. An ICC of more than 0.8 was found to be acceptable. We also aimed to determine the SDC, that is, change beyond measurement error. Test-retest reliability was investigated using data assessed on the same pictures at different time points. The SDC was calculated by the formula: $SDC_{95} = 1.96 \times \sqrt{2} \times Standard\ Error\ of\ Measurement\ (SEM)$; where $SEM = SD \times \sqrt{1-ICC}$.
Two types of settings to perform the scoring (face-to-face scoring vs. scoring of digital pictures) were compared by using the ICC and the Wilcoxon signed-rank test to exclude a systematic over- or underestimation of one of the assessment methods. For not normal distributed data, logarithmic transformation was carried out if a normal distribution could be reached.

**RESULTS**

**Interrater reliability**

**Session 1**
The initial version of the VES template was tested by 11 observers including 3 vitiligo experts, 4 dermatologists, and 4 trainees. Pictures of 31 patients with vitiligo (non-segmental) (Table 1) were scored with a mean area of depigmentation of 3.65% (range: 0.02-58.72%). The intraclass correlation (ICC) was 0.960 (95% CI: 0.937-0.978). No marked differences were found between the ICC of the vitiligo experts (ICC: 0.973 [95% CI: 0.952-0.983]), the dermatologists (nonvitiligo experts) (ICC: 0.963 [95% CI: 0.934-0.980]), and the trainees (ICC: 0.956 [95% CI: 0.926-0.977]) (Table 2).

**Session 2**
Twenty patients with vitiligo (19 nonsegmental and 1 mixed vitiligo) took part in a 1-day face-to-face evaluation session (session 2, Table 1). Assessors included six (four men and two women) clinicians with a wide range of expertise in vitiligo (three vitiligo experts, two dermatologists, and one resident).

**Table 1** - Demographic details of the patients.

<table>
<thead>
<tr>
<th></th>
<th>Session 1</th>
<th>Sessions 2 (and 3 and 4)</th>
<th>Sessions 5 (and 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>11</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Females</td>
<td>20</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Mean age</td>
<td>32.1</td>
<td>42.25</td>
<td>42.8</td>
</tr>
<tr>
<td>Age of onset</td>
<td>22.81</td>
<td>33.15</td>
<td>27.5</td>
</tr>
<tr>
<td>Photo skin type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>III</td>
<td>21</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>IV</td>
<td>5</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>VI</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>
The total scores per patient ranged for the VES (Figure 1) from 0.01% to 34.34%, with a median of 2.17% (interquartile range [IQR]: 0.65-4.72%), and for the VASI from 0.004% to 35.45%, with a median of 2.62% (IQR: 0.85-5.55%). The majority of patients (17 of 20) included in this session had a body surface area (BSA) of less than 10% (Figure 2a). Wilcoxon analysis showed a significant higher estimation of the BSA with the VASI compared with the VES (P < 0.001). This was illustrated by a decreased slope of the regression line in the scatter plot (Figure 2a). However, the median difference (0.29% [IQR:0.2 to 1.13%]) in estimated BSA between the two methods was limited. Interrater reliability analysis showed for the VES an ICC of 0.924 (0.862-0.965) and for the VASI an ICC of 0.846 (0.737-0.926). This corresponded to an SDC95 of 4.68% for the VES and 7.79% for the VASI (Table 2).

The Bland-Altman plot (Figure 3a) demonstrated a tendency of increased spread of the data in patients with more extensive vitiligo. This suggests that the accuracy of at least one of the measurement tools is dependent on the magnitude of the measurements.

**Session 3**

This session involved analysis of digital pictures of 20 patients from session 2. The above-mentioned six assessors (from session 2) independently evaluated these pictures. In this session, the scores of the VES ranged from 0.01% to 34.21%, with a median of 1.97% (IQR: 0.64-5.22), and of the VASI from 0.01% to 29.25%, with a median of 2.40% (IQR: 0.84-6.20%) (Figure 2b). Wilcoxon analysis showed higher values for the VASI compared with the VES (P = 0.001). In total, 85 of 120 cases (70.83%) were rated higher with the VASI compared with the VES. However, the median difference between the two score methods was again only 0.32% (IQR: 0.05 to 1.26). In addition, the ICC was 0.922 (95% CI: 0.859-0.965) for the VES and 0.829 (95% CI: 0.701-0.919) for the VASI. This resulted in a very similar SDC95 compared with the live scoring (Table 2).

**Session 4**

During this reassessment session, the ICC of the VASI improved from 0.829 to 0.904 although this was still within the limits of the 95% CI from session 3 and remained slightly inferior to the VES (ICC: 0.943).
**Figure 1** - Scoring sheet Vitiligo Extent Score (VES).

Example of the new scoring system. The 19 different body areas are scored separately depending on the vitiligo extent (online version available at: www.vitiligo-calculator.com)
Table 2 - Interrater reliability.

<table>
<thead>
<tr>
<th>Scoring session</th>
<th>Number of patients</th>
<th>Evaluation of patients</th>
<th>Number of raters</th>
<th>ICC (95% CI)</th>
<th>SDC&lt;sub&gt;95&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pilot</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sessions to assess optimal tool</td>
<td>1</td>
<td>31</td>
<td>Pictures</td>
<td>11</td>
<td>VES: 0.960 (0.937-0.987)</td>
</tr>
<tr>
<td><strong>Assessing interrater reliability of VES versus VASI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scoring of live patients</td>
<td>2</td>
<td>20</td>
<td>Live patients</td>
<td>6</td>
<td>VES: 0.924 (0.862-0.965)</td>
</tr>
<tr>
<td>Scoring patients of session 2</td>
<td>3</td>
<td>20</td>
<td>Pictures</td>
<td>6</td>
<td>VES: 0.922 (0.859-0.965)</td>
</tr>
<tr>
<td>Retest after 2 wk</td>
<td>4</td>
<td>20</td>
<td>Pictures</td>
<td>6</td>
<td>VES: 0.943 (0.897-0.974)</td>
</tr>
<tr>
<td><strong>Assessing interrater reliability of VES versus VASI in patients with advanced vitiligo</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scoring of extensive vitiligo patients</td>
<td>5</td>
<td>10</td>
<td>Pictures</td>
<td>5</td>
<td>VES: 0.923 (0.828-0.977)</td>
</tr>
<tr>
<td>Retest of patients with extensive vitiligo</td>
<td>6</td>
<td>10</td>
<td>Pictures</td>
<td>5</td>
<td>VES: 0.924 (0.817-0.978)</td>
</tr>
</tbody>
</table>

CI, confidence interval; ICC, intraclass correlation; SDC, smallest detectable change; VASI, Vitiligo Area Scoring Index; VES, Vitiligo Extent Score.

**Sessions 5 and 6**

Ten new patients with extensive vitiligo were evaluated by five of the six previously mentioned physicians. For the VES, the scores varied from 4.48% to 63.76%, with a median of 14.27% (IQR: 8.81-21.92%), and for the VASI from 8.60% to 71.00%, with a median of 20.38% (IQR: 12.99-27.05%) (Figure 2c). Similar to previous sessions, paired analysis showed higher values for the VASI as compared with the VES (P < 0.001). In total, 38 of 50 cases (76%) were rated higher by the VASI compared with the VES with a median difference of 4.85% (IQR: 0.80-10.22).

The Bland-Altman plot (Figure 3b) displayed an increased dispersion of the differences between the VES and the VASI compared with the face-to-face scoring session (session 2). The data are more spread in patients with extensive vitiligo with higher limits of
agreement: limits of agreement extensive vitiligo (session 5): 17.3 to 7.4 compared with limits of agreement live scoring (session 2): 5.3 to 6.7. The higher scoring of the VASI compared with the VES becomes more pronounced in patients with extensive vitiligo as the mean difference between the VES and VASI is 5.0% BSA in patients with extensive vitiligo compared with 0.7% BSA in the live session. Analysis of these new patients in session 5 demonstrated an ICC of 0.757 (95% CI: 0.422-0.927) for the VASI and 0.923 (95% CI: 0.828-0.977) for the VES. For the VASI the SDC$_{95}$ was 20.60% and for the VES 11.60%. Reassessment of the pictures 2 weeks later showed very similar results that are summarized in Table 2.

**Figure 2** - Correlations between the measured percentages of affected BSA.

Correlation analysis between VES and VASI for live scoring of patients (=session 2) (a) and scoring on pictures of consecutive patients and patients with extensive vitiligo (session 3 (b) and session 5 (c)). Correlations between live scoring (=session 2) and picture scoring (=session 3) for (d) VES and (e) VASI and between scoring of patients with extensive vitiligo at two time points (=sessions 5 and 6) for (f) VES and (g) VASI. BSA, body surface area; VASI, Vitiligo Area Scoring Index; VES, Vitiligo Extent Score.
Intrarater reliability

The intrarater reliability between the face-to-face scoring (session 2) and scoring of pictures (session 3) was determined to assess the reliability of scoring of digital pictures. The ICC was good for the VASI (ICC = 0.829; 95% CI [0.701-0.919]) and excellent for the VES (ICC = 0.922; 95% CI [0.859-0.965]), which confirms the validity of scoring of digital pictures (Figure 2d and e). In addition, Bland-Altman plots and Wilcoxon analysis showed no significant difference between patients’ or pictures scoring. This confirmed the accuracy to validate the developed scoring tool further by picture analysis.

The intrarater reliability between session 3 and session 4 (=test-retest on pictures) was comparable for the VES and the VASI (VES: ICC = 0.943 [95% CI: 0.897-0.974]; VASI: ICC = 0.904 [95% CI: 0.802-0.959]). There was no statistically significant difference between the mean scores obtained in the two different sequences.

In patients with extensive vitiligo (sessions 5 and 6), the intrarater reliability tended to be better for the VES compared with the VASI, illustrated by a closer clustering of the data around the regression line of the scatter plot (Table 3, Figure 2f and g).

Table 3 – Intrarater reliability.

<table>
<thead>
<tr>
<th>Scoring session</th>
<th>Number of scoring session</th>
<th>Number of patients</th>
<th>Evaluation of patients</th>
<th>Number of raters</th>
<th>ICC (95% CI)</th>
<th>SDC_95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessing intrarater reliability of VES versus VASI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Live scoring versus picture scoring</td>
<td>2 versus 3</td>
<td>20</td>
<td>Live patients versus pictures</td>
<td>6</td>
<td>VES: 0.922 (0.859-0.965) VASI: 0.829 (0.701-0.919)</td>
<td>VES: 4.63 VASI: 7.91</td>
</tr>
<tr>
<td>Picture scoring versus retest on pictures after 2 wk</td>
<td>3 versus 4</td>
<td>20</td>
<td>Pictures versus pictures</td>
<td>6</td>
<td>VES: 0.943 (0.897-0.974) VASI: 0.904 (0.802-0.959)</td>
<td>VES: 3.83 VASI: 5.81</td>
</tr>
<tr>
<td>Assessing intrarater reliability of VES versus VASI in patients with extensive vitiligo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test versus retest after 2 wk</td>
<td>5 versus 6</td>
<td>10</td>
<td>Pictures versus pictures</td>
<td>5</td>
<td>VES: 0.924 (0.817-0.978) VASI: 0.812 (0.539-0.944)</td>
<td>VES: 11.52 VASI: 18.16</td>
</tr>
</tbody>
</table>

CI, confidence interval; ICC, intraclass correlation; SDC, smallest detectable change; VASI, Vitiligo Area Scoring Index; VES, Vitiligo Extent Score
Chapter 3.1

Feasibility and subjective evaluation by the raters

The mean time for evaluation of digital pictures by using the VES for all 31 patients (session 1) was 1.39 minutes/patient (range: 1.03-1.7 minutes/patient). No differences between the raters with variable experience could be observed. Both the patients with mild-moderate (session 3) and extensive vitiligo (session 5) were scored faster with the VES compared with the VASI (session 3 = VES: 1 minute and 17 seconds vs. VASI: 1 minute and 42 seconds \(P = 0.089\); session 5 = VES: 1 minute and 30 seconds vs. VASI: 2 minutes and 34 seconds \(P = 0.015\), based on 5 and 4 observers, respectively).

The subjective evaluation (user-friendliness, rapidity, feeling of reliability) scored by the six raters at the end of the live scoring day (session 2) was high for the VES (Table 4) and highlights the user-friendliness of the tool.

Table 4 - Physician’s satisfaction of using the VES and VASI.

<table>
<thead>
<tr>
<th>Question</th>
<th>VES Scale 0-10 (mean)</th>
<th>VASI Scale 0-10 (mean)</th>
<th>P-value</th>
<th>VES Likert 1-5 (mean)</th>
<th>VASI Likert 1-5 (mean)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>User-friendliness of tool</td>
<td>8</td>
<td>4.83</td>
<td>0.002</td>
<td>4.5</td>
<td>2.84</td>
<td>0.026</td>
</tr>
<tr>
<td>Rapidity of tool</td>
<td>8</td>
<td>5</td>
<td>0.015</td>
<td>4.3</td>
<td>2.67</td>
<td>0.041</td>
</tr>
<tr>
<td>Feeling of reliability of tool</td>
<td>7.5</td>
<td>5.67</td>
<td>0.015</td>
<td>4</td>
<td>3</td>
<td>NS</td>
</tr>
</tbody>
</table>

VASI, Vitiligo Area Scoring Index; VES, Vitiligo Extent Score

Figure 3 - Bland-Altman plots.

Differences in the measured affected body surface area between the VES and VASI for (a) the live scoring session and (b) the scoring of patients with extensive vitiligo. VASI, Vitiligo Area Scoring Index; VES, Vitiligo Extent Score.
DISCUSSION

In this study, we developed a practical measurement instrument called the VES. After a careful construction process, we validated this tool. A comparison with a previously developed scale (VASI) was included. In all sessions, the ICCs of the VES were well exceeding the cutoff point of acceptable interrater reliability. The ICCs of the VES remained very high (>0.90) also in patients with extensive vitiligo. Moreover, we could demonstrate the very good intrarater reliability of the VES (test-retest).

As could be expected, the SDC were correlated to the affected BSA. In patients with a BSA less than 5%, which accounts for the majority of patients with vitiligo in daily practice, the SDC$_{95}$ were small (SDC$_{95}$ VES: 1.31% vs. SDC$_{95}$ VASI: 1.85%). Although the VASI performed overall reasonably well for a tool that was essentially constructed to measure repigmentation during UVB treatment, the SDC$_{95}$ of the VES were 36-44% lower (Table 2). Our data with respect to the reliability and SDC of the VASI are in agreement with the results of the study by Komen et al. They reported an SDC of 7.1% for the VASI in a patient population similar to the study group in sessions 2-4 of this study.

The main strengths of the VES lie in the areas of clarity, userfriendliness, and intuitive use. In our opinion, clinical pictures that are used in the VES improve the accessibility of using the tool. Furthermore, the VES can be used to assess the extent of vitiligo for 19 separate areas of the body and therefore improves disease monitoring of the patients. In addition, the possible translation of the VES into categories (Grade 0-6: involved area per region 0%, 1%, 5%, 10%, 25%, 50%, 75%) may be an important advantage compared with currently available tools. This conversion into a certain degree of involvement per area may improve communication between physicians.

The estimated BSA values were higher for the VASI compared with the VES. The problem of the overestimation of the BSA with the 1% hand rule has been described in other skin disorders such as psoriasis. This is probably also due to the fact that the flat closed patient’s hand does not exactly account for 1% of the BSA, but in fact represents only 0.70-0.76% of the total BSA. Nonetheless, final statements addressing this issue may only be resolved using exact digital imaging techniques (e.g., 3D imaging) quantifying the affected BSA. Digital image instruments are an alternative to measure the vitiligo extent in an objective way, although they are often more time consuming and expensive which limits their worldwide use. Furthermore, many currently available digital tools are often still based on a 2D translation of a 3D reality.
This study assists current international efforts in determination of a core outcome set for vitiligo and assessment of instruments to measure it. The main pitfalls of scoring instruments in cutaneous disorders are the insufficient validation and lack of international consensus. The construction and validation of the VES was initiated by an international “Vitiligo Score Working group” supported by the VETF/Vitiligo Global Issues Consensus Conferences (VGICC) and was discussed at the vitiligo meetings of international congresses (23rd and 24th European Academy of Dermatology and Venereology congress and 23rd World Congress of Dermatology). This may facilitate its widespread implementation. Further validation procedures will also be performed within this international framework. For example, we aim to evaluate the responsiveness of the VES, integration of a disease activity measure as well as a comparison to a digital image analysis tool in the near future.

The comparison between the live scoring and evaluation on clinical pictures demonstrated similar results in this study. This will facilitate future validation processes that can be performed on digital pictures, allowing a larger number of observers to participate in an international setting. In addition, we plan to evaluate the VES as a patient reported outcome measure (self assessment tool). Furthermore, for further implementation and optimization of the feasibility and accessibility of the VES, an internet and smartphone application will be developed (available at: www.vitiligo-calculator.com).

Possible limitations of this study include the fact that the majority of the investigated patient population was of photo skin type 2-4. Moreover, data were collected in two tertiary centers.

**Conclusion**

This developed and validated VES is a feasible, fast, and user friendly measurement instrument for clinicians. In this study, we validated this tool, which showed an excellent inter- and intrarater reliability and confirmed its user-friendliness. The possible translation of the VES into clinically relevant categories of extent might be an important advantage compared with currently available tools and can be helpful for inclusion criteria in clinical trials and grading outcomes. It remains however necessary to establish whether the VES is sensitive enough to changes in response to treatment (including follicular repigmentation). We hope that the VES can be incorporated as a preferred instrument in the international consensus on core outcome sets for future vitiligo trials.
REFERENCES

Supplements Figure S1 - Example VES.
User instructions of the VES

1. Choose the pictures which are best representing the vitiligo extent

2. If the pictures are not fully representing the extent

   - Indicate "1/2" or "1/4" if the vitiligo extent is 1/2 or 1/4 of the first picture
   - Mark "-" or "+" if the extent is less or more than the depicted figures
   - ">75" if involvement is more than the last picture but not fully depigmented
   - 100% depigmented

3. Calculate the total BSA: www.vitiligo-calculator.com (available soon)

Supplements Figure S2 - User instructions of the VES.