

Reliability and validity of the Vitiligo Signs of Activity Score (VSAS)

N. van Geel ,¹ T. Passeron,^{2,3} A. Wolkerstorfer,⁴ R. Speeckaert ¹ and K. Ezzedine ⁵¹Department of Dermatology, Ghent University Hospital, Ghent, Belgium²Department of Dermatology, Université Côte d'Azur, CHU Nice, Nice, France³Department of Dermatology & INSERM U1065, C3M, Nice, France⁴Department of Dermatology, Institute for Pigment Disorders and Infection & Immunity Institute Amsterdam UMC, Amsterdam, the Netherlands⁵Department of Dermatology, Henri Mondor Hospital, UPEC-Université Paris-Est Créteil, Paris, France

Summary

Correspondence

Nanja van Geel.

Email: Nanja.vangeel@ugent.be

Accepted for publication

13 February 2020

Funding sources

The research activities of N.v.G. are supported by the Scientific Research Foundation – Flanders (FWO Senior Clinical Investigator: 1831512N) and a LEO Foundation grant (LEO Foundation project reference number LF16092).

Conflicts of interest

N.v.G. has been a consultant and/or investigator for Pfizer, Laboratoire Génévrier and Incyte; and was involved in the preparative phase (design and pilot testing) of the VSAS and VSAS subscores.

This work was previously presented at the Vitiligo International Symposium, Detroit, 10 November 2018 and the World Congress of Dermatology, Milan, 10 and 14 June 2019.

DOI 10.1111/bjd.18950

Background The associations between disease activity and several clinical signs in vitiligo have been described, but a widely accepted and validated scoring system is lacking.

Objectives To validate the Vitiligo Signs of Activity Score (VSAS) for physicians.

Methods Three visible clinical signs were scored on 15 body locations: confetti-like depigmentation (c), Koebner phenomenon (k) and hypochromic areas/borders (h). The inter- and intrarater reliability of the global VSAS and VSAS subscores (c-VSAS, k-VSAS and h-VSAS) were tested by four and three raters (physicians), respectively. Construct validity and feasibility were evaluated.

Results The VSAS demonstrated good inter-rater reliability, with an intraclass correlation coefficient (ICC) of 0.87 in the first round and 0.90 in the second round. The intrarater reliability ICCs were all ≥ 0.86 . The inter-rater reliabilities of the subscores were excellent for c-VSAS and fair for k-VSAS and h-VSAS (ICC 0.83, 0.51 and 0.53, respectively, in the first round). Evidence for construct validity was provided. The completion time by the raters (median 2.18 min per patient) improved during the second round (median 1.33 min per patient). A limitation of the study is the low number of patients, mainly of skin phototypes II–III, from a single tertiary centre.

Conclusions The VSAS appears to be a valid and reliable instrument to score visible clinical signs linked to disease activity in a standardized way.

What is already known about this topic?

- Evidence exists for a possible link between several visible clinical signs in vitiligo and disease activity.
- A widely accepted and validated scoring system to quantify these clinical signs is lacking.

What does this study add?

- The Vitiligo Signs of Activity Score (VSAS) underwent preliminary validation and may assist quantification of visible clinical signs linked to disease activity in a standardized way in clinical practice and trials.

What are the clinical implications of this work?

- VSAS may be used for future trials that aim to establish the clinical significance of the specific visible clinical signs in vitiligo in a more controlled setting.

Although the course of vitiligo is known to be cyclic, with flares and remission, the exact rhythmicity is still unpredictable. Currently, it is well demonstrated that the depigmentation process is linked to an immune destruction of melanocytes. Based on these data new treatments are under development.^{1,2} At the moment, there are three aims for optimal treatment of vitiligo: (i) halting disease progression, (ii) repigmentation by stimulation of melanocyte differentiation and proliferation and (iii) prevention of relapses.^{3–6} For now, the clinical assessment of disease activity at a first visit or at a specific single timepoint relies on the presence of clinical signs such as Koebner phenomenon, hypochromic areas/borders and confetti-like lesions.⁷ This information can be relevant for clinical trials and clinical practice (e.g. start a treatment to halt progression) but it does not allow quantification of disease activity.⁸

Currently, available outcome measures for disease activity in vitiligo are scarce. So far, the Vitiligo Disease Activity Index (VIDA) score is the most frequently cited measurement instrument by researchers, although it is constructed as a recall-based patient-reported outcome measure. However, this VIDA score was demonstrated to be unreliable.⁹ No validation studies on physician-reported outcome measures specifically designed to score signs linked to disease activity in a combined score have been reported so far in vitiligo.

For many years several clinical signs such as Koebner phenomenon, hypochromic or trichromic areas and confetti-like depigmentation have been suggested to be linked to disease activity in vitiligo.^{10–12} Based on a recent systematic review, evidence for a possible link was provided.⁷ However, the interpretation of the literature was hampered by the lack of uniform definitions and firmly established assessment methods to evaluate these signs. Moreover, no appropriate instrument to quantify these clinical signs could be identified that met the minimal requirements for acceptable measurement properties (e.g. validity, reliability and responsiveness). Such an instrument is required in clinical practice and trials. In addition to the general need for a validated instrument to quantify the current state of the clinical signs of disease activity this instrument is also mandatory to demonstrate the real prognostic value of these clinical signs during future prospective trials.

The aim of this study was to assess the reliability, validity and feasibility of a measurement instrument to assess and quantify clinical signs linked to disease activity in vitiligo.

Materials and methods

Study design, ethics and construction of a scoring system

This study was performed by a team of vitiligo experts on behalf of the international Vitiligo Score Working Group and was supported by the Vitiligo European Task Force and the Vitiligo Global Issues Consensus Conferences (VGICC) group. The work was presented and discussed during two international special interest vitiligo meetings: the Vitiligo International

Symposium, Detroit, 10 November 2018; and the VGICC meeting at the World Congress of Dermatology, Milan, 10 and 14 June 2019.

The design of the scoring instrument was initiated, evaluated (during clinic practice) and modified (layout-related corrections, grading more specified) in a preparatory phase at the Ghent University Hospital (N.vG.). Several key elements were considered during this phase such as clinical relevance, simplicity, feasibility and availability for use in both clinical trials and daily practice. The clinical relevance of items included in the VSAS was supported by the opinion of vitiligo experts provided during the VGICC meeting in Rome, 2016 (Appendix S1; see Supporting Information). Based on this concept and a systematic review, three items were considered by the team of European vitiligo experts (N.vG., K.E., A.W., T.P.) to be worthwhile to evaluate within this study. The selection of items was based on discussions by phone and a decisive face-to-face meeting (Nice, France, 16 August 2018; N.vG., K.E., A.W., T.P.). The selected items were (i) confetti-like depigmentation, (ii) Koebner phenomenon and (iii) hypochromic areas/borders. In addition, the definitions of each sign were discussed. The term 'trichrome vitiligo' was not included based on current confusion related to definition and uncertainties of its relevance.⁷ It was decided by the scoring team not to include 'inflammatory borders' as an item. The main reasons were that the prevalence of inflammatory vitiligo is thought to be very low and that the remaining questions related to its exact definition and assessment need to be clarified first in future trials.

The main numerical output (score) obtained by this measurement instrument was the overall VSAS. This score was based on the presence of visible clinical signs within 15 predefined body locations (Figure 1). Most of the locations are based on the anatomical regions of the Vitiligo Extent Score (VES), representing classical anatomic units.^{13,14} Evaluation of all 15 body locations was considered to be important, as vitiligo can be active in limited or several areas. The presence of at least one of the three signs in a specific area was coded 1, with a score of 0 if none of the three signs was present, resulting in a score between 0 and 15. Each body area has the same impact on the total score. A similar score (0–15) was performed for each clinical sign separately generating subscores: confetti-like depigmentations (c-VSAS), Koebner phenomenon (k-VSAS) and hypochromic areas/borders (h-VSAS) (Figures 1 and 2).

Within this study, in addition to the subscores an option for grading the degree of each clinical sign was included (grade 1–3; Figure 2), reflecting the intensity per area. This intensity was measured as the estimated number of clinical signs around a representative lesion for c-VSAS and the estimated number per demarcated area for k-VSAS and h-VSAS, with grades 1–3 corresponding overall to 'somewhat present', 'clearly present' or 'very clearly present'. Based on the grading per area, in addition one global (total body) grade per sign was assigned (considered as the grade that was most evident on average for a specific sign). Both the VSAS subscores and grading per sign were considered as secondary outcomes of the measurement instruments.

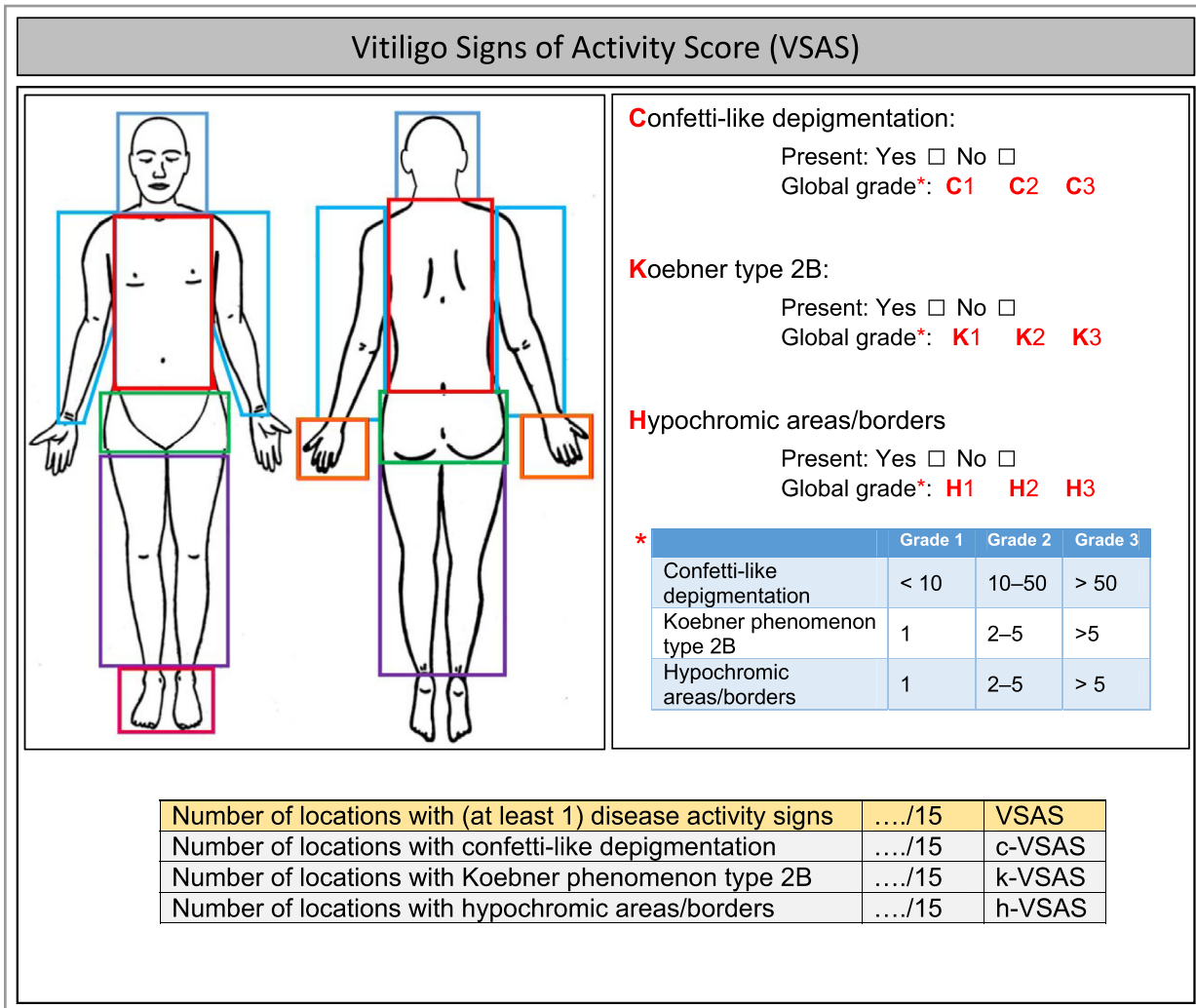


Figure 1 Measurement instrument used to score the Vitiligo Signs of Activity Score (VSAS) and subscores (c-VSAS, k-VSAS, h-VSAS) within 15 body locations, including minor modifications. The overall VSAS is based on the presence of (at least one) visible clinical signs within the 15 predefined areas resulting in a score between 0 and 15. A similar score (0–15) can be performed for each clinical sign separately generating the subscores: confetti-like depigmentations (c-VSAS), Koebner phenomenon (k-VSAS) and hypochromic areas/borders (h-VSAS). If a clinical visible sign is present one can add/write in each demarcated areas the applicable abbreviation for each sign (C, K, H) and add optionally the most applicable grade for that area (e.g. C1; C2; C3; K1; K2; K3; H1, H2, H3). Based on the grading per area, in addition one ‘global grade’ (total body grade) per sign can be assigned, which can be considered as the grade that is most evident on average for a specific sign.

Scoring was introduced during a training session (by N.vG.) of about 10–15 min, which was supported by several clinical photographs and formulated definitions, both serving as a reference for the scoring.

For each visible clinical sign (Figure 2) the definitions formulated were as follows. (i) *Confetti-like depigmentations*: ‘confetti-like depigmentations are grouped small/pinpoint-sized hypo- and depigmented macules.’^{12,14} (ii) *Koebner phenomenon* (definition included in two papers):^{15,16} ‘Koebner type 2B is considered to be positive if at clinical examination and Wood’s light inspection, depigmentations are present and clearly induced by trauma (linear, punctiform, crenate)’.^{15,16} (iii) *Hypochromic areas/borders* (translated from a discussed clinical description based on examples of clinical pictures): ‘hypochromic areas/borders are ill-defined borders or hypopigmented areas located at the contours of the vitiligo lesion’.

The study was approved by the ethics committee (reference number Ghent: B670201421409). Written informed consent was obtained from all patients for the use of the pictures within the scoring sessions. The COSMIN checklist was used as guidance for designing and reporting our study.^{17–19}

Raters and participants

The raters were all vitiligo experts from different university clinics in Europe. Scoring of the clinical signs was done independently on a standardized form (Figure 1). Independence of rating was ensured by using separate computers and scoring sheets for each rater and by providing instructions for assessment (no interindividual discussion allowed).

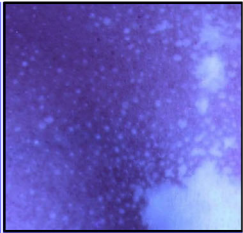
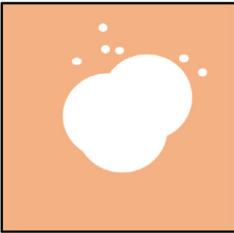
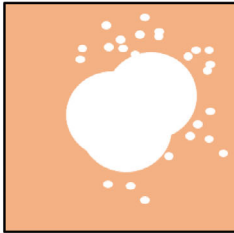
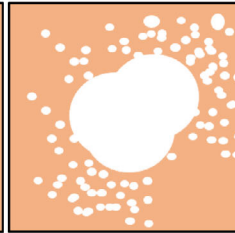

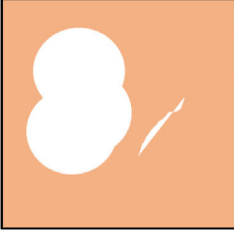
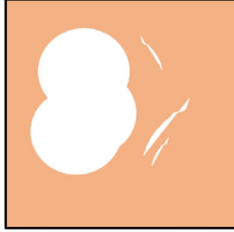
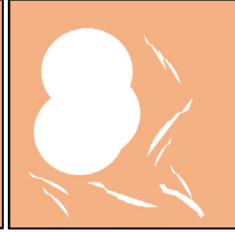

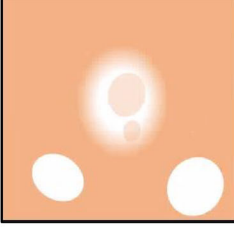
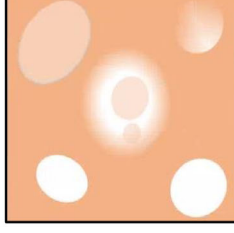
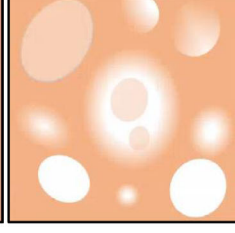
Vitiligo disease activity signs		Grade 1	Grade 2	Grade 3
C Confetti-like depigmentation				
K Koebner phenomenon				
H Hypochromic areas/borders				

Figure 2 Chart for grading each sign (grade 1–3) in the Vitiligo Signs of Activity Score (VSAS), translated from formulated definitions and photographic examples used within the context of this study. The grading reflects estimation of the intensity of each clinical sign within a specific area. For c-VSAS (confetti-like lesions) this is the estimated number of confetti-like depigmentations around a representative lesion (grade 1, < 10; grade 2, 10–50; grade 3, > 50). For k-VSAS (Koebner phenomenon) and h-VSAS (hypochromic areas/borders) this is the presence (estimated number of signs) per demarcated area: grade 1, 1; grade 2, 2–5; grade 3, > 5. The grades correspond overall to ‘somewhat present’ (grade 1), ‘clearly present’ (grade 2) or ‘very clearly present’ (grade 3). The clinical photographs represent an example of each sign. Based on the grading per area, in addition one ‘global grade’ (total body grade) per sign can be assigned, which can be considered as the grade that is most evident on average for a specific sign.

Patients were recruited at the Ghent University Hospital (Ghent, Belgium) and were selected at random from a group of voluntary clinically diagnosed patients with (nonsegmental) vitiligo, who were divided before randomization into different groups based on a global estimation of the extent of vitiligo (by N.v.G.). These patients were photographed in a standardized manner during routine clinical practice, and a set of generated ultraviolet system-based images was subsequently used for the scoring sessions. Overlap of this set exists with the set used in a previous study.²⁰

Reliability

The inter-rater reliability was assessed by comparing both the VSAS and the subscores per clinical sign (c-VSAS, k-VSAS, h-VSAS) and grading per item within each area (grade 1–3) between all raters. For the evaluation of the intrarater reliability, measurements were repeated on the same series of photographs, which were randomized in sequence using Research Randomizer (<https://www.randomizer.org>), with an interval of at least 2 weeks between rating sessions.

Validity

Face validity was assessed to evaluate the degree to which the items of the instrument are an adequate reflection of the construct to be measured, and was based on expert opinions and the literature.^{7,14,19}

Construct validity was evaluated based on testing against four hypotheses (formulated by two investigators) as no gold standard exists (Appendix S2; see Supporting Information). Sufficient evidence for construct validity was assumed if $\geq 75\%$ of the hypotheses were in accordance with the results. All hypotheses include a comparison (correlation or predictions) between the median VSAS of all raters and an expert Physician’s Global Assessment (PGA) score for clinical disease activity (performed by one expert, who was not involved in the scoring sessions). This PGA score included a five-point scale ranging from no disease activity to very severe disease activity.

Feasibility

The completion time was recorded by each rater to evaluate the feasibility of this scoring system. This scoring included

two steps: completion of the instrument (step 1) and calculation of the score (step 2). The total completion time was calculated based on the total timing of the two steps, or by the sum of the two steps separately.

Statistics and data analysis

Inter- and intrarater reliability were assessed using the two-way random (or two-way mixed for intrarater), absolute-agreement, single-measures intraclass correlation coefficient (ICC). The following guidelines for the interpretation of the ICC were used: < 0.4 was considered poor, $0.4-0.59$ fair, $0.6-0.74$ good and ≥ 0.75 excellent.²¹ For the correlation as included in the first hypothesis to test construct validity (Appendix S2; see Supporting Information), Spearman's rho correlation was used. Comparison of completion times (average ranking) between the first and second rounds was evaluated by Wilcoxon signed-rank tests. All analyses were performed using SPSS Statistics version 25 (IBM, Armonk, NY, USA).

Results

Raters and participants

The first scoring round was performed by four raters and the retest by three raters with an interval of approximately 2 months. In total, 247 ultraviolet pictures from 23 patients with vitiligo were included for the scoring. These patients had Fitzpatrick skin type II (two), III (19), IV (one) and V (one); 12 were female and 11 male. The total body surface area affected (assessed by VES) varied between the patients (range 0.13–42.7%, median 2.28%, mean 5.95%). The mean age at inclusion was 36 years (median 40, range 17–64) and the mean age at onset was 25.7 years (median 25, range 1.5–45; $n = 19$).¹³ At least one clinical sign was present in 17 of 23

cases (74%). Signs were considered 'present' if they were reported by at least three of the four raters. Presence of confetti-like depigmentations was recorded in 15 of 23 (grade 1 in seven, grade 2 in five and grade 3 in three), Koebner phenomenon in nine of 23 (grade 1 in seven and grade 2 in two) and hypochromic areas/borders in 15 of 23 (grade 1 in 14, grade 2 in one). Only two of 23 cases (9%) showed one isolated sign (both with hypochromic areas/borders), eight of 23 cases (35%) showed two signs (confetti-like depigmentations with Koebner phenomenon, and hypochromic areas/borders with confetti-like depigmentations) and seven of 23 cases (30%) showed all three clinical signs.

Reliability

Inter-rater reliability

The median overall VSAS (first round) was 4 [range 0–13, interquartile range (IQR) 1–8]. Figure 3 shows a scatter plot illustrating the distribution of the overall VSAS in mean rank order of all cases. The median scores (first round) for confetti-like depigmentations, Koebner phenomenon and hypopigmented areas/borders separately (c-VSAS, k-VSAS and h-VSAS) were 3 (range 0–12, IQR 0–7), 1 (range 0–8, IQR 0–8) and 1 (range 0–11, IQR 0–2), respectively. Table 1 lists the ICCs with 95% confidence intervals (CIs) of the overall score (VSAS) and each subscore (c-VSAS, k-VSAS and h-VSAS). The ICC inter-rater reliabilities for the overall score (VSAS) were excellent for both the first and second rounds (ICC 0.87 and 0.90, respectively). Similarly, the confetti-like depigmentation subscore (c-VSAS) also reached excellent ICC agreements. Inter-rater reliabilities for the Koebner phenomenon (k-VSAS) and hypochromic areas/borders (h-VSAS) subscores were fair (ICC 0.51 and 0.53) in the first round, while this improved to good agreement for Koebner phenomenon in the

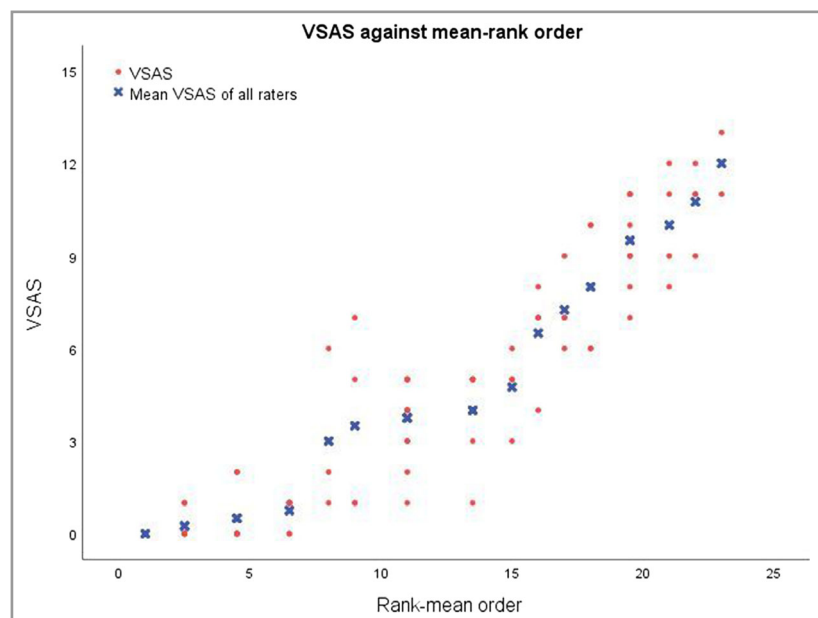


Figure 3 Vitiligo Signs of Activity Score (VSAS) scatter plot illustrating the distribution of the overall VSAS (y-axis) by four raters. The x-axis presents the rank order of the mean VSAS of all raters.

Table 1 Assessing inter- and intrarater reliability

Round	Number of raters	Score	Estimated ICC (95% CI)
Inter-rater reliability			
1	4	VSAS	0.87 (0.75–0.94)
		c-VSAS	0.83 (0.68–0.92)
		k-VSAS	0.51 (0.30–0.72)
		h-VSAS	0.53 (0.33–0.73)
2	3	VSAS	0.90 (0.80–0.95)
		c-VSAS	0.91 (0.82–0.96)
		k-VSAS	0.72 (0.52–0.86)
		h-VSAS	0.47 (0.21–0.70)
Intrarater reliability			
1 vs. 2 3	3	VSAS	Rater 1: 0.86 (0.70–0.94)
			Rater 2: 0.95 (0.88–0.98)
			Rater 3: 0.90 (0.72–0.96)
		c-VSAS	Rater 1: 0.89 (0.75–0.95)
			Rater 2: 0.95 (0.89–0.98)
			Rater 3: 0.89 (0.73–0.96)
		k-VSAS	Rater 1: 0.63 (0.29–0.82)
			Rater 2: 0.88 (0.75–0.95)
			Rater 3: 0.58 (0.22–0.80)
		h-VSAS	Rater 1: 0.60 (0.26–0.81)
			Rater 2: 0.89 (0.75–0.95)
			Rater 3: 0.73 (0.47–0.87)

The sample size was 23 patients for all assessments. CI, confidence interval; ICC, intraclass correlation coefficient; VSAS, overall Vitiligo Signs of Activity Score; c-VSAS, VSAS for confetti-like depigmentations; k-VSAS, VSAS for Koebner phenomenon type 2B; h-VSAS, VSAS for hypochromic areas/borders.

second round (ICC 0.72, 95% CI 0.52–0.86). Inter-rater agreements for grading the severity of each item within the involved areas (grade 1–3) were excellent for confetti-like lesions (ICC 0.83, 95% CI 0.71–0.92), and fair for Koebner phenomenon (ICC 0.56, 95% CI 0.35–0.76) and hypochromic areas/borders (ICC 0.51, 95% CI 0.31–0.71).

Intrarater reliability

Test–retest measurements showed excellent intrarater reliability (ICC \geq 0.86) for the overall VSAS (Table 1 and Figure 4).

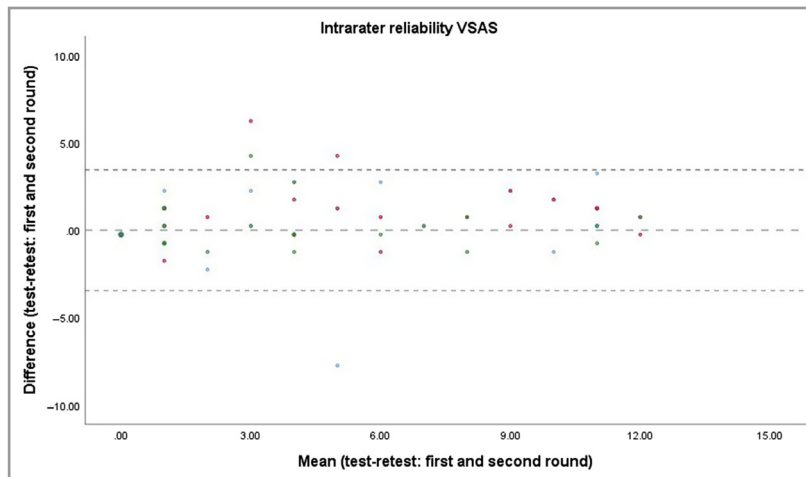


Figure 4 Bland–Altman plot illustrating the variation in Vitiligo Signs of Activity Score (VSAS) scores between the first and second scoring rounds (test–retest), which is a measure of the intrarater reliability. The intrarater intraclass correlation coefficient of the median of all raters was 0.90. The coloured dots represent the different raters.

Good median intrarater reliability was also found for the subscores confetti-like lesions (c-VSAS, ICC 0.89), Koebner phenomenon (k-VSAS, ICC 0.63) and hypochromic areas/borders (h-VSAS, ICC 0.73) (Table 1).

Validity

Face validity for the items included in the VSAS was supported by the opinion of vitiligo experts (Appendix S1; see Supporting Information) and a systematic review.^{7,14} Sufficient evidence for construct validity was provided, as three of the four hypotheses for construct validity were confirmed (Appendix S2; see Supporting Information). A strong positive monotonic correlation (median $r = 0.75$) was found between the median VSAS of the four raters per included case and the expert's PGA scores (Table 2 and Figure 5). The correlations (Spearman rho) between each rater separately and the expert's PGA scores were 0.69, 0.74, 0.68 and 0.68, respectively. For the second and fourth hypotheses the cut-off values of 50% were widely reached (78% and 75%, respectively; median of all raters). The third hypothesis was not reached, which can be explained by the low number of cases with a score of 0 (mean number of cases 5.5, median 6) and the fact that the expert rater was not informed about the predefined definitions for use of the instrument. A PGA score of 'not present' was selected only twice by the expert.

Feasibility

The median total completion time for the scoring and calculation of score in the first round was 138 s per patient (range 7–450). Significant improvement of the average ranking of completion times was observed in the second round for two of three raters (P-value from Wilcoxon signed-rank tests < 0.001). The median total completion time in the second round was 93 s per patient (range 10–323).

The median completion times of the instrument and score calculation separately in the first round (assessed by two raters) were 91 s per patient (range 3–383) for filling in the instrument (15 anatomical areas) and 18 s per patient (range 2–120) to calculate the scores.

Table 2 Spearman correlations between overall Vitiligo Signs of Activity Score and Physician's Global Assessment expert score in round 1

Rater	Spearman's rho	P-value
1	0.69	< 0.001
2	0.74	< 0.001
3	0.68	< 0.001
4	0.68	< 0.001

Discussion

In this study we evaluated the inter- and intrarater reliability, validity and feasibility of VSAS, a tool to quantify visible clinical signs linked to disease activity in a standardized way. This can be important for clinical trials (e.g. evaluating immunomodulating agents) and clinical practice (e.g. selecting specific treatments).^{1,2,8} The main strength of the VSAS is that a reproducible quantification of signs linked to disease activity in vitiligo can be proposed. The conversion into a numerical output might improve the monitoring of patients and communication between physicians. Results for reliability exceeded the cutoff point of acceptable inter- and intrarater reliability, and sufficient evidence for construct validity was provided. Based on the improved completion time in the second round we demonstrated that training seems to be useful.

Future stratification of the score is important to translate it into clinically meaningful categories (e.g. mild to very severe) and to define possible thresholds for the scores. The grading of each sign per involved area (grade 1–3) may be important to investigate in more detail before definitive implementation in the clinic or trials. This grading needs further validation especially related to its relevance. For now, it can be added as an option that will in particular be useful for studies investigating the prognostic significance of the separate signs in more depth. Within this context, combination with the

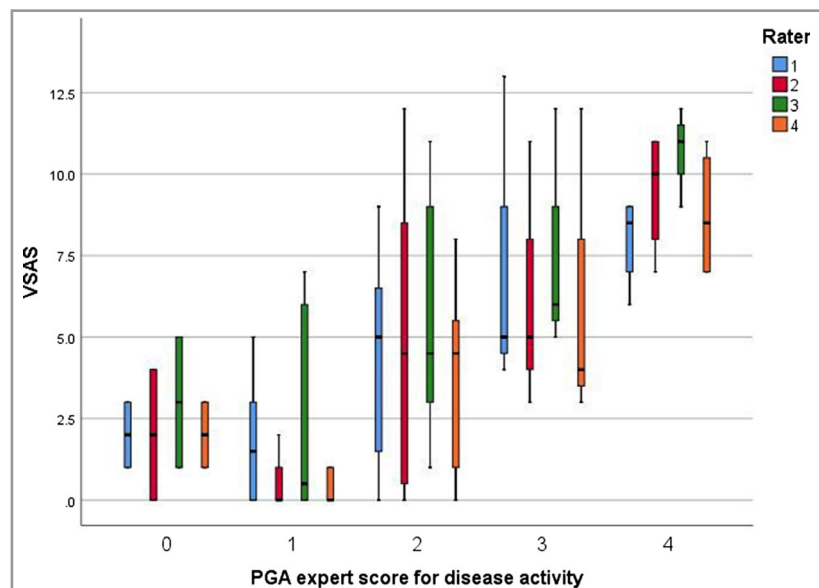
evaluation of disease progression over time might also be important (comparison of two timepoints, in general based on clinical photographs). This assessment can complete the scoring of disease activity. Future studies combining the quantification of disease activity signs with disease progression will therefore be required to provide more insight into the prognostic value of the obtained overall score, subscores and grading separately. To perform this type of study a standardized scoring system (such as the VSAS score) is required.

A strength of this measurement instrument is that it can be used during a single consultation for all patients (of all ages and with its segmental subtype). However, for the interpretation of the calculated score in segmental vitiligo we need to keep in mind the generally more limited number of locations that are involved compared with nonsegmental vitiligo. Therefore, interpretation of the scores may better be restricted to the respective subtypes.

Limitations of this study are the limited number of cases, from a single tertiary centre, and the limited variations of skin phototype (mainly II–III) of the included patient population. This limits possible generalization of the results. Photographs were taken during routine practice, which might be considered an advantage, as the quality of the pictures represents a real-life or routine situation. However, this also influenced the reduced availability of skin details (zoomed-in pictures), which might have negatively influenced the reliability. Live scoring may further improve the reliability (increase the visibility of the signs), although the feasibility aspects (e.g. complexity and timing) within this environment still need to be evaluated.

More studies within the international network are definitely required, with more investigators and raters, and on a larger cohort of patients of different origins to evaluate the comprehensibility of the scoring system and to reach a final consensus. Moreover, future studies are needed to evaluate the responsiveness of the VSAS and VSAS subscores, the exact value of each clinical sign separately in the prediction of the global disease

Figure 5 Box plots showing increasing Vitiligo Signs of Activity Score (VSAS) for higher Physician's Global Assessment (PGA) scores. Correlations between the VSAS scores and PGA expert score showed significant P-values ($P < 0.001$) for all raters, using Spearman's rho.



activity, and the relevance of VSAS from the patient perspective and/or its associations with patient-reported outcomes. To improve future implementation and accessibility a software application is currently under construction.

In conclusion, we demonstrated the reliability, validity and completion time of the VSAS. This measurement instrument enables a standardized registration of clinical signs linked to disease activity in vitiligo. Moreover, it can be used for future trials that aim to establish the clinical significance of the specific visible clinical signs in vitiligo in a more controlled setting. However, complementary studies will be required within different international settings to get more insight into the value of this instrument. These should measure content and construct validity, including evaluation of the clinical relevance of the suggested grading and evaluation of the responsiveness, and translation of the score into categories (e.g. mild to very severe).

Acknowledgments

We thank Dr Amit Pandya for critically reviewing and editing the content of the initial draft of this paper. We would like to express our gratitude to the volunteering patients for use of their pictures during the scoring rounds. We would like to thank C. Van Goethem for processing of the data.

References

- McKesey J, Pandya AG. A pilot study of 2% tofacitinib cream with narrow-band UVB for the treatment of facial vitiligo. *J Am Acad Dermatol* 2019; **81**:646–8.
- Joshiyura D, Alomran A, Zancanaro P *et al.* Treatment of vitiligo with the topical Janus kinase inhibitor ruxolitinib: a 32-week open-label extension study with optional narrow-band ultraviolet B. *J Am Acad Dermatol* 2018; **78**:1205–7.
- Cavalié M, Ezzedine K, Fontas E *et al.* Maintenance therapy of adult vitiligo with 0.1% tacrolimus ointment: a randomized, double blind, placebo-controlled study. *J Invest Dermatol* 2015; **135**:970–4.
- Eleftheriadou V, Hamzavi I, Pandya AG *et al.* International Initiative for Outcomes (INFO) for vitiligo: workshops with patients with vitiligo on repigmentation. *Br J Dermatol* 2019; **180**:574–9.
- Gan EY, Eleftheriadou V, Esmat S *et al.* Repigmentation in vitiligo: position paper of the Vitiligo Global Issues Consensus Conference. *Pigment Cell Melanoma Res* 2017; **30**:28–40.
- Eleftheriadou V, Thomas K, van Geel N *et al.* Developing core outcome set for vitiligo clinical trials: international e-Delphi consensus. *Pigment Cell Melanoma Res* 2015; **28**:363–9.
- van Geel N, Grine L, De Wispelaere P *et al.* Clinical visible signs of disease activity in vitiligo: a systematic review and meta-analysis. *J Eur Acad Dermatol Venereol* 2019; **33**:1667–75.
- Tovar-Garza A, Hinojosa JA, Hynan LS *et al.* Addition of oral minipulse dexamethasone to narrowband ultraviolet B phototherapy and topical steroids helps arrest disease activity in patients with vitiligo. *Br J Dermatol* 2019; **180**:193–4.
- Coias J, Hynan LS, Pandya AG. Lack of correlation of the patient-derived Vitiligo Disease Activity Index with the clinician-derived Vitiligo Area Scoring Index. *J Am Acad Dermatol* 2018; **78**:1015–16.
- Aboul-Fetouh N, Hinojosa J, Tovar-Garza A *et al.* The majority of patients presenting with vitiligo have a clinical sign of activity. *J Am Acad Dermatol* 2017; **77**:774–5.
- Goh BK, Pandya AG. Presentations, signs of activity, and differential diagnosis of vitiligo. *Dermatol Clin* 2017; **35**:135–44.
- Sosa JJ, Currimbhoy SD, Ukoha U *et al.* Confetti-like depigmentation: a potential sign of rapidly progressing vitiligo. *J Am Acad Dermatol* 2015; **73**:272–5.
- van Geel N, Lommerts J, Bekkenk M *et al.* Development and validation of the Vitiligo Extent Score (VES): an international collaborative initiative. *J Invest Dermatol* 2016; **136**:978–84.
- van Geel N, Boniface K, Seneschal J *et al.* Meeting report: Vitiligo Global Issues Consensus Conference workshop ‘Outcome measurement instruments’ and Vitiligo International Symposium, Rome, Nov 30–Dec 3rd. *Pigment Cell Melanoma Res* 2017; **30**:436–43.
- Ezzedine K, Lim HW, Suzuki T *et al.* Revised classification/nomenclature of vitiligo and related issues: the Vitiligo Global Issues Consensus Conference. *Pigment Cell Melanoma Res* 2012; **25**:E1–13.
- van Geel N, Speeckaert R, Taieb A *et al.* Koebner’s phenomenon in vitiligo: European position paper. *Pigment Cell Melanoma Res* 2011; **24**:564–73.
- Mokkink LB, Terwee CB, Knol DL *et al.* The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010; **10**:22.
- Terwee CB, Mokkink LB, Knol DL *et al.* Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012; **21**:651–7.
- Mokkink LB, Terwee CB, Patrick DL *et al.* The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010; **63**:737–45.
- van Geel N, Wolkerstorfer A, Ezzedine K, *et al.* Validation of a Physicians Global Assessment (PGA) tool for vitiligo extent: results of an international vitiligo expert meeting. *Pigment Cell Melanoma Res* 2019; **32**:728–33.
- Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994; **6**:284–90.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher’s website:

Appendix S1 Clinical relevance of items included in the Vitiligo Signs of Activity Score based on expert opinions during an international workshop.

Appendix S2 Hypotheses used for construct validity and results of the hypothesis testing.