The Patient and Observer Scar Assessment Scale: A Reliable and Feasible Tool for Scar Evaluation

Lieneke J. Draaijers, M.D., Fenike R. H. Tempelman, M.D., Yvonne A. M. Botman, M.D., Wim E. Tuinebreijer, M.D., Ph.D., Esther Middelkoop, Ph.D., Robert W. Kreis, M.D., Ph.D., and Paul P. M. van Zuijlen, M.D., Ph.D.

Beverwijk and Amsterdam, The Netherlands

At present, various scar assessment scales are available, but not one has been shown to be reliable, consistent, feasible, and valid at the same time. Furthermore, the existing scar assessment scales appear to attach little weight to the opinion of the patient. The newly developed Patient and Observer Scar Assessment Scale consists of two numeric scales: the Patient Scar Assessment Scale (patient scale) and the Observer Scar Assessment Scale (observer scale). The patient and observer scales have to be completed by the patient and the observer, respectively. The patient scale's consistency and the observer scale's consistency, reliability, and feasibility were tested. For the Vancouver Scar Scale, which is the most frequently used scar assessment scale at present, the same statistical measurements were examined and the results of the observer scale and the Vancouver scale were compared. The concurrent validity of the observer scale was tested with a correlation to the Vancouver scale. Furthermore, the authors examined which specific characteristics significantly influence the general opinion of the patient and the observers on the scar areas. Four independent observers have each used the observer scale and the Vancouver scale to assess 49 burn scar areas of 3×3 cm belonging to 20 different patients. Subsequently, the patients completed the patient scale for their scar areas. The (internal) consistency of both the patient and the observer scales was acceptable (Cronbach's alpha, 0.76 and 0.69, respectively), whereas the consistency of the Vancouver scale appeared not to be acceptable (alpha, 0.49). The reliability of the observer scale completed by a single observer was acceptable ($r = 0.73$). The reliability of the Vancouver scale completed by a single observer was lower ($r = 0.69$). The observer scale showed better agreement than the Vancouver scale because the coefficient of variation was lower (18 percent and 22 percent, respectively). The concurrent validity of the observer scale in relation to the Vancouver scale is high ($r = 0.89$, $p < 0.001$). Linear regression of the general opinions on scars of the observer and the patient showed that the observer's opinion is influenced by vascularization, thickness, pigmentation, and relief, whereas the patient's opinion is mainly influenced by itching and the thickness of the scar. Such an impact of itching and thickness of the scar on the patient's opinion is an important and novel finding. The Patient and Observer Scar Assessment Scale offers a suitable, reliable, and complete scar evaluation tool. (*Plast. Reconstr. Surg.* 113: 1960, 2004.)

A scar assessment scale, which subjectively evaluates the effectiveness of scar therapies, is an important evaluation tool because it describes the impression of experts on the appearance of scars. A scar assessment scale is considered suitable for the comparison of clinical results when it is tested as reliable, feasible, consistent, and valid. However, at present, no scar assessment scale is available that has proven to meet all of the aforementioned statistical requirements. A literature study shows that few of the currently available subjective scar assessment scales have been tested for reliability. It turned out that these scales could be used reliably when the scar was evaluated with more than three observers[1] or with more than nine observers.[2] Such scales are not feasible in a clinical setting because a scar assessment scale is only considered feasible when an assessment completed by a single observer has proven to be reliable. In addition, a scar assessment scale needs to be consistent to enable the

From the Burn Center and the Department of Surgery, Red Cross Hospital, and the Dutch Burn Foundation; the Department of Surgery and the Department of Dermatology, VU Medical Center; and the Department of Plastic, Reconstructive and Hand Surgery, Academic Medical Center. Received for publication April 14, 2003; revised June 30, 2003.

DOI: 10.1097/01.PRS.0000122207.28773.56

user to arrive at a total score by adding the individual scores of the separate parameters.

We feel that a subjective scar assessment scale will become even more suitable for clinical studies if the opinion of the patient is incorporated into such a scale. Some studies consider pain and itching of the scar as scar assessment features[3–5] but do not evaluate the opinion of the patient on the appearance of the scar. In an attempt to make a scar scale that also attaches weight to the opinion of the patient, we developed the Patient and Observer Scar Assessment Scale (Fig. 1). The scale consists of two numeric scales: the Patient Scar Assessment Scale (patient scale), which is to be completed by the patient, and the Observer Scar Assessment Scale (observer scale), which needs to be filled out by the observer. We conducted a study to evaluate the consistency of the patient scale and to evaluate the observer scale's consistency, reliability, and feasibility. For the Vancouver Scar Scale, which is the most frequently used scar assessment scale in clinical studies,[6–10] the same statistical measurements were established to compare the results of the observer and Vancouver scales. The concurrent validity of the observer scale was tested in relation to the Vancouver scale. Furthermore, we examined which specific characteristics significantly influence the general opinion of the patient and the observers on the scar areas.

First, the internal consistency of the patient, observer, and Vancouver scales was tested by means of Cronbach's alpha. Second, the interobserver reliability of the observer and Vancouver scales was tested by means of the intraclass correlation coefficient. An intraclass correlation coefficient of 1 indicates that measurements carried out by different numbers of observers

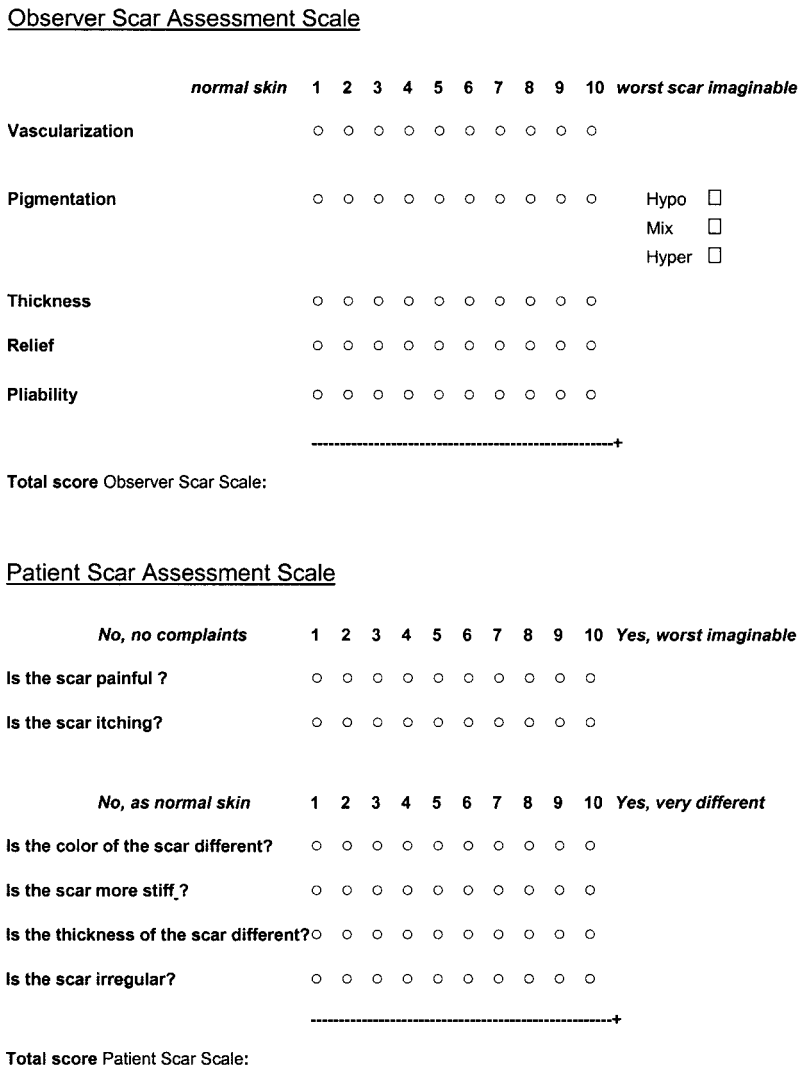

FIG. 1. The Patient and Observer Scar Assessment Scale.

produce the same results. The interobserver reliability was tested when four observers assessed the scar areas with the observer and Vancouver scales. Subsequently, the interobserver reliability was tested for each possible combination of two and three observers to study the minimal number of observers required for a reliable scar analysis. The intraclass correlation coefficient for a single observer was also examined. Third, the agreement between measurements of both the observer scale and the Vancouver scale was examined with the coefficient of variation. Fourth, the concurrent validity of the observer and Vancouver scales was established by means of a Spearman rank correlation coefficient between the two scales. Finally, the overall opinion of the patient and of the observer on the appearance of the scars was analyzed by linear regression to determine the functionality of the patient scale.

Patients and Methods

Patients

Each of the four independent observers has used the observer scale (Fig. 1) and the Vancouver scale (Table I) to assess 49 scar areas of 3 × 3 cm belonging to 20 different patients. Subsequently, each patient completed the patient scale (Fig. 1) for his or her scar areas. The age of the patients ranged from 15 to 73 years. The scars were assessed at an average of 43.7 months after the burn injury (range, 3 to 360 months; median, 13 months). All scar areas were assessed by the same four observers during the study. All observers were physicians, three of whom were regularly working with burn patients.

The POSAS

The Patient and Observer Scar Assessment Scale (Fig. 1) consists of two scales, the patient scale, which contains six items, and the observer scale, which contains five items. All items of the two scales are scored numerically. The patient scores the characteristics scar color, pliability, thickness, relief, itching, and pain, whereas the observer scores scar vascularization, pigmentation, pliability, thickness, and relief. These characteristics were selected on the basis of our clinical experience and a literature study. In our view, the pigmentation should be scored numerically. Furthermore, pigmentation was documented as hypopigmentation, hyperpigmentation, and mixed pigmentation in accordance with the documentation used in the Vancouver scale (Fig. 1). The observer has to score vascularization and pigmentation. The patient has to score color in general because it was expected that it would be too difficult for patients to make a distinction between vascularization and pigmentation. Each item has a 10-step score, whereby the score 10 reflects the worst imaginable scar or sensation. The total score of the observer scale consists of adding the scores of each of the five items (range, 5 to 50). The total score of the patient scale consists of adding the scores of each of the six items (range, 6 to 60). The lowest scores, 5 and 6, respectively, reflect normal skin. In addition to the scar assessment, the observers and the patients gave a general opinion on the appearance of the scar areas (score, 1 to 10, in which a score of 10 corresponds to the worst possible scar appearance).

TABLE I
The Vancouver Scar Scale

1. Vascularity		2. Pigmentation	
Normal	0	Normal	0
Pink	1	Hypopigmentation	1
Red	2	Mixed	2
Purple	3	Hyperpigmentation	3
3. Pliability		4. Height	
Normal	0	Flat	0
Supple	1	<2 mm	1
Yielding	2	2–5 mm	2
Firm	3	>5 mm	3
Ropes	4		
Contracture	5		

The VSS

The Vancouver scale (Table 1) consists of four variables: vascularity, height (thickness), pliability, and pigmentation. Each variable has four to six possible scores. A total score ranges from 0 to 14, whereby a score of 0 reflects normal skin.

Statistical Analysis

The data were analyzed by the statistical program SPSS for Windows 10.0 (SPSS, Inc., Chicago, Ill.). Cronbach's alpha was used to calculate the internal consistency of the patient, observer, and Vancouver scales. The average scores of all observers per item of the observer scale and of the Vancouver scale were used for calculations. Cronbach's alpha should be from 0.70 up to and including 0.90[11] to demonstrate consistency in a scale.

The reliability of the total scores of the observer and Vancouver scales was calculated by means of

the intraclass correlation coefficient.[11–13] The two-way random-effect model and the consistency type were selected for the intraclass correlation coefficient in SPSS. The intraclass correlation coefficient, with its 95% confidence interval, was calculated for the group of four observers and for one observer. Separate intraclass correlation coefficient calculations were required for each possible combination of two and three observers. Therefore, only a range of the lowest value and highest value of the intraclass correlation coefficient can be given and not a single value of the intraclass correlation coefficient. An intraclass correlation coefficient value of 0.70 was considered a minimum requirement for reliable results.

The agreement between measurements, which is related to the reliability, carried out by four observers was based on the standard error of measurement (standard error of measurement = $\sqrt{}$mean square residual). The standard error of measurement calculated the amount of error in scores and was expressed in coefficients of variation (coefficient of variation = standard error of measurement/mean $\times$ 100). Coefficients of variation were calculated to make it possible to compare the amount of error between scores of the observer scale and the Vancouver scale. The range, from the lowest coefficient of variation to the highest coefficient of variation, is produced for each possible combination of two and three observers.

The Spearman rho correlation coefficient with a two-tailed test of significance was used to establish the concurrent validity between the average total score of the observer scale and the average total score of the Vancouver scale. The significance criterion was set at 0.05.

The influence of the various scar characteristics on the overall opinion of the patient and the observer was estimated by means of linear regression. The "backward" method was used to enter the variables into the analysis. The significance criterion was set at 0.05.

RESULTS

Consistency of the Patient, Observer, and Vancouver Scales

The (internal) consistency of the patient and observer scales appeared to be acceptable (Cronbach's alpha, 0.76 and 0.69, respectively), whereas the consistency of the Vancouver scale appeared not to be acceptable (Cronbach's alpha, 0.49).

Reliability of the Observer and Vancouver Scales

The intraclass correlation coefficient values for the observer and Vancouver scales in combination with the number of observers are presented in Table II. In general, the intraclass correlation coefficient values of the observer scale were higher than those of the Vancouver scale. For both the observer and Vancouver scales, high intraclass correlation coefficient values ($r \geq 0.90$) were found when four observers assessed the scar area at the same time. The intraclass correlation coefficient of the observer scale completed by a single observer turned out to be higher ($r \geq 0.73$) than the same intraclass correlation coefficient value for the Vancouver scale ($r \geq 0.69$).

Agreement of the Observer and Vancouver Scales

The coefficients of variation for the measurements carried out by each combination of two and three observers and for four observers of the observer scale and the Vancouver scale are presented in Table III. In general, the coefficient of variation of the observer scale is lower than the coefficient of variation of the Vancouver scale. This means that the observer scale shows less variability between repeated measurements than the Vancouver scale.

Concurrent Validity of the Observer and Vancouver Scales

A significant correlation was calculated between the observer scale and the Vancouver scale (Spearman's rho = 0.89, $p < 0.001$).

Linear Regression

The linear regression of the opinion of the observer turned out to be significantly influenced by relief, thickness, pigmentation, and

TABLE II
Intraclass Correlation Coefficient Calculated for the Observer and Vancouver Scales in Combination with the Number of Observers

No. of Observers	ICC*	Observer Scale	Vancouver Scale
4	Mean (95% CI)	0.92 (0.87–0.95)	0.90 (0.84–0.94)
3	Range	0.88–0.90	0.86–0.88
2	Range	0.81–0.89	0.78–0.84
1	Mean (95% CI)	0.73 (0.62–0.82)	0.69 (0.57–0.79)

ICC, intraclass correlation coefficient; CI, confidence interval.
* The ICC together with the 95% CI is given for the scores of the observer scale and the Vancouver Scar Scale completed by a single observer and completed by four observers. For two and three observers, separate ICC calculations were required for each possible observer combination. The minimum and the maximum ICC of these combinations are given.

TABLE III

Coefficient of Variation of the Observer and Vancouver Scar Scales in Combination with the Number of Observers

No. of Observers		Observer Scale	Vancouver Scale
4	CV* (SE_{meas})	18% (3.14)	22% (1.34)
3	CV (SE_{meas} range)	16–18% (2.80–3.19)	21–23% (1.27–1.41)
2	CV (SE_{meas} range)	16–21% (2.75–3.65)	19–25% (1.18–1.54)

CV, coefficient of variation; SE_{meas}, standard error of measurement.

* Coefficient of variation is calculated as the (SE_{meas}/mean) × 100. The CV is based on the standard error of measurement, which is calculated as the square root of the mean square of the residuals of the analysis of variance.

color. Pliability was the only item that nonsignificantly influenced the opinion of the observer (Table IV). The opinion of the patient on their scar turned out to be only significantly influenced by itching and thickness of the scar (Table IV).

Discussion

A scar assessment scale can be tested on its suitability for the evaluation of the results of a clinical study. A scar assessment scale is suitable for such evaluation if it turns out to be consistent, reliable, feasible, and valid. We conducted a study to examine to what extent the newly developed Patient and Observer Scar Assessment Scale and the Vancouver Scar Scale, which is the most frequently used scar assessment scale at present, meet the aforementioned statistical requirements.

The significant correlation (concurrent validity) between the observer and Vancouver scales suggests that both scales have a comparable tendency in the overall scores. However, the other statistical requirements show several differences between them.

The consistency of the patient and observer scale was acceptable, whereas the Vancouver scale appeared not to have an acceptable consistency. However, the consistency of the patient scale could have been influenced because patients rated more than one scar on their body. As a consequence, the ratings may not have been completely independent of each other. The lack of consistency of the Vancouver scale might be explained by the fact that it includes a variable that appears to be a nominal variable (pigmentation). In contrast, each item of the Patient and Observer Scar Assessment Scale is scored on a proper numeric scale; consequently, the consistency turns out to be significantly higher. The consistency of the Patient and Observer Scar Assessment Scale indicates that the individual scores (of each scale) can be reliably summed to reach a total score.

TABLE IV

Linear Regression of the Overall Opinion of the Observer (a) and the Patient (b) with Significance Levels*

a	b	*p*
Color	0.49	<0.001
Height	0.48	<0.001
Pigmentation	0.30	<0.001
Relief	0.25	<0.001
Dependent variable: opinion of the observer		
b	b	*p*
Itching	0.44	0.001
Height	0.32	0.011
Dependent variable: opinion of the patient		

* The backward method was used and the significance criteria were set at 0.05.

The reliability of the observer scale completed by a single observer turned out to be higher than the reliability of the Vancouver scale completed by a single observer. The reliability of the observer scale was acceptable because the intraclass correlation coefficient of the observer scale assessed by a single observer was higher than the required value of 0.70. Obviously, it was not possible to measure the reliability of the patient scale by testing the reliability of measurements of different patients. The reliability of the patient scale can be measured, however, with a test-retest reliability. This will be carried out in a different study. The observer scale also has a better agreement between repeated measurements than the Vancouver scale.

The feasibility of a scar assessment scale is determined by the clinical experience of working with the scale rather than by statistical requirements. In our opinion, the observer scale is a feasible evaluation tool because a single observer can use the scale reliably. This is very practical in a clinical setting. Furthermore, this study showed us that the observers and patients could complete the Patient and Observer Scar Assessment Scale without any problem, which also contributes to the feasibility.

The linear regression showed that each of the scar characteristics had a different influence on the general opinion of the observer. Remarkably, pliability did not have a significant influence on the general opinion of the observer. All other parameters (vascularization, thickness, pigmentation, and relief), which were examined visually, did have a significant influence on the general opinion. It could be the case that the observer gives a general opinion on a scar without checking the pliability of the scar first with his or her fingers. As a result, the pliability might be underestimated in the general opinion. The general opinion of the patient is predominantly influenced by itching and by the thickness of the scar. The impact of itching on the patient's opinion underlines the importance of research on the pathophysiology of itching and nerve regeneration. Scar thickness is an important visual feature of the scar and, not surprisingly, has a significant influence on the patient's opinion. The impact of itching and thickness of the scar is an important and novel finding. Linear regression did not show an influence of pain on the general opinion of the patient, which may be caused by a low incidence of pain. Although no significant relation was shown among color, pliability, relief, and the general opinion of the patient, we feel that a conclusion that these parameters need to be eliminated from the Patient and Observer Scar Assessment Scale would be premature. It should be noted that the weight of the different parameters on the general opinion of the observer and the patient might vary among different populations (age, race, and sex) and regarding other aspects of the scar (cause, size, and location). No parameter was eliminated from the Patient and Observer Scar Assessment Scale, and each parameter has the same weight with respect to the total score. It is possible to apply a weighted score, where parameters such as itching have more weight in arriving at the total score. The weight of each parameter may be established by linear regression.

The Patient and Observer Scar Assessment Scale was developed for the evaluation of all scar types (e.g., linear scars, burn scars). However, in this study, it was only shown that the Patient and Observer Scar Assessment Scale is suitable for the rating of burn scars. Additional research on different categories of scar tissue, such as linear scars, is mandatory. We conclude that the Patient and Observer Scar Assessment Scale offers a suitable, reliable, and complete scar evaluation tool.

Lieneke J. Draaijers, M.D.
Burn Center
Red Cross Hospital
Vondellaan 13
1942 LE Beverwijk, The Netherlands
lienusd@hotmail.com

ACKNOWLEDGMENTS

This work was supported by the Dutch Burn Foundation. The authors thank Albert Jan Zonnevylle for valuable and insightful comments on this article.

REFERENCES

1. Smith, G. M., Tompkins, D. M., Bigelow, M. E., and Antoon, A. Y. Burn-induced cosmetic disfigurement: Can it be measured reliably? *J. Burn Care Rehabil.* 9: 371, 1988.
2. Beausang, E., Floyd, H., Dunn, K. W., Orton, C. I., and Ferguson, M. W. A new quantitative scale for clinical scar assessment. *Plast. Reconstr. Surg.* 102: 1954, 1998.
3. Gold, M. H. A controlled clinical trial of topical silicone gel sheeting in the treatment of hypertrophic scars and keloids. *J. Am. Acad. Dermatol.* 30: 506, 1994.
4. Nedelec, B., Shankowsky, H. A., and Tredget, E. E. Rating the resolving hypertrophic scar: Comparison of the Vancouver Scar Scale and scar volume. *J. Burn Care Rehabil.* 21: 205, 2000.
5. Ahn, S. T., Monafo, W. W., and Mustoe, T. A. Topical silicone gel for the prevention and treatment of hypertrophic scar. *Arch. Surg.* 126: 499, 1991.
6. Cheng, W., Saing, H., Zhou, H., Han, Y., Peh, W., and Tam, P. K. Ultrasound assessment of scald scars in Asian children receiving pressure garment therapy. *J. Pediatr. Surg.* 36: 466, 2001.
7. Moiemen, N. S., Staiano, J. J., Ojeh, N. O., Thway, Y., and Frame, J. D. Reconstructive surgery with a dermal regeneration template: Clinical and histologic study. *Plast. Reconstr. Surg.* 108: 93, 2001.
8. Musgrave, M. A., Umraw, N., Fish, J. S., Gomez, M., and Cartotto, R. C. The effect of silicone gel sheets on perfusion of hypertrophic burn scars. *J. Burn Care Rehabil.* 23: 208, 2002.
9. Van Zuijlen, P. P. M., Vloemans, J. F., van Trier, A. J., et al. Dermal substitution in acute burns and reconstructive surgery: A subjective and objective long-term follow-up. *Plast. Reconstr. Surg.* 108: 1938, 2001.
10. Vloemans, A. F., Soesman, A. M., Kreis, R. W., and Middelkoop, E. A newly developed hydrofibre dressing in the treatment of partial-thickness burns. *Burns* 27: 167, 2001.
11. Nunnaly, J. C., Jr. *Psychometric Theory*, 2nd Ed. New York: McGraw-Hill, 1978.
12. Streiner, D. L., and Norman, G. R. *Health Measurement Scales: A Practical Guide to Their Development and Use*, 2nd Ed. New York: Oxford University Press, 1995. Pp. 104-127.
13. Shrout, P. E., and Fleiss, J. L. Intraclass correlation: Uses in assessing rater reliability. *Psychol. Bull.* 86: 420, 1979.