Reliability of the Patient and Observer Scar Assessment Scale and a 4-point scale in evaluating linear facial surgical scars

ORIGINAL ARTICLE

Reliability of the Patient and Observer Scar Assessment Scale and a 4-point scale in evaluating linear facial surgical scars

X. Liu,[1,2,*] P.J. Nelemans,[3] M. Van Winden,[4,5] N.W.J. Kelleners-Smeets,[1,2] K. Mosterd[1,2]

[1]Department of Dermatology, Maastricht University Medical Centre, Maastricht, The Netherlands
[2]GROW, School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, The Netherlands
[3]Department of Epidemiology, Maastricht University, Maastricht, The Netherlands
[4]Department of Surgery, St. Elisabeth Hospital Tilburg, Tilburg, The Netherlands
[5]Former student Maastricht University, The Netherlands
*Correspondence: X. Liu. E-mail: xmxmliu@gmail.com

Abstract

Background In order to evaluate and improve aesthetic outcome for patients undergoing dermatological surgery, a reliable evaluation tool must be used. The 4-point scale and the Patient and Observer Scar Assessment Scale (POSAS) were both developed for this purpose.

Objective To compare the reliability of the POSAS scale with the 4-point scale for facial linear surgical scars and to assess the influence of different scar characteristics on the overall impression.

Methods Patients visiting the outpatient clinics of the Maastricht University Medical Centre with linear facial scars were included. The 4-point scale and the Observer Scar Assessment Scale (OSAS) were completed by three independent observers. The Patient Scar Assessment Scale (PSAS) was completed by the patient on the day of the visit and 2 weeks later. The intra-class correlation coefficient (ICC) with corresponding 95% confidence intervals (CI) were calculated to assess the reliability of the scales. Linear multivariate regression analyses were performed to evaluate how the score on each aspect affected the overall opinion.

Results Fifty scars in 50 patients were included. The ICC of the 4-point scale was 0.819 (95% CI: 0.708–0.892) for multiple observers and 0.602 (95% CI: 0.447–0.734) for single observer. These were superior to the ICCs of total OSAS scores (0.783 95% CI: 0.547–0.888 for multiple observers and 0.546 95% CI: 0.287–0.726 for single observer). ICCs of individual sub-items on the OSAS scored even lower. The sub-items contributed differently to the overall opinion among the different observers or patient.

Conclusion In terms of reliability, the overall opinion and the 4-point scale were superior to the total POSAS score or the sub-items. Observers do not weight individual scar characteristics equally to arrive at an overall opinion, which challenges the assumption that calculating a total POSAS score by summing of the scores on the individual items is a valid approach.

Received: 5 January 2016; Accepted: 11 April 2016

Conflicts of interest

All authors declare to have no conflict of interests.

Funding sources

Maastricht University Medical Centre, the Netherlands. No external funding sources were obtained for this work.

Introduction

The incidence of skin cancer has been on the rise over the past decades.[1–3] As surgery is the most frequently used treatment and skin cancer most often arises in the face, patients end up with permanent scars that may lead to functional and psychological distress.[4] Therefore, ensuring optimal cosmetic outcome after treatment has become a major concern.

In order to evaluate and improve aesthetic outcome for patients undergoing surgery, a reliable evaluation tool must be used. In daily practice, the 4-point scale is frequently applied: assessing the appearance of scars as excellent, good, fair or poor.[5] This is a quick and easy tool that has proved to be excellently reproducible in non-surgical scars in the past.[6] However, its reliability in surgical scars following dermatologic surgery has never

been studied. In addition, the 4-point scale does not provide information on the factors distinctive for the cosmetic appearance of the scar. This information might be valuable to improve the aesthetic outcome for patients. In a research setting, the Patient and Observer Scar Assessment Scale (POSAS) is considered the most comprehensive and widely used scale for cosmetic evaluation.[7,8] This scale allows both patients and physicians to give sub-scores on characteristics that are assumed to be important in the assessment of scars. Another major advantage of this assessment tool, above the others available, is that it incorporates both the patient and the physician's perspective of the scar. When the POSAS was firstly developed, it was meant to evaluate burn scars.[9] Later on, it was validated for linear scars after breast surgery.[10] In 2005, one item was added to the original scale and more sub-classifications were added contributing to the current POSAS 2.0.[11] This version of the POSAS was tested in 100 different linear scars all over the body, but only 10 were located on the face.

As the aesthetic results of surgery for skin cancer is especially relevant in the face, we compared the reliability of both scales for facial linear scars in this study.[6] Secondly, we studied the influence of different scar characteristics on the general impression of the cosmetic outcome as reported by the patient and the physician.

Material and methods

Patients visiting the outpatient dermatology clinics of the Maastricht University Medical Centre in the Netherlands were included if they had received surgical dermatological treatment in the face at least 1 month prior to the visit. Only linear scars were included. Patients with insufficient understanding of the Dutch language were excluded. Approval of the Ethics committee was obtained prior to the start of the study and all patients had given informed consent before participation.

Both, the 4-point and POSAS scale were used for assessment of cosmetic appearance of scars and repeated assessments were collected. At first, the patient rated his or her scar according to the Dutch Patient Scar Assessment Scale (PSAS) version 2.0 (Fig. 1) using a 1–10 scale on six aspects: colour, pliability, thickness, relief, itching and pain. This resulted in a total score of 6–60 with a higher score indicating a worse scar. In addition, the patient gave an overall opinion of the scar on a scale of 1–10 with 10 indicating the worst scar imaginable. Patients were asked to assess their own scar during the visit and again 2 weeks later. The second assessment was returned by mail. The interval of 2 weeks was chosen as it was long enough for individuals to not recall their scores and short enough for changes in the scar to not have occurred.[12]

Secondly, scars were assessed by three independent observers consisting of a dermatologist (DE), a dermatology resident (RE) and an intern (IN). The different observers assessed the scar on the same day and were blinded to each other's observations. The 4-point scale was rated before completion of the Dutch Observer Scar Assessment Scale (OSAS) as the examination of the individual scar characteristics might influence the overall opinion.

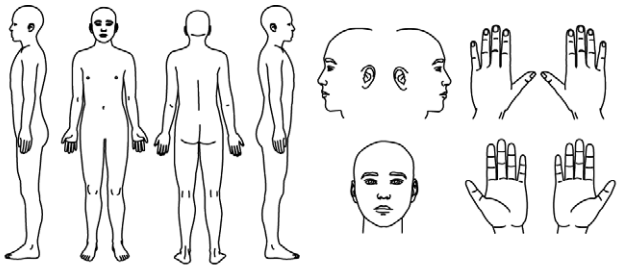
The 4-point scale classifies scars as 'excellent' (no obvious scarring, atrophy or induration and slight or no redness or change in pigmentation compared with adjacent skin), 'good' (no obvious scarring, atrophy or induration and/or moderate redness or increase in pigmentation compared with adjacent skin), 'fair' (slight to moderate occurrence of scarring, atrophy, induration and/or significant redness or increase in pigmentation compared with adjacent skin) or 'poor' (extensive occurrence of scarring, atrophy, induration and/or redness or increase in pigmentation compared with adjacent skin). Assessments with the OSAS were based on vascularity, pigmentation, pliability, thickness, relief and surface area. A Plexiglas was pressed against the scar and adjacent skin to distinguish between redness and pigmentation.[13]Each variable was scored 1–10 with 1 resembling normal skin and 10 the worst scar imaginable. Sub-classification was possible for each item, for example, to distinguish hyper- from hypo-pigmentation. The total OSAS score was the sum of all six items. In addition, an overall opinion was given on a scale of 1–10. Finally, the observer could add remarks on aspects of the scar which he or she felt not adequately addressed by the OSAS.

Statistical analysis

Reliability was quantified using the intra-class correlation coefficient (ICC) with corresponding 95% confidence intervals for both the 4-point scale and the OSAS. This reliability parameter relates the variability from repeated assessments (measurement error) to the variability between patients and expresses how well patients can be distinguished from each other despite measurement error. If the variability between patients is high compared to variability between assessments, the ICC approaches 1. If the variability between assessments is much higher than the variability between patients, the ICC approaches 0. The two-way random effect model with absolute agreement was used to calculate the ICC and assessed the reliability of single scores by one observer (single measurement) and of average scores from three observers (average measurement). The sample size calculation was aimed at obtaining a 95% confidence interval with acceptable width (± 0.1) around the estimated ICC. Based on the literature, an ICC of at least 0.8 was expected. With two measurements per patient (test-retest) the required sample size was 50 and with three measurements (three observers) the required sample size was 35.[14]

Linear multivariate regression analyses were performed to evaluate the impact of each aspect on the overall opinion. The dependent variable was the overall opinion as rated by the observer or patient, and the independent variables were the scores of the individual scar characteristics. The regression coefficient with the corresponding 95% CI was calculated. A P-value ≤ 0.05

POSAS Observer scale



	1 = no, not at all … yes, very much = 10									
	1	2	3	4	5	6	7	8	9	10
Has the scar been painful the past few weeks?	○	○	○	○	○	○	○	○	○	○
Has the scar been itching the past few weeks?	○	○	○	○	○	○	○	○	○	○

	1 = no, as normal skin … yes, very different = 10									
	1	2	3	4	5	6	7	8	9	10
Is the scar color different from the color of your normal skin at present?	○	○	○	○	○	○	○	○	○	○
Is the stiffness of the scar different from your normal skin at present?	○	○	○	○	○	○	○	○	○	○
Is the thickness of the scar different from your normal skin at present?	○	○	○	○	○	○	○	○	○	○
Is the scar more irregular than your normal skin at present?	○	○	○	○	○	○	○	○	○	○

	1 = as normal skin … very different = 10									
	1	2	3	4	5	6	7	8	9	10
What is your overall opinion of the scar as compared to normal skin?	○	○	○	○	○	○	○	○	○	○

	1 = normal skin … worst scar imaginable = 10									
Parameter	1	2	3	4	5	6	7	8	9	10
Vascularity	○	○	○	○	○	○	○	○	○	○
Pigmentation	○	○	○	○	○	○	○	○	○	○
Thickness	○	○	○	○	○	○	○	○	○	○
Relief	○	○	○	○	○	○	○	○	○	○
Pliability	○	○	○	○	○	○	○	○	○	○
Surface area	○	○	○	○	○	○	○	○	○	○
Overall opinion	○	○	○	○	○	○	○	○	○	○

Category
pale | pink | red | purple | mix
hypo | hyper | mix
thicker | thinner
more | less | mix
supple | stiff | mix
expansion | contraction | mix

Explanation

The observer scale of the POSAS consists of six items (vascularity, pigmentation, thickness, relief, pliability and, surface area).
All items are scored on a scale ranging from 1 ("like normal skin") to 10 ("worst scar imaginable").
The sum of the six items results in a total score of the POSAS observer scale. Categories boxes are added for each item. Furthermore, an overall opinion is scored on a scale ranging from 1 to 10.
All parameters should preferably be compared to normal skin on a comparable anatomic location.

Explanatory note on the items:

- **VASCULARITY** Presence of vessels in scar tissue assessed by the amount of redness, tested by the amount of blood return after blanching with a piece of Pexiglas
- **PIGMENTATION** Brownish coloration of the scar by pigment (melanin): apply Plexiglas to the skin with moderate pressure to eliminate the effect of vascularity
- **THICKNESS** Average distance between the subcutical-dermal border and the epidermal surface of the scar
- **RELIEF** The extent to which surface irregularities are present (preferably compared with adjacent normal skin)
- **PLIABILITY** Suppleness of the scar tested by wrinkling the scar between the thumb and index finger
- **SURFACE AREA** Surface area of the scar in relation to the original wound area

Figure 1 The Patient and Observer Scar Assessment Scale.

Table 1 Ratings on the 4-point scale for all observers

	DE	RE	IN
Excellent	12 (24%)	8 (16%)	8 (16%)
Good	19 (38%)	14 (28%)	24 (48%)
Average	18 (36%)	26 (52%)	16 (32%)
Bad	1 (2%)	2 (4%)	1 (2%)

DE, dermatologist; RE, resident; IN, intern.

Table 2 ICCs of single and multiple assessments regarding the PSAS

	N	Single assessment (95% CI)	Multiple assessments (95% CI)
Pain	47	0.411 (0.140–0.624)	0.583 (0.246–0.769)
Itch	48	0.868 (0.777–0.924)	0.930 (0.875–0.960)
Colour	47	0.412 (0.153–0.621)	0.584 (0.266–0.766)
Stiffness	48	0.786 (0.646–0.875)	0.880 (0.785–0.933)
Thickness	47	0.500 (0.242–0.690)	0.667 (0.390–0.817)
Surface	48	0.555 (0.323–0.724)	0.714 (0.488–0.840)
Total PSAS	45	0.736 (0.524–0.854)	0.848 (0.688–0.921)
Overall opinion	37	0.796 (0.640–0.889)	0.886 (0.780–0.941)

ICC, intra-class correlation coefficient; PSAS, Patient Scar Assessment Scale; CI, confidence interval.

was considered to be statistically significant. All data were analysed using SPSS, version 20.0.

Results

Fifty scars were assessed in 50 patients (26 males and 24 females). The mean age of the patient was 66 years (range 34–91 years). The average age of the scar was 55 months (range 1–216 months). After 2 weeks, 48 patients returned their second PSAS, however, some questions remained unanswered.

Overall, patients gave their scar a mean score of 3.8 (±2.5) on the first assessment and 3.6 (±2.2) 2 weeks later. The mean total PSAS was 13.0 (±8.2) on the first assessment and 15.5 (±9.8) on the second assessment. The mean overall opinion of the scar as rated by the observers in the OSAS was 3.7 (±2.0, DE), 4.2 (±1.9, RE) and 3.4 (±1.6, IN). The mean total OSAS score was 11.6 (±4.7, DE), 14.9 (±5.1, RE) and 15.7 (±5.2, IN) for the three observers. The ratings on the 4-point scale by the observers are shown in Table 1. The most common remark added was the presence of 'railroad track' suture marks which clearly influenced the appearance of the scar.

The ICCs of the PSAS and its individual items for single and two measurements (test and re-test after 2 weeks) are shown in Table 2. For the patient, the itchiness and stiffness had a higher ICC than other items of the scale.

Table 3 summarizes all ICCs for single and multiple observers. The 4-point scale and the overall opinion scores of the OSAS showed a higher ICC than the scores on individual items or the sum of those (total OSAS score).

Table 3 ICCs of single and multiple observers regarding the 4-point scale and the OSAS

	N	Single observer (95% CI)	Multiple observers (95% CI)
Vascularity	50	0.377 (0.205–0.549)	0.645 (0.437–0.785)
Pigmentation	50	0.380 (0.204–0.554)	0.648 (0.434–0.789)
Thickness	49	0.410 (0.220–0.588)	0.676 (0.459–0.810)
Relief	50	0.144 (0.002–0.315)	0.335 (0.006–0.580)
Pliability	50	0.396 (0.215–0.571)	0.663 (0.451–0.800)
Surface area	50	0.301 (0.118–0.487)	0.563 (0.287–0.740)
Total OSAS	49	0.546 (0.287–0.726)	0.783 (0.547–0.888)
Overall opinion	50	0.652 (0.502–0.773)	0.849 (0.752–0.911)
4-point scale	49	0.602 (0.447–0.734)	0.819 (0.708–0.892)

ICC, intra-class correlation coefficient; OSAS, Observer Scar Assessment Scale; CI, confidence interval.

The results of the linear multivariate regression analyses are shown in Table 4. For the patients, colour had the biggest impact on the total PSAS score. Relief was also important but did not reach a statistically significant level. For observers, the importance of the different aspects varied; pigmentation and colour were both of great importance to the overall opinion. One exception was the assessment made by the intern, which was not influenced significantly by pigmentation of the scar. For one of the three observers, pliability had great impact on the overall opinion and for another observer the thickness of the scar was of great importance.

Discussion

The overall opinion of the POSAS and the 4-point scale were associated with higher reliability compared to the total POSAS score or sub-items of the POSAS score. Furthermore, we have shown that observers do not seem to weight individual scar characteristics equally to arrive at an overall opinion, which challenges the assumption that calculating a total POSAS score by summing of the scores on the individual items is a valid approach.

ICCs on the 4-point scale and the overall opinion of the POSAS score were consistently higher than the ICC for the total OSAS score. The reliability for both scales increases when average results of multiple measurements are used for cosmetic assessment. Averaging the results of two measurements by patients and three measurements by observers gives more stable results, and assessment becomes less dependent on the method used for evaluating cosmetic outcome. This is in line with earlier findings for non-surgical facial scars.[6]

Both investigated scales have their own pros and cons. The 4-point scale is easy to use and has proved to be reliable. However, it lacks specific information on individual scar characteristics. The POSAS, on the other hand, do obligate observers to

Table 4 The regression coefficients with the corresponding 95% CI and *P*-values for the scores in individual characteristics of the OSAS and the PSAS

	Dermatologist				Resident				Intern			
	B	Lower 95% CI	Upper 95% CI	*P*	B	Lower 95% CI	Upper 95% CI	*P*	B	Lower 95% CI	Upper 95% CI	*P*
Vascularity	0.545	0.164	0.925	0.01	0.702	0.408	0.997	<0.01	0.299	0.014	0.584	0.04
Pigmentation	0.629	0.333	0.926	<0.01	0.373	0.097	0.649	0.01	0.145	−0.167	0.456	0.35
Thickness	0.015	−0.595	0.626	0.96	0.399	0.085	0.713	0.01	0.251	−0.046	0.547	0.10
Relief	0.162	−0.362	0.686	0.54	0.090	−0.218	0.398	0.56	0.178	−0.114	0.471	0.23
Pliability	0.738	−0.061	1.538	0.07	−0.063	−0.342	0.215	0.65	0.261	−0.043	0.565	0.09
Surface	−0.591	−1.248	0.066	0.08	0.264	−0.09	0.618	0.14	0.224	−0.027	0.476	0.08

	Patient			
	B	Lower 95% CI	Upper 95% CI	*P*
Pain	0.192	−0.703	1.087	0.67
Itching	0.131	−0.528	0.790	0.69
Colour	0.485	0.106	0.865	0.01
Pliability	0.102	−0.657	0.861	0.79
Thickness	−0.230	−1.146	0.687	0.62
Relief	0.509	−0.176	1.195	0.14

CI, confidence interval; OSAS, Observer Scar Assessment Scale; PSAS, Patient Scar Assessment Scale.

explicitly assess individual scar characteristics that determine the cosmetic outcome. This approach is intuitively appealing, because measurements are performed in a more standardized way, but the scores on individual POSAS sub-items showed great variability among observers. Furthermore, the underlying assumption that scores on individual items can be summed to get a total POSAS score is challenged by the findings from the multivariate regression analyses. These findings suggest that observers attach different weights to items to arrive to an overall opinion. This overall opinion is also a part of the POSAS questionnaire and when this overall opinion is formed in a different way than the total POSAS score, discrepancies may arise making interpretation difficult. An Additional disadvantage of the current scoring system is that the presence of 'railroad track' suture marks and 'dog ear' formation on the edges of the scar are not examined, although they tended to increase the score on overall opinion. In a comprehensive evaluation of scars, those items need to be incorporated in the grading system.

There are some limitations of the study. First of all, the three observers who assessed the scars varied in their levels of experience which could have contributed to lower inter-observer correlation rates on different items. However, this reflects the clinical setting in which scars are evaluated and does not affect the comparison between different scales. Of note, all observers have read and discussed the instructions of the two scales prior to assessment and reached consensus on the way of evaluation. Secondly, at the second assessment, not all items were completed by all the patients who returned their questionnaire, limiting the sample size for overall opinion of the PSAS. Apparently, for some patients, the PSAS is too difficult to complete.

Our findings raise the question whether the use of the POSAS score for assessment of cosmetic outcome for linear scars in the face has additional value when compared to the 4-point scale and the POSAS overall opinion score. The latter tools are easy and quick to use and are not inferior in their ability to discriminate between patients with good and poor cosmetic outcome. However, both scales have their limitations and the findings point to the need for improved and more optimal scales for cosmetic evaluation of linear scars.

References

1 Rogers HW, Weinstock MA, Harris AR *et al.* Incidence estimate of nonmelanoma skin cancer in the United States, 2006. *Arch Dermatol* 2010; **146**: 283–287.

2 Staples MP, Elwood M, Burton RC, Williams JL, Marks R, Giles GG. Non-melanoma skin cancer in Australia: the 2002 national survey and trends since 1985. *Med J Aust* 2006; **184**: 6–10.

3 Nikolaou V, Stratigos AJ. Emerging trends in the epidemiology of melanoma. *Br J Dermatol* 2014; **170**: 11–19.

4 Rumsey N, Clarke A, White P. Exploring the psychosocial concerns of outpatients with disfiguring conditions. *J Wound Care* 2003; **12**: 247–252.

5 Rhodes LE, de Rie MA, Leifsdottir R *et al.* Five-year follow-up of a randomized, prospective trial of topical methyl aminolevulinate photodynamic therapy vs surgery for nodular basal cell carcinoma. *Arch Dermatol* 2007; **143**: 1131–1136.

6 Mosterd K, Arits AH, Nelemans PJ, Kelleners-Smeets NW. Aesthetic evaluation after non-invasive treatment for superficial basal cell carcinoma. *J Eur Acad Dermatol Venereol* 2013; **27**: 647–650.

7 Bae SH, Bae YC. Analysis of frequency of use of different scar assessment scales based on the scar condition and treatment method. *Arch Plast Surg* 2014; **41**: 111–115.

8 Vercelli S, Ferriero G, Sartorio F, Stissi V, Franchignoni F. How to assess postsurgical scars: a review of outcome measures. *Disabil Rehabil* 2009; **31**: 2055–2063.

9 Draaijers LJ, Tempelman FRH, Botman YAM *et al.* The Patient and Observer Scar Assessment Scale: a Reliable and Feasible Tool for Scar Evaluation. *Plast Reconstr Surg* 2004; **113**: 1960–1965.
10 Truong PT, Lee JC, Soer B, Gaul CA, Olivotto IA. Reliability and validity testing of the Patient and Observer Scar Assessment Scale in evaluating linear scars after breast cancer surgery. *Plast Reconstr Surg* 2007; **119**: 487–494.
11 van de Kar AL, Corion LU, Smeulders MJ, Draaijers LJ, van der Horst CM, van Zuijlen PP. Reliable and feasible evaluation of linear scars by the Patient and Observer Scar Assessment Scale. *Plast Reconstr Surg* 2005; **116**: 514–522.
12 Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use, 3rd edn. Oxford University Press, New York, 2003.
13 Baryza MJ, Baryza GA. The Vancouver Scar Scale: an administration tool and its interrater reliability. *J Burn Care Rehabil* 1995; **16**: 535–538.
14 Giraudeau B, Mary JY. Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Stat Med* 2001; **20**: 3205–3214.

