Patient-Related Keloid Scar Assessment and Outcome Measures

Rebecca S. Nicholas, M.B.B.S.
Hannah Falvey, B.Sc.(Hons.), D.Clin.Psychol.
Pambos Lemonas, M.D., M.R.C.S.
Gopinath Damodaran, B.Sc., Ph.D.
Ali Ghannem, M.R.C.S., M.D., Ph.D.
Florendyn Selim, Harshad Navsaria, Ph.D.
Simon Myers, Ph.D., F.R.C.S.(Plas.)
London, United Kingdom

Background: Keloid scars cause pain, itching, functional limitation, and disfigurement, leading to psychological distress. Progress in treatment regimens is hindered by the lack of a universally accepted outcome measure. The Patient and Observer Scar Assessment Scale is a tool for the assessment of scars that includes scar assessment by both clinician and patient. This study evaluates the application of the scale to keloid scars, in addition to comparing it to the widely used Vancouver Scar Scale which, although originally designed for burns, is in wide use in many units across the world as the standard mode of assessment for scars.

Methods: Thirty-four patients with 41 keloid scars were assessed independently by three observers using the Vancouver Scar Scale and the Patient and Observer Scar Assessment Scale. Patients evaluated their own scars simultaneously using the patient component of the Patient and Observer Scar Assessment Scale. Internal consistency, interobserver reliability, and convergent validity were examined.

Results: Both components of the Patient and Observer Scar Assessment Scale had high internal consistency (0.82 and 0.86 for patient and observer components, respectively), and were higher than the Vancouver Scar Scale (0.65). Interobserver reliability was “substantial” for the Vancouver Scar Scale (0.65) and “almost perfect” for the observer component of the Patient and Observer Scar Assessment Scale (0.85). Convergent validity between the two was very strong (0.83, \( p < 0.01 \)), although the patient component did not correlate well with either of the observer scales. Patients rated their scars worse than the observer average for 83 percent of the scars, and were influenced by color, stiffness, thickness, and irregularity (\( p < 0.05 \)).

Conclusion: The findings support the use of the Patient and Observer Scar Assessment Scale as a reliable and valid method of assessing keloid scars in a clinical context. (Plast. Reconstr. Surg. 129: 1, 2012.)

CLINICAL QUESTION/LEVEL OF EVIDENCE: Diagnostic, II.

Keloid scars arise from pathologic wound healing following trauma and inflammation, and occur most commonly in those with darker pigmented skin. The resulting raised, irregular clusters of scar tissue extend well beyond the original wound margins, and their color is often discrepant to the surrounding skin, leading to significant disfigurement. The scars are associated with physical discomfort (pruritus and pain), functional limitation (particularly when located over joints), and significant psychological morbidity associated with their disfiguring appearance.\(^1\)\(^2\)

Keloid scars are notoriously difficult to treat. Extralesional surgical excision alone results in recurrence that can be worse than the original lesion. Other therapeutic methods include intralesional excision (debulking), extralesional excision and external superficial radiotherapy, interstitial irradiation, intralesional steroid injections, systemic chemotherapy, cryotherapy, ultrasound, zinc tape strapping, pressure garments, and silicone gels. Despite the wide range of therapeutic methods available, keloid scars typically remain refractory to treatment and have a high rate of recurrence. In addition, treatment complications such as local atrophy following steroid injections and late tu-

From the Academic Burn and Microsurgery Group, Blizard Institute of Cell and Molecular Science.
Received for publication May 6, 2011; accepted September 21, 2011.
Copyright ©2012 by the American Society of Plastic Surgeons
DOI: 10.1097/PRS.0b013e3182402c51

Disclosure: The authors have no financial interest to declare in relation to the content of this article.
Mor induction following radiotherapy are well recognized.\textsuperscript{1,3,4}

Progress in the identification of effective treatment regimens through clinical research has been hampered by the lack of a universally accepted method of measuring outcome.\textsuperscript{5,6} In addition, much research has focused on improving physical outcomes as judged by the clinician, although very little is known about whether such outcomes are perceived by patients as satisfactory. A suitable, feasible, and reproducible outcome measure in wide use would both allow comparison of keloid scar treatment outcomes across different clinical sites and also be useful in the local clinical setting for measuring individual patient responses to treatment.

The specialist keloid service at the Royal London Hospital is probably the largest multidisciplinary keloid clinic in Europe. The ethnically diverse communities surrounding the hospital present a high incidence of patients with keloid scars.

The Vancouver Scar Scale, first described by Sullivan et al. in 1990 (Fig. 1),\textsuperscript{7} has become one of the most widely applied scar assessment scales in clinical research.\textsuperscript{5} Although originally designed for burns, the Vancouver Scar Scale has been found to be clinically useful in evaluating a wide range of scar types and has become established in many units across the world as the standard mode of assessment. It allows assessment of pigmentation, vascularity, pliability, and scar height. Each of these variables is relevant to keloid scars which, without treatment, are typically pigmented, vascular (especially apparent in patients with light skin tones), stiff, and raised. In our clinical experience, each of these scar characteristics can respond positively to effective treatment, producing paler, less red, more supple, and flatter scars. Although the Vancouver Scar Scale is simple to use in a clinical setting and is well received by clinicians, it does not allow assessment of the patient perspective or symptoms associated with the scar, which are particularly relevant to keloid scar patients. The newer Patient and Observer Scar Assessment Scale is a scar evaluation tool that allows assessment of the same observed scar characteristics as the Vancouver Scar Scale in addition to incorporating the patient perspective (Figs. 2 and 3).\textsuperscript{8,9} For this reason, it is a potentially promising tool for the assessment of keloid disease. This study evaluates the application of the Patient and Observer Scar Assessment Scale to keloid scars and compares it to the widely used Vancouver Scar Scale in a clinical setting.

PATIENTS AND METHODS

Thirty-four patients attending the specialist keloid service at The Royal London Hospital, Whitechapel, Barts and The London NHS Trust were assessed as part of routine clinical practice between June 15 and July 15, 2010. Their mean age was 33 years (median, 29 years; range, 17 to 56 years). A total of 41 keloid scars were evaluated: 11 on the ear, 10 presternal keloids, eight on the shoulder, four on the anterior abdominal wall, four on the lower limb, two on the face, one on the breast, and one nuchal keloid. The surface area of the scars ranged from \(25 \text{ mm}^2\) (5\textsuperscript{th} and 5-mm scar) to \(21,700 \text{ mm}^2\) (155\textsuperscript{th} and 140-mm scar). Median surface area was \(825 \text{ mm}^2\). Five patients (15 percent) had Fitzpatrick skin type I to IV (white-beige skin tone), and 29 patients (85 percent) had Fitzpatrick skin type V to VI (brown-black skin tone).\textsuperscript{10} Seven patients were new to the clinic and had received no previous treatment. Most patients were already undergoing treatment with intralosional steroid injections: on average, patients had received six previous injections.

Scar Assessment Scales

All scars were assessed independently by three observers (one clinical nurse specialist, one spe-
PATIENT & OBSERVER SCAR ASSESSMENT SCALE

**PATIENT SCALE:**

**SCALe DEFINITION:**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
</tbody>
</table>

1. Has the scar been **painful** in the past few weeks?
   
2. Has the scar been **itching** in the past few weeks?
   
3. Is the scar **colour** different from the colour of your normal skin at present?
   
4. Is the scar **stiffness** different from the colour of your normal skin at present?
   
5. Is the scar **thickness** different from the colour of your normal skin at present?
   
6. Is the scar more **irregular** (bumpy) than your normal skin at present?
   
7. What is your **overall opinion** of the scar compared to your normal skin?

---

**Fig. 2.** The Patient and Observer Scar Assessment Scale: Patient scale.

---

Patient-related keloid scar assessment

A specialist registrar in plastic surgery, and one final-year medical student) on the same day and using both the Vancouver Scar Scale and the Patient and Observer Scar Assessment Scale (observer component). The three observers were blinded to each others’ scores. These scoring scales are shown in Figures 1 through 3. Each scar scale allowed the observer to rate a number of physical characteristics on a numerical scale. The Vancouver Scar Scale had four components—vasculature, pigmentation, pliability, and height—to give a total scar score ranging from 0 to 13 (where 0 represents normal skin). The Observer Scar Assessment Scale rates six physical characteristics—vascularity, pigmentation, pliability, thickness, relief, and surface area—in addition to an “overall opinion” component. The total score was calculated as a sum of all seven components of the scale, to give a score of 7 to 70 (where 7 represents normal skin with no associated pain or itching).9

**Statistical Analysis**

The data were analyzed using the statistical program PASW 18.0 (SPSS, Inc., Chicago, Ill.). Cronbach’s α, a reliability coefficient, was used to measure internal consistency for each of the two scales.11 Internal consistency is defined as the extent to which a set of items—in this case, the scale components—behave as a group. If the value is high, this is evidence that the scale components measure the same underlying construct. A reliability coefficient of 0.70 or higher is considered “acceptable.”12,13

Interobserver reliability is the degree to which assessments from two or more observers agree.
This was measured by calculating the intraclass correlation coefficient using a two-way mixed model with measures of consistency. The intraclass correlation coefficient was calculated both for single observer (single measure) and for the three observers (average measure). In evaluating this, a correlation coefficient of 0 to 0.2 was interpreted as “slight”; 0.21 to 0.40, “fair”; 0.41 to 0.60, “moderate”; 0.61 to 0.80, “substantial”; and 0.81 to 1.0, “almost perfect.”

Convergent validity refers to the degree to which a rating is similar to another independently gathered rating to which it should be theoretically similar. We used the Pearson correlation coefficient to determine the degree of correlation between each of the components of the Vancouver Scar Scale and both parts of the Patient and Observer Scar Assessment Scale.

We performed simple linear regression analysis to identify factors that significantly influenced patients’ overall opinion of their scars. In addition to examining each component of the Patient Scar Assessment Scale (scar-related pain, itchiness, color, stiffness, thickness, and irregularity), the amount of treatment received to date (number of steroid injections—the mainstay of treatment in the keloid clinic), and the scar surface area were also tested in the regression model.

RESULTS

Internal Consistency

Both components of the Patient and Observer Scar Assessment Scale were found to have relatively high internal consistency, with reliability coefficients of 0.82 for the patient component and 0.86 for the observer component. However, a reliability coefficient of only 0.65 was found for the Vancouver Scar Scale, suggesting that the scale items have relatively low internal consistency.
Interobserver Reliability

The interobserver reliability of the Vancouver Scar Scale and the observer component of the Patient and Observer Scar Assessment Scale are summarized in Table 1. The intraclass correlation coefficient was consistently greater for the group of three observers (average measure) compared with the single observer (single measure). Looking at the total score for each of the scales, interobserver reliability was “substantial” for the Vancouver Scar Scale, and “almost perfect” for the observer component of the Patient and Observer Scar Assessment Scale. For the individual Vancouver Scar Scale variables, the interobserver reliability was “almost perfect” for pigmentation and height, “moderate” for pliability, but only “slight” for vascularity. For the observer component of the Patient and Observer Scar Assessment Scale, interobserver reliability was “almost perfect” for pliability, surface area, and overall opinion; “substantial” for pigmentation, thickness, and relief; and “moderate” for vascularity.

Convergent Validity

As Table 2 shows, the two scales had a strong correlation with each other, with values of $p < 0.01$ for all. The scores for pliability and height had the strongest correlation coefficients (0.88 and 0.85, respectively), whereas correlations for pigmentation and vascularity scores were the lowest of the group of variables (0.64 and 0.70, respectively).
Table 1. Interobserver Reliability of the Vancouver Scar Scale and the Observer Component of the Patient and Observer Scar Assessment Scale

<table>
<thead>
<tr>
<th></th>
<th>Single Measure ICC (95% CI)</th>
<th>Average Measure ICC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vancouver Scar Scale</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vascularity</td>
<td>0.054 (–0.11–0.26)*</td>
<td>0.14 (–0.43–0.51)*</td>
</tr>
<tr>
<td>Pigmentation</td>
<td>0.39 (0.43–0.74)$</td>
<td>0.92 (0.69–0.89)$</td>
</tr>
<tr>
<td>Pliability</td>
<td>0.34 (0.15–0.54)$</td>
<td>0.60 (0.34–0.78)$</td>
</tr>
<tr>
<td>Height</td>
<td>0.69 (0.55–0.81)$</td>
<td>0.87 (0.78–0.95)$</td>
</tr>
<tr>
<td>Total score</td>
<td>0.38 (0.19–0.57)$</td>
<td>0.65 (0.41–0.80)$</td>
</tr>
<tr>
<td><strong>Observer component of the</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POSAS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vascularity</td>
<td>0.24 (0.05–0.45)$</td>
<td>0.49 (0.15–0.71)$</td>
</tr>
<tr>
<td>Pigmentation</td>
<td>0.54 (0.36–0.7)$</td>
<td>0.77 (0.62–0.87)$</td>
</tr>
<tr>
<td>Pliability</td>
<td>0.59 (0.43–0.74)$</td>
<td>0.81 (0.69–0.89)$</td>
</tr>
<tr>
<td>Thickness</td>
<td>0.57 (0.4–0.72)$</td>
<td>0.80 (0.66–0.88)$</td>
</tr>
<tr>
<td>Relief</td>
<td>0.43 (0.24–0.61)$</td>
<td>0.70 (0.50–0.83)$</td>
</tr>
<tr>
<td>Surface area</td>
<td>0.60 (0.43–0.75)$</td>
<td>0.82 (0.70–0.90)$</td>
</tr>
<tr>
<td>Overall opinion</td>
<td>0.66 (0.51–0.79)$</td>
<td>0.85 (0.76–0.91)$</td>
</tr>
<tr>
<td>Total score</td>
<td>0.66 (0.51–0.79)$</td>
<td>0.85 (0.76–0.91)$</td>
</tr>
</tbody>
</table>

CI, confidence interval; Single Measure ICC, intraclass correlation coefficient for a single observer; Average Measure ICC, intraclass correlation coefficient for the group of three observer; POSAS, Patient and Observer Scar Assessment Scale.

Key: *slight; †fair; ‡moderate; §substantial; ‡almost perfect.

Correlation between each of the observer scales and the patient component of the Patient and Observer Scar Assessment Scale were not statistically significant for any of the variables on either scale (Table 3). However, it was notable that patients’ total scores were higher than the average observer total for the same scar in 83 percent of cases (n = 33).

Patient Scar Assessment Scale

Forty-three percent of patients reported that their scar had been painful over the preceding few weeks, and scar-related itching was reported by 70 percent of patients. Almost all patients reported that scar color (93 percent), thickness (95 percent), stiffness (98 percent), and surface irregularity (98 percent) were different from their normal skin, and at least half of the patients (53, 50, and 50 percent, respectively) rated these physical scar characteristics as very different from their normal skin (8 to 10 of 10). Seventy percent of patients had an overall opinion of their scar of 8 to 10 of 10.

Simple Linear Regression Analysis

Table 4 shows that patients’ overall opinion of their scars was significantly influenced by scar color, stiffness, thickness, and irregularity (p <
0.05 for all), with color, thickness, and irregularity being the strongest determinants of the variation in overall opinion [coefficients of determination ($r^2$) of 0.24, 0.21, and 0.32 respectively]. The influence of scar-related pain, itchiness, number of treatments, or scar surface area on patients' overall opinion was not shown to be significant ($p > 0.05$).

**DISCUSSION**

A simple and reproducible way of assessing keloid scars in routine clinical practice is needed to monitor treatment outcomes. An accurate and universally accepted method of keloid scar assessment will assist the development of improved treatment regimens. Such an assessment tool must be valid, consistent, reliable, and feasible for use in the outpatient clinic setting.

The Vancouver Scar Scale is a widely respected and commonly used scar assessment scale. Before this study, it was the scar assessment tool of choice in the Royal London Hospital Keloid Clinic. However, the Vancouver Scar Scale fails to capture clinical information from the patient’s perspective, the importance of which is increasingly being recognized as we seek to attain patient satisfaction with scar outcome. For this reason, the Patient and Observer Scar Assessment Scale is a promising new scar assessment scale, as it combines an observer’s clinical assessment with patient evaluation.

This study is the first to evaluate the application of the Patient and Observer Scar Assessment Scale for the assessment of keloid disease specifically. It provides evidence to support the use of the Patient and Observer Scar Assessment Scale for the assessment of keloid scars; when compared with the current best available observer scale, the Vancouver Scar Scale, the Patient and Observer Scar Assessment Scale observer component showed excellent concurrent validity. In addition, the Patient and Observer Scar Assessment Scale was shown to have a number of advantages over the Vancouver Scar Scale. In particular, it scored higher on measures of internal consistency and reliability: both components of the Patient and Observer Scar Assessment Scale were found to have excellent internal consistency, whereas the results for the Vancouver Scar Scale failed to demonstrate acceptable internal consistency. A lack of consistency of the Vancouver Scar Scale was also found by Draaijers et al. in 2004. They suggested that this may be attributable to the pigmentation variable on the Vancouver Scar Scale appearing nominal, rather than being a “proper numerical scale” as seen on the Patient and Observer Scar Assessment Scale. However, on reviewing the original article that first describes the Vancouver Scar Scale, the rationale for the pigmentation score is explained: hypopigmentation scored 1, as it was deemed to be less significant than hyperpigmentation (scoring 2) for white, Indian, and Asian populations. However, as keloid disease predominates in extreme Fitzpatrick skin types (those with darker skin), this original rationale for pigmentation in the Vancouver Scar Scale does not seem applicable: hypopigmentation can be just as disfiguring for patients with dark skin as extreme hyperpigmentation in lighter skin. The Patient and Observer Scar Assessment Scale scores pigmentation ranging from the patient’s normal skin color (score of 1) to the opposite or most different extreme of pigmentation (score of 10) and therefore seems more suitable for the assessment of altered scar pigmentation in a population comprising diverse Fitzpatrick skin types. Furthermore, the interobserver reliability of the Patient and Observer Scar Assessment Scale completed both by multiple observers and by a single observer was significantly better than the Vancouver Scar Scale. Perhaps, as would be expected, the interobserver reliability for both the Patient and Observer Scar Assessment Scale and the Vancouver Scar Scale was better for three observers than for a single observer, but the reliability for three observers was “almost perfect” for the Patient and Observer Scar Assessment Scale and only “substantial” for the Vancouver Scar Scale. The reliability of the observer scale for a single observer was 0.66 for the Patient and Observer Scar Assessment Scale—higher than for the Vancouver Scar Scale (0.38). Although this does not quite meet the recommended value of 0.70, it is worth noting that it is higher than the results for a comparable study that concluded the Patient and Observer Scar Assessment Scale had “acceptable interobserver reliability” with 0.33 for a single measure and 0.60 for an average measure with three observers. Consequently, it seems there needs to be greater agreement in such evaluations regarding what degree of reliability is “acceptable” for clinical practice.

The Patient Scar Assessment Scale showed excellent internal consistency; however, it did not correlate well with the observer scales. This in itself is a valuable finding because, as clinicians, we may think we know what the patient perceives as a successful outcome when in fact our views can be different from those of our patients. On the whole, patients thought worse of their scars than did the...
observers, and most had a very negative overall opinion. A similar discrepancy between observer and patient scores was noted in the 2005 article by van de Kar et al., who suggested this could be partly attributed to itching and pain. However, unlike the study by van de Kar et al., here patients’ overall opinions were not shown to be significantly influenced by pain and itching. Instead, we suggest that patients’ impression of their scar may be related to the degree of psychological distress associated with the visible disfigurement caused by keloid disease, as patients’ overall opinion was most strongly influenced by the visible characteristics of their scars. Anecdotally, when completing the self-assessment scale, many of the patients involved in this study expressed a wish to describe the wider impact of their keloid scars on their life, mentioning factors such as clothing, relationships, and social activities. This merits further study, with a view to possibly incorporating such factors into the patient assessment scale for keloid scars.

It is important that any tool that forms part of routine clinical assessment is feasible for use in a busy outpatient clinic setting. In this study, given initial training and some practice, we found that the Patient and Observer Scar Assessment Scale was quick and easy to use. From our experience of using the scale in a busy clinical setting, we recommend some amendments to the scale pro forma that we found helpful: to make a clear distinction between the numerical score and the additional clinical information under the six category headings, we reformatted the scale to make the order of completion clear. In addition, observers found it helpful to annotate the pro forma with the definitions and clarifications provided by the authors (Figs. 2 and 3). As the meaning of the term “relief” was not immediately clear to all users of the scale, we recommend renaming this category “surface irregularities” or “bumpiness,” as defined by the original authors. Figure 5 shows guidelines that we adapted from the original pub-
application after consultation with the authors on areas of potential confusion. This was used for training observers. We recommend that the original authors might supplement this with a formal training package, complete with “anchor photographs” correlating to each 1-to-10 scale component. This would improve training and enhance the reliability of the scale. We also suggest a patient-friendly instruction guide to assist patients who are new to using the patient component the scale. Finally, “overall opinion”, although not strictly a clinical parameter, was a particularly useful indicator of response to treatment in a clinical setting. This article demonstrates that the inclusion of this category in the total score does not detract from the scale’s reliability or validity, and we would therefore advocate its inclusion in the total score.

CONCLUSIONS

For the assessment of keloid scars, the Patient and Observer Scar Assessment Scale is associated with high internal consistency and “almost perfect” interobserver reliability. It is more comprehensive than the Vancouver Scar Scale, which had lower internal consistency and lower interobserver reliability by comparison. Neither scale correlated well with patients’ ratings, although the patient component of the Patient and Observer Scar Assessment Scale provides valuable clinical information about the patient’s symptoms and perspective on their scar, which is currently missing from any other scar assessment scoring system. The findings of this study support the use of the Patient and Observer Scar Assessment Scale as a more useful, reliable, and valid method of assessing keloid scars in a clinical context.

Pambos Lemonas, M.D., M.R.C.S.
The Barts and the London NHS Trust Whitechapel London E1 1BB, United Kingdom

ACKNOWLEDGMENT

This study was supported by the Healing Foundation in association with the British Association of Plastic and Reconstructive Surgeons.

REFERENCES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1: AUTHOR—Author lines: Degrees were taken from the Enkwell transmittal. Please confirm that they are correct, and supply a degree for Florendyn Selim, if wanted.

AQ2: AUTHOR—Correspondence information: Please provide a street address, if possible, and an email address if you would like it to appear in the printed article.

AQ3: AUTHOR—Per Journal style, the abstract should not exceed 250 words. The current abstract has 293 words. Please shorten it as needed to meet Journal word limit requirements. Please note that the Journal prefers to avoid the use of abbreviations and acronyms.

AQ4: AUTHOR—Correct as edited (“Pearson correlation coefficient," adding the word coefficient)? Please advise/revise as needed.

AQ5: AUTHOR—Key footnote to Table 1: Shades of gray were replaced by footnote symbols; please check that this has been done correctly.