

# Reliability and Photographic Equivalency of the Scar Cosmesis Assessment and Rating (SCAR) Scale, an Outcome Measure for Postoperative Scars

Jonathan Kantor, MD, MSCE, MA

 Supplemental content

**IMPORTANCE** Until recently, no ideal valid, feasible, and reliable scar scale existed to effectively assess the quality of postoperative linear scars. The Scar Cosmesis Assessment and Rating (SCAR) scale was developed and validated as a tool to assess the quality of postoperative scars in clinical and research settings.

**OBJECTIVE** To assess the reliability of using photographs in lieu of live patient scar rating assessments, and to determine the interrater and intrarater reliability of the SCAR scale.

**DESIGN, SETTING, AND PARTICIPANTS** This was a reliability study to assess clinicians' interrater and intrarater reliability, as well as the reliability of using high-quality macrophotographs of postoperative scars. Patients were from a private practice dermatology clinic, with assessed scars representing a range of surgical procedures including those performed by dermatologists, plastic surgeons, and general surgeons. Assessments were performed by an international multidisciplinary team from dermatology, plastic surgery, surgical oncology, emergency medicine, and physiatry, using photographs and live patient assessments. A single photograph was assessed for each patient's scar. Data were obtained between August 3, 2015, and January 18, 2016. Data analysis occurred between January 18, 2016, and July 29, 2016. Using the intraclass correlation coefficient (ICC), the scale was tested for photographic equivalency as well as interrater reliability and intrarater reliability by 5 raters on a set of 80 total patient scars, 20 of which were analyzed for photographic equivalency and the remaining 60 of which were analyzed for interrater and intrarater reliability.

**MAIN OUTCOMES AND MEASURES** The SCAR scale that measures postoperative scar cosmesis, with scores ranging from 0 (best possible scar) to 15 (worst possible scar), based on 6 clinician and 2 patient items was used. Of those 60 in the photographic subgroup, 10 were rated using not only the SCAR scale but also the Patient and Observer Scar Assessment Scale and the Vancouver Scar Scale, and 10 were assessed twice by the same rater at different times to assess intrarater reliability.

**RESULTS** Patients' ages ranged from 18 to 96 years, with Fitzpatrick skin types I through VI. Thirty-seven were male, and 43 were female. A set of 20 live patient scars with associated photographs, as well as a separate set of 60 photographs, were rated; 10 patients were assessed twice for intrarater reliability. The SCAR scale ratings using photographs were found to be largely equivalent to live patient assessments, with ICCs of 0.99 (95% CI, 0.96-0.99) and 0.98 (95% CI, 0.96-0.99). The interrater reliability of the overall scale showed an ICC of 0.95 (95% CI, 0.96-0.99) using a 2-sample random-effects model. Intrarater reliability found ICCs ranging from 0.96 to 0.99 with 5 separate raters. Modeling the overall SCAR score predicted whether the rater would consider the scar undesirable, with an odds ratio of association of 1.76 (95% CI, 1.24-2.2). A secondary analysis of Fitzpatrick skin types IV, VI, and VI demonstrated a sustained interrater reliability, with an ICC of 0.93 (95% CI, 0.86-0.98).

**CONCLUSIONS AND RELEVANCE** The SCAR scale is a reliable rating scale for postoperative linear scars, and photographs may reliably be used in lieu of live patient assessments. The SCAR scale therefore represents a reliable standard rating scale for postoperative scar cosmesis.

*JAMA Dermatol.* 2017;153(1):55-60. doi:10.1001/jamadermatol.2016.3757  
Published online November 2, 2016. Corrected on January 11, 2017.

**Author Affiliations:** Department of Dermatology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia; Florida Center for Dermatology, PA, St Augustine.

**Corresponding Author:** Jonathan Kantor, MD, MSCE, MA, Florida Center for Dermatology, PA, PO Box 3044, St Augustine, FL 32085 (jonkantor@gmail.com).

All desirable scars are alike; all undesirable scars are undesirable in different ways. Assessing the merits of various surgical approaches and outcomes is predicated on the existence of clinically valid and statistically reliable measurement scales. Many scales, of varying length and complexity, have been described for assessing scars.<sup>1-4</sup> Historically, the Vancouver Scar Scale (VSS)<sup>5</sup> was used to assess scarring, although this 4-item scale was initially developed to assess burn scars rather than postoperative linear scars, which have a very different set of clinical considerations. More recently, the Patient and Observer Scar Assessment Scale (POSAS) was developed for burn scars,<sup>6</sup> although several years after initial scale development it was reliability tested for linear scars as well.<sup>7</sup> Other scar scales in use include the Hamilton Scale,<sup>8</sup> the Seattle Scale,<sup>9</sup> the Manchester Scar Scale,<sup>10</sup> the Stony Brook Scar Evaluation Scale,<sup>11</sup> the University of North Carolina 4P Scar Scale,<sup>12</sup> and their variants.

Until recently, however, no psychometrically rigorous scale existed that was designed from the ground up to assess the evolution of postoperative linear scar cosmesis and function.<sup>13</sup> The Scar Cosmesis Assessment and Rating (SCAR) scale, whose initial development and validation has been described separately,<sup>14</sup> was developed as a rating scale for postoperative linear scars that could be used with both live patients and photographs, while capturing change in a particular scar component over time. This scale was built on prior work, including the development of a patient rating scale that highlighted the need to differentially weight scar hypertrophy, spread, and erythema when assessing scar quality.<sup>15</sup> The SCAR scale is thus unique in that the items are weighted based on the degree of clinical importance assigned by the multispecialty validity committee, as well as patients.

The aim of this study was to (1) assess whether high-quality clinical photographs may be used in lieu of live SCAR scale assessments by raters and (2) to assess both the interrater and intrarater reliability of the SCAR scale on a range of postoperative scars.

## Methods

Several steps are required to develop a psychometrically rigorous rating scale; the details of this methodology have been previously described.<sup>3,16</sup> Scoring of the SCAR scale (Table 1) is based on 6 observer-scored items and 2 patient-scored items.<sup>14</sup> No identifying characteristics were recorded, so no written informed consent was received other than a blanket photographic release that was signed by all patients. No compensation was given to participants. The institutional review board at St Vincent's Medical Center considered the study exempt (see the study protocol in the Supplement). Eighty patients and scars were divided as follows: 20 were analyzed for photographic equivalency (ie, both live scar and photograph were examined), and 60 photographs of the 60 remaining patients analyzed by 5 raters. Of those 60, 10 were rated using not only the SCAR scale but also the Patient and Observer Scar Assessment Scale and the Vancouver Scar Scale, and 10 were assessed twice by the same rater at different times to assess intrarater reliability.

## Key Points

**Questions** What is the reliability of the Scar Cosmesis Assessment and Rating (SCAR) scale, and can photographs be assessed in lieu of live patient ratings?

**Findings** In this reliability study of a set of 60 patient scars, the interobserver reliability was excellent, with an intraclass correlation coefficient (ICC) of 0.95, while the intraobserver reliability showed ICCs ranging from 0.96 to 0.99 on a subset of 10 scars. Near-equivalence was seen between photographic and in-person rating on a separate set of 20 patient scars.

**Meaning** These findings, coupled with the excellent feasibility of the scale, suggest that the SCAR scale could become the new standard outcome measure for postoperative scar quality.

## Photograph and Live Patient Equivalency

Prior to completing the full reliability study, the SCAR scale was assessed for equivalency between its use on live patients and high-quality photographs. Photographs were taken with a specialized macrolens and flash (Canon EOS 70D 20.2 MP SLR; Canon EF 100-mm f/2.8 L IS USM lens; Canon MR-14EX II Macro Ring Lite). The camera settings were set to automatic, the LED flash unit was set to TTL (through the lens) metering, and all photographs were taken at the camera's maximum resolution using standard JPEG compression. Intrarater reliability to assess for photograph and live patient equivalency was calculated for 2 raters for a set of 20 live patient scars and photographs of these scars. Live patient assessments were performed in a single 1-day session, and assessors were able to palpate and closely observe the scars. Grading of the photographs was then performed 1 week later with photographs set in a random order.

## Assessment of Interrater and Intrarater Reliability

Interrater reliability was assessed by a group of 5 clinicians (4 board-certified dermatologists and 1 physician's assistant) who scored a separate set of 60 high-quality scar photographs of scars reflecting a broad range of patient ages, skin types, and severity. Clinicians were purposely not trained on the SCAR scale using baseline photographs, a technique consistent with the development of other scar rating scales<sup>1-3,6</sup> and other rigorous outcome measures in dermatology<sup>17</sup> to avoid favorably biasing the interrater reliability findings. All patients were rated using the SCAR scale and an overall visual analog scale (VAS) for scar quality. A subset of 10 patients was also evaluated using the OSAS and VSS. Representative photographs are shown in the Figure.

To assess for intrarater reliability, 10 patients were assessed twice by each rater. To minimize the chance of recall, raters were not informed beforehand that selected scars would be rated twice, which were randomly inserted into the photographic set.

## Secondary Logistic Regression Analysis

A secondary analysis of internal consistency was performed on both the full data set as well as the subset of 10 scars assessed by all 3 scales to assess whether the overall SCAR scale, as well as the OSAS and VSS, predicted the presence of a desirable scar.

### Secondary Analysis for Fitzpatrick Skin Types IV to VI

A secondary analysis of the interrater reliability of SCAR scale was performed for scars of patients with Fitzpatrick skin types IV, V, and VI.

### Statistical Analysis

Interrater reliability was assessed with the intraclass correlation coefficient (ICC) using a 2-way random-effects model.<sup>18</sup> Intrarater reliability was calculated using the ICC with a 1-way random-effects model.<sup>19</sup> All modeling was performed both with individual scale components and the overall SCAR score, and results are reported with 95% CIs. Univariate and multivariate logistic regression was used as a secondary analysis of internal consistency by modeling a desirable scar (as defined by a single desirability question) as the dependent variable. All statistical analyses were performed using Stata software (version 13 for Mac; Stata Corp).

## Results

### Scar and Patient Characteristics

Data were obtained between August 3, 2015, and January 18, 2016. Data analysis occurred between January 18, 2016, and July 29, 2016. Assessed scars included a wide range of clinical outcomes, from nearly undetectable scars to large keloids, and in patients with a wide age range (18-96 years) and varying skin types (Fitzpatrick types I-VI). Thirty-seven were male, and 43 were female. SCAR scale clinician scores found in this study ranged from 0 to 13, reflecting the full range of possible scores.

### Photograph and Live Patient Equivalency

There was near equivalence between SCAR scale ratings for live patients and for high-quality photographs. The ICC for each of 2 raters on a separate 20-scar set of live patients and photographs was 0.99 (95% CI, 0.96-0.99) and 0.98 (95% CI, 0.96-0.99), suggesting clinical equivalence between the SCAR scale's use on live patients and high-quality macrophotographs. The lowest ICC for any 1 component of the scale was for the hypertrophy and atrophy component, which was 0.90 (95% CI, 0.73-0.96) for one rater and 0.96 (95% CI, 0.91-0.99) for the other.

### Interrater and Intrarater Reliability

For interrater reliability, the ICC for the overall scale, using a 2-sample random-effects model, was found to be outstanding at 0.95. Descriptive statistics for all scar assessment scales and each of their components, as well as the ICC for each scale component, are included in **Table 2**. Overall, the interrater reliability of the scale was equal to or better than previously reported scales. This finding suggests that there is remarkable agreement between different clinicians when scoring the SCAR scale.

Intrarater reliability for the SCAR scale was similarly excellent, with 1-sample random-effects model ICCs ranging from 0.96 to 0.99 for each rater using a 1-way random effects model ( $P < .001$  for all comparisons). This highlights the outstand-

**Table 1. The Scar Cosmesis Assessment and Rating (SCAR) Scale**

Clinician Items	Scale Ratings
Scar spread	0, None to near-invisible
	1, Pencil-thin line
	2, Mild spread, noticeable on close inspection
	3, Moderate spread, obvious scarring
Erythema	4, Severe spread
	0, None
	1, Light pink, some telangiectasias may be present
Dyspigmentation (includes hyperpigmentation and hypopigmentation)	2, Red, many telangiectasias may be present
	3, Deep red or purple
Track marks or suture marks	0, Absent
	1, Present
Hypertrophy/atrophy	0, None
	1, Mild: palpable, barely visible hypertrophy or atrophy
	2, Moderate: clearly visible hypertrophy or atrophy
	3, Severe: marked hypertrophy or atrophy or keloid formation
Overall impression	0, Desirable scar
	1, Undesirable scar
Patient items	
Have you been bothered by any itch from the scar in the past 24 h?	0, No
	1, Yes
Have you been bothered by any pain from the scar in the past 24 h?	0, No
	1, Yes
Total score range	0 (best possible scar) to 15 (worst possible scar)

ing test-retest reliability of the SCAR scale, meaning that the same observer is likely to rate the same scar with the same overall score when assessed a second time.

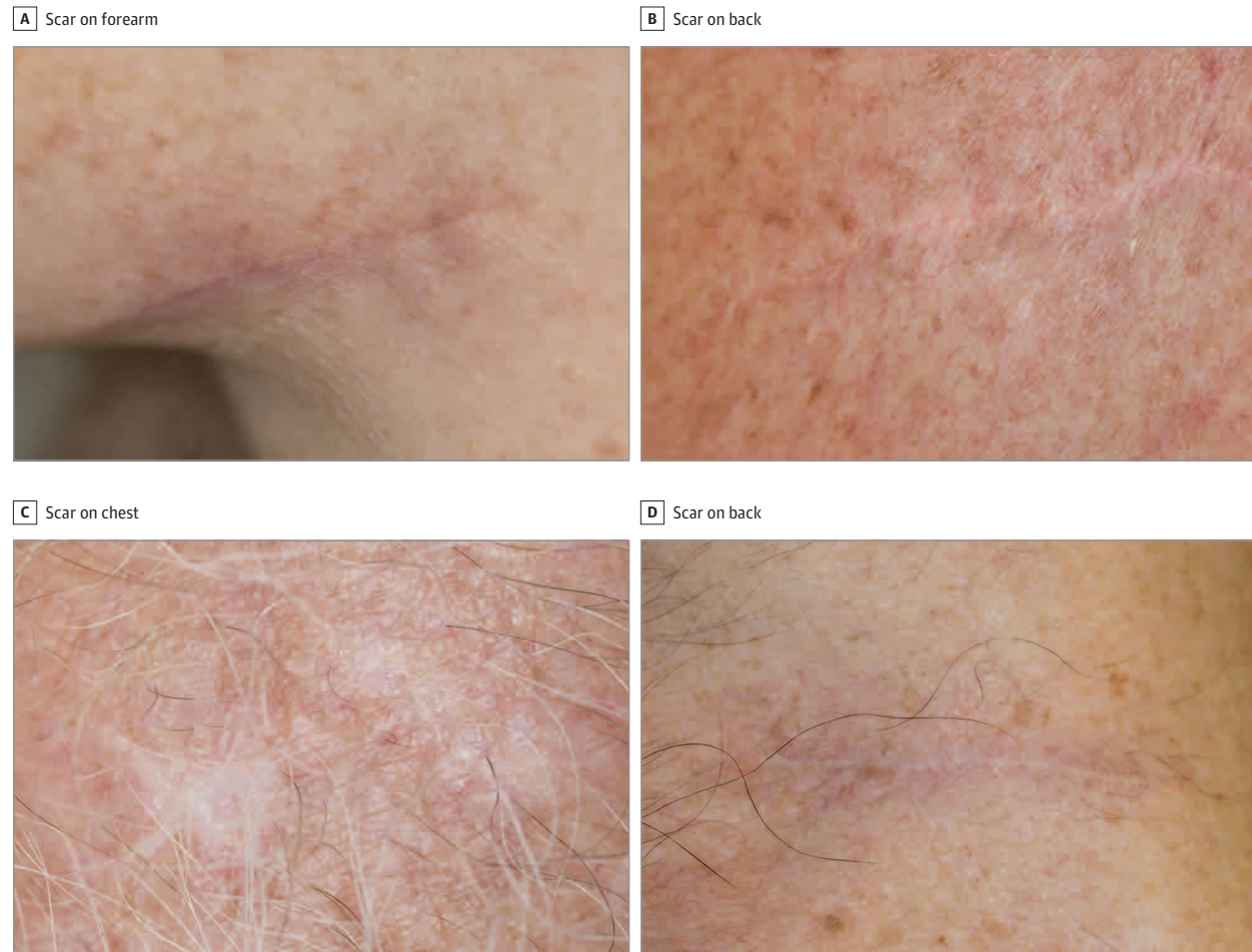
### Secondary Logistic Regression Analysis

Internal consistency, which was previously found to be very good (Cronbach  $\alpha = .77$ , average interitem covariance of 0.11), was further demonstrated by a secondary logistic regression analysis. Modeling the overall SCAR score significantly predicted whether the rater would consider the scar desirable with an odds ratio of association of 1.76 (95% CI, 1.24-2.2;  $P < .001$ ). Thus for each 1-point increase in the SCAR scale there is a 76% increased risk of the scar being considered undesirable. Moreover, in a multivariate logistic regression model, with a desirable scar as a binary outcome and including the SCAR scale, OSAS, and VSS, only the SCAR scale demonstrated statistical significance. These secondary analyses highlight the robustness of the SCAR scale as a clinically valid outcome measure.

### Secondary Analysis for Fitzpatrick Skin Types IV to VI

The interrater reliability of the SCAR scale on scars in patients with Fitzpatrick type IV, V, and VI skin ( $n = 20$ ) was excellent, with an ICC of 0.93 (95% CI, 0.86-0.98). The ICC for the erythema

Figure. Representative Scar Photographs



Values are represented as means (95% CIs) for all 5 raters. A, Observer scar assessment scale (OSAS), 11.0 (4.6-17.4); Vancouver Scar Scale (VSS), 3.2 (1.8-4.6); Scar Cosmesis Assessment and Rating (SCAR) scale, 3.8 (2.2-5.4).

B, OSAS, 6.8 (0-14.8); VSS, 1.2 (0.2-2.2); SCAR, 2.4 (1.0-3.8). C, OSAS, 17.4 (6.5-28.3); VSS, 2 (1.1-2.9); SCAR, 4.6 (2.9-6.3). D, OSAS, 10.6 (5.4-15.8); VSS, 2.4 (1.0-3.8); SCAR, 3.4 (1.9-4.8).

component in this patient subset was similarly excellent at 0.92 (95% CI, 0.84-0.96), suggesting that the SCAR scale is appropriate for use in diverse populations with multiple skin types.

## Discussion

The SCAR scale represents a reliable outcome measure for linear scar assessment that may be used to assess both live patients and high-quality scar photographs. The ability to reliably assess photographs of scars is a significant advantage of this scale because it allows future outcome studies and clinical trials to include a wide array of raters from varying geographic locales.

The SCAR scale should be understood in the context of preexisting scar assessment scales. It includes 6 observer components and 2 patient components, which adds significantly to its feasibility. Including additional components in the SCAR scale, such as scar age and anatomic location, would result in possibly overfitting the statistical model while losing

track of the goal of the outcome measure: to measure the overall quality of an individual scar to permit valid, reliable, and clinically meaningful conclusions to be drawn when comparing interventions and techniques.

Most of the SCAR scale components are graded as either binary (yes/no) outcomes or are linked to a clinically objective outcome. This approach is shared with other scales, such as the VSS. The OSAS, in contrast, uses a 1-10 rating for each item. While the advantage of the latter approach is that a broad range of choices are available, ultimately this results in a series of multiple VASs that capture the assessors' *overall impression* of the scar rather than the presence or absence of *specific objective* clinical findings.

The SCAR scale's overall score is generated by combining the clinician and patient sections; while these may be reported separately, as in the POSAS, the relative rarity of patient-reported symptoms of itch or pain means that this is not necessary in clinical practice. Unlike the POSAS, the SCAR scale does not include both observer and patient assessments of the same outcomes. Studies have consistently demonstrated a high



Table 2. Descriptive Statistics and Interrater Reliability of Scar Assessment Scales<sup>a</sup>

Scale	Component	Mean (SE)	(95% CI)	ICC (95% CI)
SCAR	Total score	4.15 (0.15)	(3.85-4.44)	0.95 (0.93-0.97)
	Spread	1.61 (0.06)	(1.49-1.72)	0.92 (0.89-0.95)
	Erythema	0.49 (0.04)	(0.42-0.57)	0.92 (0.88-0.95)
	Dyspigmentation	0.74 (0.02)	(0.69-0.79)	0.66 (0.51-0.77)
	Track marks	0.27 (0.02)	(0.22-0.32)	0.77 (0.66-0.85)
	Hypertrophy/atrophy	0.69 (0.04)	(0.61-0.78)	0.92 (0.88-0.94)
	Overall impression	0.34 (0.03)	(0.29-0.39)	0.84 (0.77-0.90)
OSAS	Total score	14.76 (2.02)	(10.69-18.83)	0.95 (0.88-0.99)
	Vascularity	2.20 (0.35)	(1.50-2.90)	0.89 (0.73-0.97)
	Pigmentation	2.08 (0.22)	(1.63-2.53)	0.73 (0.36-0.92)
	Thickness	1.84 (0.35)	(1.13-2.55)	0.94 (0.85-0.98)
	Relief	1.66 (0.35)	(0.96-2.36)	0.95 (0.87-0.99)
	Pliability	1.74 (0.36)	(1.02-2.46)	0.96 (0.90-0.99)
	Surface area	2.28 (0.35)	(1.58-2.98)	0.93 (0.83-0.98)
	Overall opinion	2.96 (0.36)	(2.24-3.68)	0.91 (0.77-0.98)
VSS	Total score	2.96 (0.41)	(2.13-3.79)	0.97 (0.91-0.99)
	Pigmentation	0.84 (0.10)	(0.63-1.04)	0.40 (0-0.82)
	Vascularity	0.76 (0.11)	(0.53-0.99)	0.92 (0.81-0.98)
	Pliability	1.02 (0.19)	(0.64-1.40)	0.97 (0.91-0.99)
	Height	0.34 (0.10)	(0.14-0.54)	0.96 (0.91-0.99)
VAS	Overall	66.57 (1.43)	(63.75-69.38)	0.91 (0.84-0.95)

Abbreviations: ICC, intraclass correlation coefficient; OSAS, observer scar assessment scale; SCAR, Scar Cosmesis Assessment and Rating scale; SE, standard error; VAS, visual analog scale; VSS, Vancouver Scar Scale.

<sup>a</sup>A 2-way random-effects model was used for all analyses.  $P < .001$  for all comparisons.

correlation between observer and patient measures,<sup>6,7,20</sup> and thus an opinion-based measure for overall patient assessment of the scar is unlikely to add significantly to the value of a rating system, while concomitantly introducing increased complexity and potential bias. Moreover, as has been recently explored with psoriasis outcome measures, the redundancy inherent in longer and more complex scoring systems may be counterproductive.<sup>21</sup>

Preexisting scar assessment scales have significant shortcomings. The Stony Brook scar evaluation scale was developed in the context of emergency medicine primarily for assessing the quality of laceration repairs.<sup>11</sup> While it shares some features with the VSS,<sup>22</sup> its main advantage—binary yes/no outcomes for several important scar features, such as spread and hypertrophy—is also an important drawback because it fails to capture degrees of quality difference in the scar or responsiveness to change, and also does not capture the differential importance of certain scar characteristics.

The Manchester Scar Sale,<sup>10</sup> also developed for linear scars, has several significant drawbacks as well.<sup>2</sup> It does not differentially weight certain scar characteristics, includes some unclear item definitions, and fails to include scar spread, perhaps the most important measure of linear scar quality. It also demonstrated significant intrarater variability between in-person and photographic assessments.

Other disadvantages of previously developed scales include the nebulousness of some of the terminology used; clinicians, and even those adept at rating scars, may not be familiar with the differences between “pliability,” “relief,” and “thickness,” yet all 3 of these items are included in the POSAS. Similar challenges exist in other scales as well.

By including a separate rating for “desirable scar,” the SCAR scale is able to take into account myriad subtle changes that affect scar quality, such as involvement of multiple cosmetic subunits, without including a large number of separate items. This allows the SCAR scale to capture the overall cosmetic and functional outcome and improves its validity while not detracting from its feasibility.

Another strength of this scale is the incorporation of a multidisciplinary team in its development, because a group including specialists from dermatology, plastic surgery, surgical oncology, physiatry, and emergency medicine was involved in the validation and reliability assessments. Of note, the SCAR scale was not developed in the context of a particular clinical trial, which is helpful because some authors have questioned whether instruments developed for a particular trial may bias its outcomes.<sup>17,23</sup> Another strength of the SCAR scale is its use of separate scores for erythema and dyspigmentation, which should allow investigators to capture improvement in 1 or both components as scars evolve over time or are treated with adjuvant approaches.<sup>12,24</sup>

### Limitations

This study has several limitations. Numerous other outcome measures exist for linear scars, and some—such as VSS and POSAS—have been used widely in a variety of settings. Still, these preexisting scales were not initially developed for linear postoperative scars, and therefore some vestiges of their prior incarnation as burn scar assessment tools persist. Moreover, in the POSAS the large number of measures all sharing a 10 item choice response lends itself to several important biases, such as central tendency (avoidance of the

extremes of response), positive skew (satisfaction or acceptability questions tend to elicit more positive responses), underlying cause (an overall poor scar may receive low assessments on all components), and others.<sup>25</sup> The value of an alternative outcome measure therefore should be seen within this context.

Other limitations of this study include the overall generalizability of the raters' responses (although including 5 raters from varying clinical backgrounds helps mitigate this concern) and whether the high reliability of photographic equivalency found in this study would generalize to the non-macrophotographs occasionally used in clinical practice. Another possible limitation is the number of raters and scars included in the study, although the number of raters and assessed scars in this reliability study are higher than those used in most previously developed rating scales.<sup>1</sup> The SCAR scale also does not include objective measures of the physical properties of the scar because requiring specialized measurement equipment for color, thickness, and other components would drastically reduce the

feasibility of this scale and would make it inappropriate for use as a clinical (as opposed to research) outcome measure. The SCAR scale also does not address functional impairment and the psychosocial impact of scars, although both of these are very infrequently seen in postoperative linear scars.

## Conclusions

The SCAR scale is a feasible and reliable scar assessment instrument. Further studies may help delineate ideal cutpoints for acceptable, desirable, and unacceptable scarring, as well as explore its responsiveness, or sensitivity to change over time. The SCAR scale provides a unique combination of an outcome measure designed for linear scars that may be used by examining photographs, rather than live patients, and that may be completed in less than 30 seconds by most raters. These attributes contribute to its potential use as a tool in both daily practice and clinical research.

### ARTICLE INFORMATION

**Accepted for Publication:** August 16, 2016.

**Published Online:** November 2, 2016.  
doi:10.1001/jamadermatol.2016.3757

**Conflict of Interest Disclosures:** None reported.

**Correction:** This article was corrected online January 11, 2017, to fix a typographical error in Table 1.

**Additional Contributions:** I thank the reliability and validity committee: Michael J. Dans, MD, PhD; Alan Durkin, MD; George Evangelou, MD; Yoram D. Gutfreund, MD; Adam Z. Hammer, MD; David T. Harvey, MD; Giorgos Karakousis, MD; Mark Tang, MD; and Kristine A. Waters, PA-C. No compensation was provided.

### REFERENCES

- Vercelli S, Ferriero G, Sartorio F, Cisari C, Bravini E. Clinimetric properties and clinical utility in rehabilitation of postsurgical scar rating scales: a systematic review. *Int J Rehabil Res.* 2015;38(4):279-286.
- Vercelli S, Ferriero G, Sartorio F, Stissi V, Franchignoni F. How to assess postsurgical scars: a review of outcome measures. *Disabil Rehabil.* 2009;31(25):2055-2063.
- Durani P, McGrouther DA, Ferguson MW. Current scales for assessing human scarring: a review. *J Plast Reconstr Aesthet Surg.* 2009;62(6):713-720.
- Nguyen TA, Feldstein SI, Shumaker PR, Krakowski AC. A review of scar assessment scales. *Semin Cutan Med Surg.* 2015;34(1):28-36.
- Sullivan T, Smith J, Kermodie J, McIver E, Courtemanche DJ. Rating the burn scar. *J Burn Care Rehabil.* 1990;11(3):256-260.
- Draaijers LJ, Tempelman FR, Botman YA, et al. The patient and observer scar assessment scale: a reliable and feasible tool for scar evaluation. *Plast Reconstr Surg.* 2004;113(7):1960-1965.
- 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(2):514-522.
- Crowe JM, Simpson K, Johnson W, Allen J. Reliability of photographic analysis in determining change in scar appearance. *J Burn Care Rehabil.* 1998;19(2):183-186.
- Yeong EK, Mann R, Engrav LH, et al. Improved burn scar assessment with use of a new scar-rating scale. *J Burn Care Rehabil.* 1997;18(4):353-355.
- Beausang E, Floyd H, Dunn KW, Orton CI, Ferguson MW. A new quantitative scale for clinical scar assessment. *Plast Reconstr Surg.* 1998;102(6):1954-1961.
- Singer AJ, Arora B, Dagum A, Valentine S, Hollander JE. Development and validation of a novel scar evaluation scale. *Plast Reconstr Surg.* 2007;120(7):1892-1897.
- Hultman CS, Friedstat JS, Edkins RE, Cairns BA, Meyer AA. Laser resurfacing and remodeling of hypertrophic burn scars: the results of a large, prospective, before-after cohort study, with long-term follow-up. *Ann Surg.* 2014;260(3):519-529.
- Maher IA, Fosko S, Alam M. Experience vs experiments with the purse-string closure: unexpected results. *JAMA Dermatol.* 2015;151(3):259-260.
- Kantor J. The SCAR scale (Scar Cosmesis Assessment and Rating scale): development and validation of a new outcome measure for postoperative scar assessment [published online June 13, 2016]. *Br J Dermatol.* doi:10.1111/bjd.14812
- Kantor J. Utilizing the Patient Attitudes to Scarring Scale (PASS) to develop an outcome measure for postoperative scarring: a study in 430 patients. *J Am Acad Dermatol.* 2016;74(6):1280-1281.e2.
- Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002;11(3):193-205.
- Albrecht J, Taylor L, Berlin JA, et al. The CLASI (Cutaneous Lupus Erythematosus Disease Area and Severity Index): an outcome instrument for cutaneous lupus erythematosus. *J Invest Dermatol.* 2005;125(5):889-894.
- Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas.* 1973;33:613-619. doi:10.1177/001316447303300309
- Sullivan GM, Artino AR Jr. Analyzing and interpreting data from Likert-type scales. *J Grad Med Educ.* 2013;5(4):541-542.
- van der Wal MB, van de Kar AL, Tuinebreijer WE, et al. The modified patient and observer scar assessment scale: a novel approach to defining pathologic and nonpathologic scarring? *Plast Reconstr Surg.* 2012;129(1):172e-174e.
- Robinson A, Kardos M, Kimball AB. Physician Global Assessment (PGA) and Psoriasis Area and Severity Index (PASI): why do both? a systematic analysis of randomized controlled trials of biologic agents for moderate to severe plaque psoriasis. *J Am Acad Dermatol.* 2012;66(3):369-375.
- Baryza MJ, Baryza GA. The Vancouver Scar Scale: an administration tool and its interrater reliability. *J Burn Care Rehabil.* 1995;16(5):535-538.
- Charman C, Williams H. Outcome measures of disease severity in atopic eczema. *Arch Dermatol.* 2000;136(6):763-769.
- Jared Christophel J, Elm C, Endrizzi BT, Hilger PA, Zelikson B. A randomized controlled trial of fractional laser therapy and dermabrasion for scar resurfacing. *Dermatol Surg.* 2012;38(4):595-602.
- Choi BC, Pak AW. A catalog of biases in questionnaires. *Prev Chronic Dis.* 2005;2(1):A13.