Evidence-based and clinical views on acute wound healing and scar formation
Brölmann, Fleur

Citation for published version (APA):
Brölmann, F. E. (2013). Evidence-based and clinical views on acute wound healing and scar formation
Are digital photographs reliable to assess donor site scars? An inter-method analysis and validity testing

F.E. Brölmann
A.M. Eskes
A.L. van de Kar
C.M.A.M. van der Horst
F.B. Niessen
R. Lindeboom
H. Vermeulen
D.T. Ubbink

Adapted from short communication:
Chapter 7

Abstract

Background

The Observer Scar Assessment Scale (OSAS) is used to judge scars in vivo and on digital photographs. It is questionable whether these different methods influence the results of the scar assessment.

The aim of this study was to assess the inter-method reliability, score agreement and validity testing of in vivo and digital photographic assessments of donor site scars.

Methods

Donor site scars of 119 patients were investigated. Three months after complete epithelialization of the donor site the six items of the OSAS were rated in vivo and from digital photographs. Intra-class correlation coefficients (ICC) were calculated to judge inter-method reliability. Score agreement for each item was assessed using Bland-Altman plots. Validity of the digital photographic assessment was tested by comparing the OSAS outcomes of six wound dressings evaluated in a trial using the in vivo judgements as a reference standard.

Results

Reliability was moderate at best for the ‘total observer score’ (ICC 0.47) and ‘pigmentation’ (ICC 0.45). Bland-Altman plots showed wide limits of agreement between the in vivo and photographic assessment for ‘vascularity’, ‘pigmentation’, ‘overall opinion’ and ‘total observer scale’. The item ‘pigmentation’ as assessed on digital photographs received a lower score than the in vivo assessment (p=0.027, Mann-Whitney U test).

Conclusions

Reliability between digital photographs and in vivo assessment of donor site scars is limited when using the OSAS. Correspondingly, both methods result in different scar judgements of donor site scars. Therefore, in clinical practice, in vivo assessment of donor site scars should remain the policy of choice.
**Introduction**

Donor site wounds (DSW) after split skin grafting can be a considerable burden to patients, not only during, but also after the healing process.\(^1\)-\(^3\) Often patients endure a wide range of problems related to their donor site scar, varying from aberrant colour and itching to overall quality of life issues.\(^4\)-\(^8\) Moreover Cubison et al. found that prolonged healing time increases the risk of hypertrophic scarring.\(^9\) Consequently the interest for scar research, prevention, monitoring and treatment is rising.\(^10\)-\(^14\)

For research purposes, scar quality can be assessed in vivo, i.e. real life, or from digital photographs, each with its pros and cons. Conventional outpatient clinics are suitable for monitoring DSWs and donor site scars through in vivo observations. However, these visits may not be worthwhile when its sole purpose is wound inspection and may be time-consuming for both patients and healthcare professionals, and cumbersome for the patient due to the wound that required the transplantation.

As an alternative, digital photography or, in a broader sense, telemedicine have been proposed as a potential (consultation) tool in the management of donor site scars, as they accommodate the visual nature of skin examination.\(^15\)-\(^17\) Digital photographs have been successfully used for repeated wound size measurements in (randomised) clinical trial design\(^18\)-\(^20\), and inter-observer studies\(^21\)-\(^22\). The advantages of digital photographs are their high-quality colour images, which can be easily captured and transferred to a computer, analysed, stored, printed or sent by email. This enables communication across the disciplines involved, which is particularly valuable in wound care due to its multidisciplinary nature. Furthermore, these photographs can be stored as part of an electronic patient record and used in a telemedicine system for shared patient management by community nurses, specialised wound nurses and plastic surgeons.

If digital photographs are used for a uniform judgment and management of donor site scars, valid scar assessment tools are needed. The Patient and Observer Scar Assessment Scale (POSAS) is such a validated tool for the judgment of various types of scars, i.e. burn- and linear scars.\(^23\)-\(^24\) The POSAS is a subjective, consistent and feasible tool\(^23,25\), which can be used reliably even by inexperienced observers\(^6\).

However, little is known about the observer agreement of digital photographs versus in vivo observation when monitoring donor site scars.\(^17\),\(^26\) If we decide to rely on photographs in the management of DSWs, we need to know whether these are as reliable and valid as in vivo judgements. From contiguous research on wound assessment it is known that the inter-observer reproducibility of wound measurements made by using
digital photographs can be as accurate as contact tracing.\textsuperscript{27} Others emphasize that lack of live assessment fails to evaluate the wound from multiple visual angles and under various lighting conditions.\textsuperscript{28} Furthermore, it does not allow palpation of the wound, which makes three-dimensional defects less obvious.\textsuperscript{28} In current research the observer scale of the POSAS, also called the Observer Scar Assessment Scale (OSAS) is used for in vivo and digital photographs as well.\textsuperscript{29-30} However, it is questionable whether these different settings influence the results of the scar judgment and eventually treatment policy.

Therefore, the aim of this study was to assess the inter-method reliability and score agreement of in vivo and digital photographic assessments with the OSAS and validity testing of the digital photographic observations for healed DSWs.

\section*{Patients and methods}

\subsection*{Patients and equipment}

For this study we used data from a multicentre randomised clinical trial regarding the treatment of donor site wounds (REMBRANDT trial, registered as NTR1849). In this trial 289 adult patients, who underwent split skin grafting leaving a donor site area of at least 10 cm\textsuperscript{2}, were included to evaluate the effectiveness of six different wound dressing materials (alginate, film, gauze, hydrocolloid, hydrofibres or silicone). The REMBRANDT trial and its study protocol have been described in detail previously.\textsuperscript{31-32} All patients gave written informed consent for the REMBRANDT trial and this inter-method analysis. This study was approved by the medical ethics review boards of the contributing centres.

For this study we selected all patients with complete follow-up data regarding their scar evaluation, conducted three months after complete epithelialisation of the DSW. Complete wound healing was defined as re-epithelialisation of the total wound surface without remaining scabs.

Photographs were taken with digital cameras, with a minimal resolution of five megapixels, in the outpatient clinics or during home visits, by a total of eleven medical doctors, specialized wound nurses, surgical nurses, or wound researchers from five centres. The images were stored in JPEG format in the camera and transferred to a personal computer. Those who took the pictures were also the in vivo observers. Two pictures were taken; one with and one without using flashlight.
Chapter 7

Observers and scar assessment

The eleven healthcare professionals performed the in vivo scar judgment independently, using the OSAS. Furthermore, they recorded skin type by means of the Fitzpatrick skin classification, a numerical classification for the colour of skin. Their judgments and photographs were then submitted to the study coordinators (FB, AE). Subsequently, one plastic surgeon (AK), with extensive experience with the POSAS to judge donor sites, independently judged the digital photographs of the donor site scars using the observer scale and recorded the skin type.

The observer scale contains six items, including ‘vascularity’, ‘pigmentation’, ‘pliability’, ‘thickness’, ‘relief’ and ‘surface area’. Each item is scored numerically, ranging from one (best possible outcome) to ten (worst possible outcome), and results in a ‘total score’ of the OSAS. In addition, an ‘overall opinion’ on the dissatisfaction with the cosmetic appearance of the scar was given, also on a scale of one to ten, where ten corresponds with the worst imaginable scar.

Data gathering and analysis

Data and data entry were cross-checked by two investigators independently (AE and FB). Basic demographic data of the patients contained age, sex, smoking status, indication for the split-skin grafting, location of the DSW, thickness of the graft, and skin type. To assess the inter-method reliability of the OSAS scores obtained from digital photographs and in vivo judgments, intra-class correlation coefficients (ICC), including their 95% confidence intervals (CI) were calculated. ICCs of the separate OSAS items, ‘total scores’ and ‘overall opinions’ were calculated using the one-way random single measures model. We interpreted an ICC above 0.8 as ‘very good’, between 0.8 and 0.6 as ‘good’, between 0.6 and 0.4 ‘moderate’, below 0.4 ‘poor’ and negative values mean no reliability at all.

Score agreement was further assessed using the method of Bland&Altman, which plots the score differences between in vivo and photographic judgments against their averages. This plotting technique allows detection of both random and systematic differences across the range of values measured. Bland and Altman suggested the calculation of “limits of agreement” to indicate the absolute magnitude of the deviation scores. When the measurement error is random, 95% of the deviation scores is expected to fall within ± 1.96 standard deviations. As the OSAS scale for each item varies from one to ten, we considered arbitrarily a two-point variation (i.e. 20% of the total scale) as acceptable. Similarly, we considered a maximum of 20% of the total OSAS scale, i.e. an 11-point difference, as an acceptable variation for the total OSAS score, which varies from six to 60.
Chapter 7

After establishing the inter-method reliability, validity testing of the digital photograph assessment was performed by comparing the OSAS results in vivo with the digital photographs using the in vivo observations as a reference standard. We hypothesized that a (very) good reliability between both methods would result in similar effectiveness outcomes of the wound dressing materials used in the REMBRANDT trial. The differences between the OSAS scores were analysed using the Mann-Whitney U test due to their non-normal distribution. Data analysis was performed using SPSS v. 18 (IBM, Armonk, NY, USA).

Results

Patients

A total of 119 patients with complete follow-up data regarding their scar evaluation was investigated. They had a mean age of 60 years (SD 16.1, range 18-90), 81 (68.6%) of whom were males (Table 1). Patients needed a split-skin grafting mainly for surgical or traumatic wounds (66.4%). The graft had a median thickness of 0.30 millimetres and was harvested usually from the upper thigh. The in vivo and photograph observers classified

| Table 1. Patient and peri-operative characteristics of the 119 patients investigated |
|---------------------------------|--------|--------|
| Mean age ± SD, years          | 60.0 ± 16.2 |
| Males, n (%)                  | 81 (68.1)  |
| Smokers, n (%)                | 36 (30.3)   |
| Indication for SSG, n (%)     |          |
| - Chronic wound               | 28 (23.5)  |
| - Burn wound                  | 3 (2.5)    |
| - Surgical/traumatic wound    | 79 (66.4)  |
| - Tumour excision             | 7 (5.9)    |
| - Other                       | 2 (1.7)    |
| Location of the DSW, n (%)    |          |
| - Thigh                       | 115 (96.6)|
| - Other                       | 4 (3.3)    |
| Median thickness of graft, mm (range) | 0.30 (0.10-0.80) |

<table>
<thead>
<tr>
<th>Fitzpatrick skin classification</th>
<th>In vivo</th>
<th>Photograph</th>
</tr>
</thead>
<tbody>
<tr>
<td>- I</td>
<td>14 (11.8)</td>
<td>-</td>
</tr>
<tr>
<td>- II</td>
<td>65 (54.6)</td>
<td>102 (85.7)</td>
</tr>
<tr>
<td>- III</td>
<td>22 (18.5)</td>
<td>9 (7.6)</td>
</tr>
<tr>
<td>- IV</td>
<td>6 (5.0)</td>
<td>-</td>
</tr>
<tr>
<td>- V</td>
<td>1 (0.8)</td>
<td>2 (1.7)</td>
</tr>
<tr>
<td>- VI</td>
<td>7 (5.8)</td>
<td>6 (5.0)</td>
</tr>
</tbody>
</table>

SD, Standard Deviation; SSG, Split Skin Graft; DSW, Donor Site Wound; mm, millimetres.
the majority of the patients as Fitzpatrick skin type II (white skin with blond or red hair; blue, green or hazel eyes). However, the Fitzpatrick skin classification scores showed a wider range when assessed in vivo than using the photographs (Table 1).

Reliability and agreement on OSAS items and overall opinion

Table 2 shows the reliability of judging photographs using the observer scale items, the ‘total OSAS score’ and the ‘overall opinion’ as compared to in vivo judgments. A moderate reliability was observed for the ‘total OSAS score’ (ICC 0.47, 95% CI 0.32-0.60) and ‘pigmentation’ (ICC 0.45, 95% CI 0.29-0.58). Other items scored poorly or, in case of ‘surface area’, there was no reliability at all (i.e., a negative ICC was observed), as shown in Table 2.

Table 2. Inter-method reliability between digital photographs and in vivo assessment of donor site scars, expressed as interclass correlation coefficients for the OSAS items using a one-way random model (single measurements)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Inter-method reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MD</td>
</tr>
<tr>
<td>Total OSAS score</td>
<td>-1.6</td>
</tr>
<tr>
<td>Vascularity</td>
<td>-0.9</td>
</tr>
<tr>
<td>Pigmentation</td>
<td>-0.6</td>
</tr>
<tr>
<td>Thickness</td>
<td>0.02</td>
</tr>
<tr>
<td>Relief</td>
<td>-0.3</td>
</tr>
<tr>
<td>Pliability</td>
<td>0.1</td>
</tr>
<tr>
<td>Surface area</td>
<td>0.1</td>
</tr>
<tr>
<td>Overall opinion</td>
<td>-1.3</td>
</tr>
</tbody>
</table>

MD, mean difference between in vivo and photographic OSAS scores; SD, Standard Deviation; ICC, Intra-class Correlation Coefficient; CI, Confidence Interval; OSAS, Observer Scar Assessment Scale.

Scores of the following items (‘thickness’, ‘relief’, ‘pliability’ and ‘surface area’) showed nearly always a single value for the in vivo assessments, which precluded the production of Bland-Altman plots. The remaining items; ‘total OSAS score’, ‘vascularity’, ‘pigmentation’ and ‘overall opinion’, were analysed using the Bland-Altman technique (Figure 1). Total OSAS scores ranged between six and 23. The in vivo judgments led to a systematically lower OSAS score, which means better scar quality, for most items investigated (‘total OSAS score’, ‘vascularity’, ‘pigmentation’, ‘relief’ and ‘overall opinion’). Also, the differences between the in vivo and photograph judgments of the items ‘vascularity’, ‘pigmentation’ and ‘overall opinion’ were larger than four and, thus, crossed the predefined limits. This is in line with the poor to moderate agreement regarding each of these items. However, for the ‘overall OSAS score’ the differences were within our predefined limits.
Chapter 7

Validation of the OSAS based on digital photographs

In vivo scar assessment scores did not show significant differences among the six wound dressing materials compared in the REMBRANDT trial on any OSAS item or total OSAS score. This was in contrast with the photographic scar assessment, which showed that patients who received alginate dressings scored lower on the OSAS item 'pigmentation' (p=0.027, Mann-Whitney U test) compared to the other treatment arms taken together.

Discussion

The results of this study show that OSAS assessed from photographs are not a valid tool to judge donor site scars, based on the limited reliability and agreement between digital photographs and in vivo assessment using the OSAS for donor site scars. The inter-method reliability was moderate at best on only two of the six items, 'pigmentation' and the 'total OSAS score'. As expected, considering the characteristics of a photograph,
we found hardly any agreement as to ‘thickness’, ‘relief’, ‘pliability’ and ‘surface area’. Overall, the judgment of scars in vivo and on digital photographs appeared to disagree, resulting in a discrepancy in the judgment of scar quality.

To our knowledge the OSAS has not been validated when using digital photographs, yet they are interchangeably used for scar assessment in post-surgical and DSWs. Because in vivo agreement among caregivers has shown to be valid and reliable, the (P)OSAS is generally accepted as a scar assessment tool in clinical practice and is rationally used as an outcome measure for scar formation research.

Our validation analysis showed that the lack of agreement can lead to a different perception of the clinical outcome and consequently to a different patient satisfaction. If digital photographs were used in the REMBRANDT trial, alginate would have shown better results based on the OSAS score than the in vivo judgments. We assumed that the in vivo judgment, as reference standard, is superior to the judgement of photographs because the OSAS has been validated in vivo. In any case, both methods are not interchangeable and the consequences of differences in judgment could result in different research results and possibly different clinical choices for wound dressing regimes.

Previous studies, using different (standardised) methods of photographic assessment, report contradicting evidence as to the use of digital photographs in wound care. This makes the value of digital photographs debatable, however popular they may be for, for instance, tele-dermatology. Aggregated evidence on the merits of tele-dermatology for various skin conditions shows this remote practice has an inferior diagnostic accuracy. Moreover, limited evidence exists regarding tele-dermatology as to clinical outcomes and management compared to clinical dermatologists.

Other studies have shown good inter- and intra-observer reliability for the use of digital photographs. Only one study compared the wound area of chronic wounds as assessed by digital photographs and by means of contact tracing and found no significant differences. These contradicting findings may be due to differences in study population, method of assessing images, or study design.

Four out of the six items of the observer scale; i.e. ‘thickness’, ‘relief’, ‘pliability’ and ‘surface area’, had a low reliability because these are fairly difficult to assess from digital photographs, irrespective of the quality of the images. Because these items contribute substantially to the total observer scale judgement, it is remarkable that the observer scale has not been validated for digital photographs in previous studies before its acceptance in the clinical setting.
Some limitations of our study should be mentioned. First, the photographs used for assessing donor site scars were taken with different digital cameras in different settings (e.g. in-hospitals, outpatient clinics, nursing homes, and patients’ homes). Although we believe that today’s modern digital cameras as used in this study warrant auto-focused, high-resolution pictures with low varying quality, it still may have affected observer scale judgement. Also, standardization of the photographic equipment could have reduced the generalizability of our study results.

Secondly, the OSAS scores when judged from the digital photographs were systematically lower than when judged in vivo. This could be due to a lower quality of the digital photographs or to the fact these pictures were judged by a single observer (AK). Whichever reason may be true, the possible risks of this discrepancy as mentioned above remain lurking. The digital photographs were judged by one observer (AK), which could have influenced the reliability of the results. Yet, the proven inter- and intra-observer reliability of the POSAS tool has been the reason to choose for this study design.6

Finally, we used a sample of 119 patients with complete follow-up data regarding their scar evaluation. Previous studies used samples varied from 20 to 80 patients for inter- and intra observer analysis and validity testing.12,30,41,42 In order to achieve more precise results we used a sample of more than 100 patients as advised by the COSMIN group.43 We were able to confirm the precision of our results by means of a post-hoc power analysis. This analysis showed that the 95% CI would range from 0.36 to 0.64 given an expected ICC of 0.50. This seems a relatively small range, as it would not have changed the conclusion of our study that the agreement is ‘moderate’ at best.

Further research with the OSAS should consider the poor inter-method reliability of digital photographs for scar assessment of donor site scars and, consequently, its failing ability to measure what it is supposed to measure. For other types of scars, reliability and validity testing should be performed before relying on digital photographs.

Our findings suggest that digital photographs are no valid substitute for in vivo judgment of donor site scars using the OSAS. Hence, in this case a picture is not worth more than a thousand words. Therefore, in clinical practice, in vivo assessment of donor site scars should remain the policy of choice. Research findings based on digital images may be inconsistent or invalid, and clinically, this could result in different treatment decisions for various scars. Because these conclusions were drawn merely from donor site scars, similar questions remain to be answered for other types of wound.
Acknowledgement

We owe our gratitude to the patients, specialised wound care nurses, and surgical nurses collaborating in the REMBRANDT study for their contribution to the assessment of the donor site scars. Furthermore, we would like to thank the REMBRANDT study group for including the patients in the REMBRANDT trial.
References

Digital photographs to assess donor site scars


