Digital image analysis versus clinical assessment of wound epithelialization: a validation study

Monica C.T. Bloemen
Bouke K.H.L. Boekema
Marcel Vlig
Paul P.M. van Zuijlen
Esther Middelkoop
Abstract

Introduction: To evaluate the progress in wound healing, wound assessment is mandatory. Epithelialization is traditionally assessed subjectively by the clinician. In a previous study, subjective evaluation of epithelialization was shown to be reliable. Aim of this study was to investigate validity of this subjective evaluation. Therefore, reliability of epithelialization measured by digital image analysis was tested in order to analyze the correlation between the subjective evaluation and the digital image analysis.

Methods: Clinicians assessed epithelialization in 50 burn wounds that were treated with a split-skin graft. Epithelialization of these wounds was also measured by three observers using digital image analysis. Reliability of digital image analysis was tested using the intraclass correlation (IC). To test validity, subjective clinical assessment was correlated with digital image analysis (IC).

Results: Interobserver reliability of epithelialization measured by digital image analysis was good (IC coefficient 0.74). Subjective clinical assessment of epithelialization showed a strong correlation with digital image analysis (IC coefficient 0.80).

Conclusions: Subjective clinical evaluation of wound epithelialization is highly correlated with the assessment by digital image analysis. Since digital image analysis is more time-consuming, we consider the subjective wound assessment as a valid technique for daily practice.
Introduction

Implementation of evidence-based medicine requires valid and reliable tools for assessment of different wound healing parameters. It not only allows clinicians to follow progression of wound healing in an individual patient; wound assessment is also necessary to evaluate the effectiveness of a specific treatment, especially in clinical trials. In the majority of the reported studies on wound healing, a subjective wound assessment is used, usually performed by one or more clinicians. Other subjective methods have been described to estimate wound healing, such as qualitative index scores and scales. However, also quantitative and objective measurements are preferred when comparing the effects of different wound treatments. Therefore, methods such as ruler-based measurements, tracings, and more modern technologies have been developed, e.g. computerized planimetry, digital image analysis, three-dimensional laser imaging.

A relevant parameter for the evaluation of wound healing, especially in wounds treated with a meshed split-skin graft, is the rate of epithelialization. Although much research has been performed on other aspects of the wound, e.g. wound size, there is a paucity of research on the assessment of epithelialization. Some reports describe the measurement of epithelialization by means of photoplanimetry, optical coherence tomography, or measuring water evaporation. However, these techniques were not evaluated sufficiently on the required clinimetric properties reliability, validity, and feasibility. Epithelialization has also been evaluated histologically by means of a biopsy. This method requires an invasive action and can cause pain, infection, and scarring. Objective tools which are feasible appear to be scarce. For this reason, we are still searching for methods to determine the true value of epithelialization as good as possible. The most important and optimal evaluation of epithelialization still seems to be the clinician’s subjective assessment.

In a previous study, we investigated the reliability of subjective assessment. The intra- and interobserver reliability of subjective assessment of epithelialization was shown to be good when performed by an experienced observer. However, the validity of the subjective clinical assessment was not tested, which means it is not known if the measurements of the experienced observer represent the true rate of epithelialization. In this study, we investigated the validity of the subjective clinical assessment of epithelialization. In order to do this, the evaluation of epithelialization performed by an experienced clinician was compared with epithelialization measured by a computerized technique, i.e. digital image analysis. Preceding the investigation of the validity of the subjective assessment, the interobserver reliability of digital image analysis was tested, as this was not performed previously for this parameter.
**Methods**

*Study Design*

The wounds assessed in this study were originating from consecutive burn patients treated in an ongoing multicenter randomized controlled trial (RCT). This trial, performed in the three Dutch Burn Centers, investigates dermal substitution in combination with topical negative pressure in the treatment of acute burns that require skin grafting. The study protocol was approved by the medical ethics committee (M07-035) and registered at Clinical Trials (ID NCT00548314). Intraoperatively, patients (with informed consent) were allocated to the following four treatment groups: 1. dermal substitute in combination with a split-skin graft (SSG) and topical negative pressure; 2. dermal substitute in combination with a SSG; 3. SSG and topical negative pressure; 4. SSG alone. All skin grafts were expanded with an 1:1.5 ratio. Four to seven days after skin transplantation, each patient was visually assessed by the treating clinician. Due to the multicenter setting, this bedside assessment was performed by different experienced clinicians. Several wound parameters, such as the percentage of epithelialization, were documented and a digital photograph of the wound was taken. Photographs were obtained with commercially available digital cameras and were judged to be of high quality, i.e. in focus, showing the relevant wound area, perpendicular to the center point of the wound, and of high enough resolution to be able to judge wound aspects in detail. The digital photographs of the first 50 patients of the above mentioned clinical trial were selected for the computer-aided image analysis. This assessment, in which the percentage of epithelialization was determined, was performed by three independent observers. Figure 1 illustrates an overview of the study design.

*Measurements with Digital Image Analysis*

Macroscopic quantification of the wounds was achieved using the computer-assisted image analysis software NIS-Elements Ar (Nikon Instruments Europe B.V., Amstelveen, The Netherlands). This semi-automatic software can analyze a digital photograph and measure the number of pixels of the selected area. Independent of each other, the observers first marked the total wound area. The total wound area was the area of the initial burn wound that was transplanted with a SSG. Secondly, the observer manually marked the parts of the wound that were not epithelialized or were not covered with a vital SSG, termed open wound area. The marked total and open wound area were calculated and presented by the number of pixels. Wound parameter epithelialization was defined as the percentage of the wound with a vital skin graft and healed graft interstices. Accordingly, necrotic or granulation tissue was interpreted as non-epithelialized. For each wound, the percentage of wound epithelialization was calculated using the following formula: $1 - \frac{\text{pixels open wound}}{\text{total wound area}}$. 

94
area/ pixels total wound area) X 100%. Figure 2a-d shows an example of a wound analyzed by means of digital image analysis and Table 1 shows the digital and subjective data of this wound.

**Figure 1: Overview of the study design**

Statistical Analysis
Data were analyzed with the statistical program SPSS for Windows 18.0 (SPSS Inc., Chicago, USA). The interobserver reliability of digital image analysis measures the agreement between several observers. Analysis was performed by means of the intraclass correlation with a 95% confidence interval. The two-way-random effect model and the absolute agreement type for a single and average measurement were selected for the calculations of the intraclass correlation coefficient (ICC)\(^2^3\). This calculation provides a single measure ICC which is based on a single measurement and an average measure ICC based on the average measurements of all observers. An ICC below 0.4 represents poor agreement, values above 0.75 represent excellent agreement and values between 0.4 and 0.75 stand for fair to good agreement\(^2^4\). The standard error of measurement (SE\(_{\text{meas}}\)) was used to calculate the number of errors between measurements. The coefficient of variation (CV) was calculated using the following formula: \(CV = \frac{SE_{\text{meas}}}{\text{mean}} \times 100\). A low CV percentage represents less variation in error than a high CV. To analyze validity of the subjective epithelialization assessment, this score was compared to the measurements of the digital image analysis, by means of the ICC. The ICC was chosen for this analysis as it reflects both the degree of correspondence and agreement among the two measurement techniques.
Figure 2: Example of a measurement with digital image analysis

A, burn wound with overlying split-skin graft; B, C, D, transplanted burn wound 5 days postoperatively; marked in grey the ‘total wound area’; marked in yellow the ‘open wound area’.

Table 1: Digital image analysis measurements of a wound by three observers and the subjective clinical assessment

<table>
<thead>
<tr>
<th>Patient example</th>
<th>Wound area</th>
<th>Open wound area</th>
<th>Epithelialization (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer 1</td>
<td>2240000 pixels</td>
<td>220483 pixels</td>
<td>90.2</td>
</tr>
<tr>
<td>Observer 2</td>
<td>2201954 pixels</td>
<td>216022 pixels</td>
<td>90.2</td>
</tr>
<tr>
<td>Observer 3</td>
<td>2154363 pixels</td>
<td>286380 pixels</td>
<td>86.7</td>
</tr>
<tr>
<td>Mean score observers</td>
<td>2198772 pixels</td>
<td>240961 pixels</td>
<td>89.0</td>
</tr>
<tr>
<td>Subjective clinical assessment</td>
<td>-</td>
<td>-</td>
<td>90.0</td>
</tr>
</tbody>
</table>

Data were obtained from the wound shown in Figure 2.
Results
The first 50 patients of the above mentioned RCT were treated in the burn centers of Beverwijk and Rotterdam in the Netherlands and were used for this clinimetrical study. The mean percentage of epithelialization evaluated during a bedside procedure by the clinician was 88.0% ± 13.5. The mean percentage of epithelialization measured with digital image analysis by three observers was 90.0% ± 10.0.

Reliability of Digital Image Analysis
The interobserver reliability of epithelialization measured with digital image analysis was analyzed using the measurements of three observers. The single and average measure ICCs with 95% confidence interval were respectively 0.48 (0.32 - 0.64) and 0.74 (0.58 - 0.84), which represents fair to good interobserver reliability. The coefficient of variation was 9.9%.

Validity of Subjective Clinical Assessment
A good to excellent agreement was found between the percentage of epithelialization assessed during a bedside procedure and the percentage of epithelialization measured with digital image analysis (single measure ICC 0.67 (0.49 - 0.80), average measure ICC 0.80 (0.66 - 0.89), 95% confidence interval).

Discussion
In evidence-based medicine, valid and reliable tools for the assessment of wound healing parameters are crucial, not only to record the progress of healing, also to compare the effect of applied treatments. The subjective wound assessment is still most frequently used in clinical practice. For this reason, reliability and validity of this assessment should be examined. In clinimetric research, reliability of measurement methods is investigated first, after which validity is tested. Subjective assessment was already shown to be reliable. In this study, we investigated the validity of the subjective assessment of epithelialization. As there is no standard tool to assess epithelialization, we considered assessment with digital image analysis to be the most valid tool. Therefore, the subjective results of epithelialization were compared with digital image analysis in order to test the validity.

An excellent validity, indicated by the high average ICC value, was found for the subjective bedside epithelialization score when compared with the data of the digital image analysis; the two wound evaluation methods were shown to give comparable results. Therefore, we consider the subjective clinical assessment of epithelialization to be a valid tool. In this study design, wounds were evaluated by a
variable clinician. The correlation between the two assessments would have even be higher, if bedside assessment was performed by the same clinician in all cases. The results of this study can be compared with results of several other studies. First of all, Mekkes et al. studied venous ulcers and showed similar results in their comparison of the subjective scores with the measurements of digital image analysis. However, this system was designed to measure red granulation tissue, and yellow and black necrotic tissue based on a color analysis\textsuperscript{25}. In a study of Hauser et al., the subjective evaluation of epithelialization of two observers was correlated with photoplanimetric assessment\textsuperscript{17}. The subjective assessment appeared to have a good correlation with the measurements of the photoplanimetry, however statistical analysis was not provided. Finally, visual assessment was compared with computerized planimetry to measure fibrin percentage\textsuperscript{26}. Average visual estimations were very similar to the computerized planimetry, therefore bedside evaluation of fibrin percentage was considered to be reliable\textsuperscript{26}. Although our study only investigated the parameter wound epithelialization, it is plausible that subjective assessment is also appropriate for the evaluation of other wound parameters, such as the percentage of eschar, granulation or necrotic tissue.

Preceding the correlation of the subjective bedside assessment with the scores obtained by digital image analysis, we tested the reliability of the digital image analysis. Our data show that interobserver ICC values of parameter epithelialization measured by digital image analysis were good, therefore we can safely assume the intraobserver reliability is good as well. Intraobserver reliability is expected to be higher than the interobserver reliability, as there is less variability within one observer\textsuperscript{27}. Therefore, in this study only interobserver reliability was tested. In the majority of the reported studies, reliability of other wound parameters and wound types were investigated. Nevertheless, some studies showed comparable results to ours. In one study, a high agreement was found between observers assessing the percentage of necrosis and granulation tissue using digital image analysis\textsuperscript{25}. In addition, Laplaud et al. investigated the interobserver reliability of computerized planimetry for measuring fibrin percentage and also found high ICC scores\textsuperscript{26}.

Fortunately, this study implies that the subjective clinical assessment of epithelialization can be used for clinical purposes, given that assessment with digital image analysis has some shortcomings. Even digital image analysis of epithelialization rate requires some involvement of observers and cannot be entirely automatic; the labeling of the wound aspects still relies on the diagnostic view of the clinician or researcher. Due to the variable aspect of wounds, the system is not capable of distinguishing different wound tissues based on colors. Therefore, the observer defines a classification of the wound and determines which parts of the wound are
open or non-healed. As it is dependent on the observer’s clinical judgment, the data might show interobserver variability. Nevertheless, the standard deviation of the mean percentage of epithelialization assessed with digital image analysis was lower than the standard deviation of the clinician’s assessment (10.0 and 13.5, respectively). Furthermore, digital image analysis is time-consuming which limits its clinical usefulness in daily practice. In general, the shortcomings of digital image analysis emphasize the importance of investigating the validity of the subjective clinical assessment.

In conclusion, feasible, reliable, and valid wound assessment is a must in wound care to evaluate and therefore improve wound treatment. Reliability of digital image analysis was investigated and shown to be good for the assessment of epithelialization. However, it is relatively time-consuming and for this reason less feasible in a clinical setting. After comparison with the data obtained by digital image analysis, subjective clinical assessment was shown to be a valid method to evaluate epithelialization in transplanted wounds. We conclude that subjective clinical evaluation remains the most feasible method for wound epithelialization.

Acknowledgements

The authors would like to thank J. Dokter and H. Hofland for their contribution to the photographs. This study was partially financed by the Dutch Burns Foundation (grant number 07.109).
Chapter 5

References
