Clinical and patient reported outcomes in vitiligo

Linthorst Homan, M.W.

Citation for published version (APA):
DIGITAL IMAGE ANALYSIS VERSUS CLINICAL ASSESSMENT TO EVALUATE REPIGMENTATION AFTER PUNCH GRAFTING IN VITILIGO

M.W. Linthorst Homan
A. Wolkerstorfer
M.A.G. Sprangers
J.P.W. van der Veen

SUMMARY

Background: Repigmentation as a treatment outcome in vitiligo is not assessed in a standard way, making results of clinical trials hard to compare. Different types of repigmentation assessments after punch grafting have not been compared so far.

Objective: To compare assessments of repigmentation by a digital image analysis system (DIAS) with those of clinical observers and patients after punch grafting for vitiligo.

Methods: One vitiligo patch was selected in each patient (n=21). This patch was treated with the punch grafting technique. The grade of repigmentation (%) after 3 months was assessed by 1) DIAS; 2) 3 clinical observers and 3) the patient, scoring the grade of repigmentation on photographs. Physicians and patients also evaluated the global result on a 7-point scale.

Results: There was an almost perfect agreement between the three clinical observers and the DIAS (ICC 0.83). As expected, variation was found between the clinical observers. Patients’ scores showed a moderate agreement with the DIAS (ICC 0.49) and a poor agreement with the physicians (ICC 0.28). Overall, the patients were more satisfied with the results than the physicians.

Conclusions: Whereas the results of the digital and clinical assessments were comparable, patients’ ratings diverged. The DIAS can overcome the inevitable differences between observers, which are intrinsic to a visual grading method, and is advisable for clinical trials on vitiligo to objectively assess repigmentation in limited lesions.
INTRODUCTION

Vitiligo is a chronic skin disease characterised by the development of depigmented macules, due to loss of melanocytes. Several treatment options are available; non-surgical (e.g. local immunosuppressive treatments, phototherapy) and surgical therapies (e.g. punch grafting).\(^1\)

Reliable outcome measures are important to compare studies and to assess the changes over time. A variety of scoring systems is used to evaluate the treatment outcome in terms of repigmentation;\(^2\) making cross-study comparisons difficult. Moreover, most outcome measures rely on a subjective clinical assessment, which cannot exclude inter-observer bias and may therefore have limited accuracy and reproducibility. An objective measurement tool of repigmentation, such as a digital image analysis system (DIAS),\(^3-13\) can overcome some of these limitations. To date DIAS has not been evaluated for the assessment of the outcome after punch grafting.

The aim of this study was to compare assessments of repigmentation by computerised measurement (DIAS, objective scoring), different clinical observers (clinical scoring) and patients’ assessment as treatment outcome after punch grafting in vitiligo patients.

MATERIALS AND METHODS

Patients

Consecutive adult vitiligo patients (18 years or older) of the Netherlands Institute of Pigment Disorders (NIPD), who were scheduled for punch grafting were invited to participate in the study. An informed consent form was signed before inclusion. As punch grafting is a routine treatment in the NIPD and treatment was given as part of routine medical care, approval by the local research ethics committee was not needed.

Treatment

In each patient one vitiligo patch was selected. This patch was treated with the punch grafting technique.\(^5\)

Prior to treatment and after three months, clinical photographs of the treated lesion were taken with a Canon Power Shot G6 digital camera. Conditions were standardised throughout the study, including the room, lighting, background and distance between camera and patient.

Digital Image Analysis

Objective measurement of the grade of repigmentation was performed with a digital image analysis system (DIAS). To assess the size of the vitiligo lesion, the contours of the lesion were copied on a transparent sheet before and after treatment. This technique has the advantage that the curvature of the body surface will not bias the measurement. Afterwards these sheets were scanned using a
predefined resolution. DIAS involved an application based on Matlab (Mathworks, Inc, Natik, MA, USA). By comparing pre- and post-treatment pictures we computed the relative surface showing repigmentation expressed as percentage of the selected vitiligo patch.

**Clinical assessment**
Three different observers (all dermatologists) evaluated the percentage of repigmentation on a 10-point scale (1-10%, 11-20%....-91-100%) by comparing the photographs on a computer monitor before and after treatment. The observers were additionally asked to give their opinion about the result of treatment on a 7-point scale (excellent - very good – good - not good/ not bad - rather poor - poor - very poor).

**Patients’ assessment**
Patients were asked to assess the degree of repigmentation on a 10-point scale (1-10%, 11-20%....-91-100%) by comparing their own photographs, before and after treatment, on a computer monitor. The patients were additionally asked to give their opinion about the result of treatment on a 7-point scale (excellent - very good – good - not good/ not bad - rather poor - poor - very poor).

**Statistical analysis**
The scores of digital image analysis, clinical and patients’ assessment were compared and correlations were calculated. To measure the inter-observer reliability between the three physicians and patients, we used the intra-class correlation (ICC) coefficient. We used the absolute agreement definition for all ICC models. ICC values of 0.80–1.0, 0.60–0.80, and 0.40–0.60 are generally considered to indicate “almost perfect,” “substantial,” or “moderate” agreement, respectively.

We needed 21 photographs of patients with vitiligo to estimate the ICC with 0.15 accuracy and 95% certainty around the assumed point estimate of ICC= 0.80.

All analyses were done using a Statistical Package for Social Sciences (SPSS) database, the SPSS for Windows version 12.01.

**RESULTS**

**Patients**
Twenty-one adult patients with vitiligo (8 males/13 females) were included in this study. The mean age was 44.3 years (SD 14.7, Median 45.3). The body sites affected varied, including trunk (28.6%), leg (19.1%), feet (14.3%), arm (14.3%), hand (9.5%), wrist (9.5%) and face (4.8%).

**Digital, clinical and patients’ assessment**
The repigmentation scores provided by the DIAS, observers and patients as well as the scores of the 7-point scale are displayed per patient in Table 1. Agreement
Table 1. Scores of repigmentation (%) and 7-point scale. NA Not available

<table>
<thead>
<tr>
<th>Gender</th>
<th>Bodysite</th>
<th>Observer 1</th>
<th>Observer 2</th>
<th>Observer 3</th>
<th>DIAS</th>
<th>Opinion Observer 1</th>
<th>Opinion Observer 2</th>
<th>Opinion Observer 3</th>
<th>Opinion Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>Arm</td>
<td>0-10%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>Very poor</td>
<td>Very poor</td>
<td>Poor</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>Hand</td>
<td>31-40%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>Very poor</td>
<td>Rather poor</td>
<td>Very poor</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>Face</td>
<td>0-10%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>Very poor</td>
<td>Very poor</td>
<td>Very poor</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>Trunk</td>
<td>NA</td>
<td>51-60%</td>
<td>21-30%</td>
<td>11-20%</td>
<td>Not good-not bad</td>
<td>Rather good</td>
<td>Rather poor</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>Leg</td>
<td>61-70%</td>
<td>51-60%</td>
<td>31-40%</td>
<td>0-10%</td>
<td>Rather good</td>
<td>Good</td>
<td>Poor</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>Arm</td>
<td>0-10%</td>
<td>21-30%</td>
<td>11-20%</td>
<td>0-10%</td>
<td>Not good-not bad</td>
<td>Poor</td>
<td>Poor</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>Hand</td>
<td>11-20%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>Poor</td>
<td>Poor</td>
<td>Very poor</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>Arm</td>
<td>0-10%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>Very poor</td>
<td>Very poor</td>
<td>Very poor</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>Trunk</td>
<td>NA</td>
<td>91-100%</td>
<td>81-90%</td>
<td>71-80%</td>
<td>Rather good</td>
<td>Very good</td>
<td>Rather good</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>Arm</td>
<td>NA</td>
<td>21-30%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>Rather poor</td>
<td>Poor</td>
<td>Very poor</td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>Arm</td>
<td>71-80%</td>
<td>51-60%</td>
<td>21-30%</td>
<td>11-20%</td>
<td>Rather good</td>
<td>Rather good</td>
<td>Not good-not bad</td>
</tr>
<tr>
<td>12</td>
<td>F</td>
<td>Trunk</td>
<td>51-60%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>Very poor</td>
<td>Very poor</td>
<td>Very poor</td>
</tr>
<tr>
<td>13</td>
<td>M</td>
<td>Foot</td>
<td>NA</td>
<td>11-20%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>Poor</td>
<td>Very poor</td>
<td>Very poor</td>
</tr>
<tr>
<td>14</td>
<td>F</td>
<td>Leg</td>
<td>81-90%</td>
<td>71-80%</td>
<td>51-60%</td>
<td>0-10%</td>
<td>Rather good</td>
<td>Good</td>
<td>Poor</td>
</tr>
<tr>
<td>15</td>
<td>M</td>
<td>Trunk</td>
<td>71-80%</td>
<td>61-70%</td>
<td>31-40%</td>
<td>21-30%</td>
<td>Rather good</td>
<td>Good</td>
<td>Rather good</td>
</tr>
<tr>
<td>16</td>
<td>M</td>
<td>Foot</td>
<td>0-10%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>Very poor</td>
<td>Very poor</td>
<td>Poor</td>
</tr>
<tr>
<td>17</td>
<td>F</td>
<td>Trunk</td>
<td>41-50</td>
<td>21-30%</td>
<td>11-20%</td>
<td>0-10%</td>
<td>Rather poor</td>
<td>Not good-not bad</td>
<td>Poor</td>
</tr>
<tr>
<td>18</td>
<td>F</td>
<td>Trunk</td>
<td>21-30%</td>
<td>11-20%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>Poor</td>
<td>Very poor</td>
<td>Poor</td>
</tr>
<tr>
<td>19</td>
<td>M</td>
<td>Leg</td>
<td>11-20%</td>
<td>51-60%</td>
<td>31-40%</td>
<td>0-10%</td>
<td>Rather good</td>
<td>Good</td>
<td>Very bad</td>
</tr>
<tr>
<td>20</td>
<td>M</td>
<td>Leg</td>
<td>0-10%</td>
<td>21-30%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>Rather poor</td>
<td>Very poor</td>
<td>Very poor</td>
</tr>
<tr>
<td>21</td>
<td>F</td>
<td>Foot</td>
<td>0-10%</td>
<td>11-20%</td>
<td>0-10%</td>
<td>0-10%</td>
<td>Rather poor</td>
<td>Very poor</td>
<td>Poor</td>
</tr>
</tbody>
</table>
between clinical observers and DIAS was almost perfect for observers 1 and 2 (ICC 0.90 and 0.85 respectively), but moderate for observer 3 (ICC 0.47). Overall, the agreement between the three clinical observers combined (average rating) and DIAS was almost perfect (ICC 0.83) (Table 2). Patients on the other hand scored the degree of repigmentation in general higher than did the physicians or the DIAS (Table 1). Patients’ scores showed a moderate agreement with the DIAS (ICC 0.49) and a poor agreement with the physicians (ICC 0.28). Overall, the patients were more satisfied with the results than the physicians. For example, seven patients rated the change as very good and good whereas two of the three observers did not endorse these two top categories.

**DISCUSSION**

In this study we compared an objective digital assessment system with the more subjective assessment by physicians and patients of repigmentation after punch grafting for vitiligo. We found an almost perfect agreement between the clinical observers and the DIAS. However, there was a substantial variation among the clinical observers, a finding that is common for visual assessment. Consequently, clinical observers should be trained to calibrate their scoring of repigmentation.

Conversely, DIAS does not introduce inter-rater and intra-rater observer variability. It is therefore a good objective assessment of repigmentation in vitiligo patches, thus enabling reliable cross-treatment and cross-study comparisons.

DIAS has some limitations though. Firstly, on curved body surfaces, the DIAS has a certain degree of distortion resulting from the sloped surface, unless a transparent sheet is used to trace the lesion's contours, which we did. Furthermore, perifollicular repigmentation or other complex patterns are difficult to catch by using boundary tracing system. Moreover, for the evaluation of widespread vitiligo lesions DIAS is time consuming. For this treatment the VASI score\(^\text{16}\) may be more appropriate. Furthermore, observers have to be trained to be able to use the DIAS which includes a learning curve. Without adequate training, DIAS might be subject to some intra-and inter-rater variability as well.\(^\text{17}\) Finally, as the technique is time consuming, it is less feasible in daily practice.

A limitation of our study is that we did not explore other, more complex, DIAS options (e.g. analysis of colour or ultraviolet photographs). These options could also overcome the difficulty in analyzing vitiligo areas of larger extension.

In our study, patients scored the degree of repigmentation generally higher than the physicians or the DIAS, which may reflect wishful thinking and is in line with their higher treatment satisfaction as compared to physicians. It is important to include the patients’ assessment of the treatment outcome, because their satisfaction is the most important goal of the treatment. For atopic dermatitis a few self-assessment scoring systems have been developed, for example the patient oriented SCORAD.\(^\text{18}\) These scoring systems can be used in daily clinical practice,
e.g. to screen patients for their need for (additional) care and in the context of research. For vitiligo, self-assessment instruments for treatment outcome have not been developed yet.

Other methods of repigmentation scoring, such as point counting\(^9\) have been proposed and may provide a simple alternative to measure the vitiligo surface area. Further research is needed to compare this method with DIAS.

Although it was not an objective of our study to evaluate percentage of repigmentation after punch grafting, we noticed a remarkable variation of repigmentation. This is partly a result of the inclusion of locations that are known to react poorly to treatment, such as the feet. Moreover, intrinsic patient-related factors may be involved.

From this study we can conclude that:

1. Clinical scoring in general is as adequate as digital image analysis, though inter-observer variability can be great and training should be given to observers in the context of clinical trials.
2. Patients’ assessments of repigmentation differ from those of clinicians and are more positive as is their satisfaction with treatment (compared to those of physicians).
3. Digital image analysis is the most objective measure of repigmentation and for this reason could be regarded as the gold standard for assessing repigmentation in clinical research. Since the technique is complex and laborious, it is only feasible for monitoring limited areas of vitiligo.

**Acknowledgement:** We thank B.L. Kallimanis – King and Robert Lindeboom for the help with the statistical analyses of the data.

| Table 2. ICC: Intraclass Correlation Coefficients, C.I: Confidence Interval |
|-------------------------------|-------------------|
|                               | ICC   | 95% C.I. |
| Observer 1 vs DIAS\(^o\)       | 0.90  | 0.73 – 0.96 |
| Observer 2 vs DIAS\(^o\)       | 0.85  | 0.63 – 0.94 |
| Observer 3 vs DIAS\(^o\)       | 0.47  | 0.06 – 0.75 |
| (Observer 1, 2 and 3) vs DIAS\(^1\) | 0.83  | 0.64 – 0.93 |
| Patients vs DIAS\(^2\)         | 0.49  | 0.09 – 0.75 |
| Patients vs (Observer 1, 2 and 3)\(^2\) | 0.28  | -0.16 – 0.63 |

\(^o\) = model 3 (two-way fixed), single measure
\(^1\) = model 2 (two-way random), single measure
\(^2\) = model 1 (one-way random), single measure
REFERENCE LIST


