Analysis of portwine stain disfigurement and pulsed dye laser treatment results
Koster, P.H.L.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 2

Assessment of clinical outcome following flashlamp pumped pulsed dye laser treatment of portwine stains: a comprehensive questionnaire

Petra HL Koster, MD, Amsterdam
Patrick MM Bossuyt, Prof PhD, Amsterdam
Chantal MAM van der Horst, MD, Amsterdam
Geert HM Gijsbers, PhD, Rotterdam
Martin JC van Gemert, Prof PhD, Amsterdam


1 Department of Plastic, Reconstructive and Hand Surgery, Academic Medical Center, University of Amsterdam, NL
2 Department of Clinical Epidemiology and Biostatistics, University of Amsterdam, NL
3 Laser Laboratory, Thorax Center, De Wez Hospital, Rotterdam, NL
4 Laser Center, Academic Medical Center, University of Amsterdam, NL

This study was supported by grant 06-02-03 from the Dutch Health Insurance Council.
Abstract

A generally accepted method to assess the clinical outcome of laser treatment of portwine stains is not available. This paper describes the development and evaluation of a comprehensive questionnaire for the assessment of the following portwine stain characteristics: color (hue and lightness), sharpness of boundary, pigmentation, size, shape, skin surface structure and hypertrophy of underlying tissue. The questionnaire was applied to photographs of 54 patients with previously untreated portwine stains in the head-neck area. Photographs were taken before treatment and after five treatments of the entire portwine stain with a flashlamp pumped pulsed dye laser. Each photograph was evaluated separately by a panel of five professionals: the treating physician, two plastic surgeons, a dermatologist, and a clinical photographer. Treatment results were assessed by comparing ratings before and after treatment. Agreement amongst raters was acceptable for all portwine stain characteristics, as was shown by weighted kappa analysis. The reliability of the answers was further improved by taking the average rating of the five panel members. The scores thus obtained were very reliable with Cronbach alpha coefficients of 0.8 on average. After five treatments of the entire portwine stain, the most considerable changes were measured in the ratings for color (lightening of the stain by 32 percent), boundary (sharpness reduced by 38 percent), and size (13 percent decrease). Using the questionnaire can be helpful in the design of comparative clinical trials on portwine stain treatment, and may facilitate comparison of treatment results between different treatment centers and/or different lasers.
Introduction

Although lasers have been used since the early 1960s to treat portwine stains, consensus on a method to compare treatment results between different types of lasers or between groups of patients has never been achieved. Methods used to assess laser treatment results range from objective, such as color measurement with a colorimeter, to subjective, such as grading the results as either satisfactory or non-satisfactory without using strict criteria.

The most extensive methods to assess laser treatment results were designed for the argon laser, which has been the laser of choice for many years. For example, Quaba designed a protocol to measure color changes, scarring, textural skin changes, and changes in the use of make-up. Unfortunately, these methods turn out to be less useful to evaluate more recent treatment modalities, such as the flashlamp pumped pulsed dye laser, because virtually no scar tissue formation or textural skin changes occur with these lasers. Because flashlamp pumped pulsed dye laser treatment offers the possibility of treating (young) children, the make-up issue also looses importance.

Several alternative methods have been designed, most of them focusing on the color of the portwine stains. However, restricting the assessment of clinical outcome to color disregards other characteristics of a portwine stain that possibly contribute to its disfiguring effect, such as size of the portwine stain and sharpness of its boundary. Quaba already mentioned the favorable effect of argon laser treatment on feathering out the sharp outline of a portwine stain, thus achieving better blending of the portwine stain with surrounding normal skin.

The goal of this study was to develop a rating system to quantify the different aspects of a portwine stain that can contribute to its disfiguring effect. Repeated ratings during the course of treatment would enable us to objectively monitor treatment progress. We, therefore, composed a questionnaire addressing an array of portwine stain characteristics, providing a detailed description of the portwine stain. Reliability of the questionnaire was tested by means of weighted kappa analysis and by
calculation of Cronbach alpha coefficients. In this paper, we present the results obtained for a group of 50 portwine stain patients, of whom pictures were taken under standardized conditions before treatment and after the entire portwine stain had been treated five times with a flashlamp pumped pulsed dye laser. From these pictures, the characteristics of the portwine stains were rated by a team of five professionals. Afterward, pretreatment and post-treatment ratings were compared to assess treatment results.

**Materials and methods**

**Portwine stain characteristics**

Based on our experience with portwine stain treatment, we postulated a number of characteristics that contribute to the distinguishing effect of a portwine stain and that are potentially affected by treatment. Obviously, color was considered to be an important characteristic and was, therefore, included as the first item of the questionnaire. Panel members were asked to assess two aspects of this characteristic: hue (pink, red, or purple), and lightness (pale, bright, or dark). A third aspect of color was included as patchiness. This patchiness also refers to the reticular pattern that may occur after treatment.

*Sharpness of the boundary* was taken into account because we expected the boundary of a portwine stain to play an important role in its conspicuousness.

*Presence of changes* may play a role in the assessment of post-treatment portwine stains. Treated areas sometimes show hyperpigmentation or hypopigmentation. The first condition usually is a transient phenomenon, but the latter condition might be permanent.

An impression of the size of the portwine stain was asked for, using a range of subjective categories, from "very small" to "very large".

*Shape* of the portwine stain, i.e., its irregularity, was also taken into account, on the assumption that an irregularly shaped portwine stain would cause more disfigurement than a regularly shaped one.
Surface-structure was considered indispensable in rare cases of scar tissue formation.

Hypertrophy of subcutaneous tissue is present in a relatively small number of patients only, but we expected it to contribute largely to disfigurement in case it is present. Whether it can be improved by flashlamp pumped pulsed dye laser treatment is doubtful.

Questionnaire

With the eight characteristics mentioned in the previous section a questionnaire was constructed. For each characteristic, a series of rating options was given, ordered according to increasing contribution to disfigurement. An initial questionnaire was tested with a panel of five professionals, using the slides of 40 patients. Based on comments of this panel, a few alterations were introduced regarding the sequence of questions and the required range of options per characteristic. The most important of these changes concerned the range of options for the characteristic color, i.e. the combination of hue and lightness. The panel members found some of the rating options for this characteristic to be interchangeable, e.g. what some would call dark-pink others might call light-purple. In the resulting questionnaire we therefore considered the following options to be comparable: bright-pink and pale-red, dark-pink and pale-purple, as well as dark-red and bright-pink. Comparable sets were combined into one rating option, reducing the possible number of 10 rating options to 7. The questionnaire is shown in Table 1.

Patients

The study population consisted of 70 white patients (24 male, 46 female), ages ranging from 6 to 50 yr (average, 11.4 yr; sd, 8.6 yr), with previously untreated portwine stains in the head/neck region. Patients were treated with a Candela SPTL-1 flashlamp pumped pulsed dye laser (wavelength, 585 nanometer; pulse duration, 450 microsecond; repetition rate, 0.3 Hz; and spot diameter, 5 mm).

Before treatment and after five treatments of the entire portwine stain, standardized color slides were taken by a clinical photographer, in a
Table 1: Questionnaire

1. COLOR (macular degeneration)
   1. normal
   2. pale pink
   3. pale pink with few patches
   4. pale purple with dark spots
   5. dark pink
   6. bright purple or dark red
   7. dark purple

2. PATTERN
   1. not at all patchy
   2. a little patchy
   3. rather patchy
   4. very patchy

3. BOUNDARY
   1. vague
   2. rather sharp
   3. very sharp

4. HYPO/HYPER-PIGMENTATION
   1. no white, brown, or discoloration
   2. a little brown or white
   3. rather brown or white

5. SIZE
   1. very small
   2. small
   3. medium
   4. large
   5. very large

6. SHAPE
   1. round
   2. somewhat irregular
   3. very irregular

7. SURFACE
   1. smooth
   2. a little uneven
   3. rather uneven

8. HYPERTRPHY
   1. rather hypertrophic
   2. a little hypertrophic
   3. rather hypertrophic
   4. very hypertrophic
Rating procedure
The slides of the patients were presented to a panel with five members: the treating physician, two plastic surgeons, a dermatologist, and a clinical photographer. The slides were projected in pairs, each pair consisting of one slide of the patient in full face and one in profile. Of all 70 patients, two such slide pairs were available: one pair taken before treatment and a similar pair taken after 5 treatments of the entire portwine stain. To judge the pretreatment and posttreatment slide pairs of one patient independently from each other, the two slide pairs were presented with a maximal separation between them. In half of the randomly selected cases, the pretreatment slide pair preceded the posttreatment slide pair in the presentation order; in the remaining half, the pretreatment slide pair followed the posttreatment pair. The slide pairs were projected in a random sequence with respect to patient's age.

Before the actual procedure was started, the panel members were given an impression of the portwine stains to be judged by viewing the first 20 slide pairs without having to fill out a questionnaire; projection was then started again with the first slide pair. For each slide pair, all panel members individually filled out one questionnaire.

Changes in portwine stain characteristics due to laser treatment were calculated by subtracting ratings before treatment from ratings after 5 treatments. First, the average rating of all five panel members for each portwine stain characteristic was calculated per slide pair. Then, the average rating per characteristic was calculated for the 70 pretreatment and posttreatment slide pairs separately. Changes were tested for significance using the paired t-test statistic.

Reliability
The reliability of the panel assessment, as well as the reliability of the questionnaire as a measurement instrument, were evaluated by means of weighted kappa analysis\(^1\) for the individual ratings, and by calculating the Cronbach alpha coefficient\(^1\) for the average ratings.

The weighted kappa represents the amount of agreement between 2 individual panel members beyond the amount that can be expected based
on coincidence, taking into account the proportion of disagreement. The kappa statistic takes values between -1 and 1, a weighted kappa of 0 indicating no more agreement than that based on sheer coincidence, and a weighted kappa of 1 meaning that the two panel members completely agree with each other. When panel members completely agree with each other, the rating of each panel member is as trustworthy as the ratings of the others, and one rating will do.

In cases of less perfect agreement, where kappa values are lower than 1, the reliability of the assessment results can be increased by taking the average rating of all panel members. In order to be used, a coefficient was used as a measure of the reliability of these average ratings. This method is similar to those used in correlation to study the reliability of the results. In our case, the panel members are the replicates of the test and scale. The value of the Cronbach alpha coefficient can range from 0 to 1, where 1 means the results are completely reliable and 0 indicates perfect reliability of the results. The reliability of the results is a reflection of the reliability of the questionnaire as a measurement instrument.

Results

**Reliability of portwine stain characteristic ratings**

For all portwine stain characteristics mentioned on the questionnaire, interrater agreement statistics were calculated. The results for pretreatment and posttreatment sides were highly similar. In Table 2, the results based on the pretreatment sides are shown. In the first column, the characteristics are listed in order of appearance on the questionnaire. The second column shows the average percentage of exact agreement among panel members, calculated by averaging the percentages of exact agreement between the 10 combinations of two panel members. The third and fourth columns, respectively, show the average weighted kappa and the Cronbach alpha coefficient. The average weighted kappa was computed by averaging the outcomes of the 10 possible combinations of the five panel members.
According to the guidelines for the interpretation of kappa values as derived from Altman, the weighted kappa values in Table 2 indicate moderate to good interrater agreement for the characteristics color, boundary, size, and hypertrophy (kappa's between 0.44 and 0.64). Intercrater agreement on patchiness and shape can be considered fair (kappa's of 0.32 and 0.28, respectively). For the characteristics pigmentation and surface-structure, consensus among panel members is low according to the kappa values of 0.11 and 0.22. The Cronbach alpha coefficients display an acceptable degree of reliability of the average ratings for all characteristics, with values exceeding 0.6 for all characteristics except pigmentation. For pigmentation, the Cronbach alpha coefficient is somewhat lower (0.43), but also in this case an improvement of the reliability was achieved as a result of taking average panel ratings.

Table 2: Average interrater agreement values for characteristics of untreated portwine stains

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Average agreement</th>
<th>Average weighted kappa</th>
<th>Cronbach alpha coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>45</td>
<td>0.46</td>
<td>0.92</td>
</tr>
<tr>
<td>Patchiness</td>
<td>54</td>
<td>0.72</td>
<td>0.81</td>
</tr>
<tr>
<td>Boundary</td>
<td>61</td>
<td>0.77</td>
<td>0.83</td>
</tr>
<tr>
<td>Pigmentation</td>
<td>55</td>
<td>0.41</td>
<td>0.43</td>
</tr>
<tr>
<td>Size</td>
<td>50</td>
<td>0.51</td>
<td>0.56</td>
</tr>
<tr>
<td>Shape</td>
<td>82</td>
<td>0.28</td>
<td>0.76</td>
</tr>
<tr>
<td>Surface structure</td>
<td>59</td>
<td>0.22</td>
<td>0.66</td>
</tr>
<tr>
<td>Hypertrophy</td>
<td>71</td>
<td>0.41</td>
<td>0.86</td>
</tr>
</tbody>
</table>
Panel ratings and the influence of treatment

In Table 4, the averaged ratings for all 75 patients are listed per portwine stain characteristic before treatment and after five treatments of the entire portwine stain. Also shown are the percent change as a result of treatment and the corresponding p-values. To calculate the percent change, the pretreatment rating was subtracted from the rating after five treatments, and expressed as a percentage relative to the pretreatment rating. A positive change represents an improvement of the portwine stain characteristic, whereas a positive value represents a deterioration of the characteristic.

Table 4: Rating values of portwine stain characteristics in 75

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Before Treatment</th>
<th>After Treatment</th>
<th>Percent Change</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.001</td>
</tr>
<tr>
<td>Intensity</td>
<td>0.5</td>
<td>0.3</td>
<td>0.2</td>
<td>0.01</td>
</tr>
<tr>
<td>Boundary</td>
<td>0.5</td>
<td>0.4</td>
<td>0.1</td>
<td>0.03</td>
</tr>
<tr>
<td>Presentation</td>
<td>0.5</td>
<td>0.4</td>
<td>0.1</td>
<td>0.04</td>
</tr>
<tr>
<td>Size</td>
<td>0.5</td>
<td>0.3</td>
<td>0.2</td>
<td>0.04</td>
</tr>
<tr>
<td>Shape</td>
<td>0.5</td>
<td>0.4</td>
<td>0.1</td>
<td>0.02</td>
</tr>
<tr>
<td>Surface structure</td>
<td>0.5</td>
<td>0.4</td>
<td>0.1</td>
<td>0.04</td>
</tr>
<tr>
<td>Hypertrophy</td>
<td>0.5</td>
<td>0.3</td>
<td>0.2</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Table 4 shows that the largest reductions due to treatment were achieved for color and boundary; the distinguishing effect of these characteristics was diminished by 50 and 40 percent, respectively. The third most important improvement was achieved for size, this aspect of disfigurement being reduced with 35 percent.
Patchiness was not altered in a statistically significant way, whereas pigmentation seemed to become somewhat worse with treatment. This finding could well be because pigmentary changes are an adverse side effect of treatment. Shape showed a small but statistically significant decrease. Neither surface-structure nor hypertrophy showed significant changes with treatment. Most portwine stains included in the study were smooth and treatment did not result in any scar tissue formation. In the case of hypertrophy, which was present in 24 percent of the patients, we conclude that we have not been able to improve this aspect of the portwine stain with flashlamp pumped pulsed dye laser treatment.

Discussion

The aim of our study was to develop a method for objective and quantitative assessment of clinical results achieved with laser treatment of portwine stains. Ideally, such a method should detect relevant changes in portwine stain skin in an unambiguous way, making comparison of results possible between different treatment centers and different laser systems.

By discerning a number of characteristics that may contribute to a portwine stain's disfiguring effect and that each may be affected by treatment in a different way, we developed a multi-item questionnaire. The ratings on this questionnaire proved to have acceptable interrater agreement, with fair to good agreement on all items. We concluded this despite the fact that weighted kappa values for pigmentation and surface-structure were only 0.11 and 0.22, respectively. These low interrater agreement values are caused by a peculiarity of weighted kappa analysis rather than by true unreliability of the answers, because the kappa statistic is very sensitive to a one-sided distribution of given answers. For both pigmentation and surface-structure, the frequency tables, from which the weighted kappa values are calculated, showed skewed distributions of the answers. For pigmentation, this skewedness is caused by the fact that
hypopigmentation or hyperpigmentation, being a side effect of treatment, does not occur in untreated portwine stains. As for surface-structure, most of the portwine stains included in this study were smooth, and therefore were rated as such.

Fig. 1: Frequency tables for 70 patients for the characteristic surface-structure for rater pairs X/Y and X/Z.

In Fig. 1 the frequency tables for the characteristic surface-structure are shown for rater pairs X/Y and X/Z. Although raters X and Y agreed 59 of 70 times, their answers are not distributed evenly around the diagonal of the frequency table, resulting in a kappa value of zero. Raters X and Z agreed 61 of 70 times, but their answers are more evenly spread around the diagonal of the frequency table and a kappa of 0.45 was found. The calculated kappa values in cases in which the answers show a skewed distribution suggest larger differences between raters than the frequency tables demonstrate. We, therefore, conclude that agreement among panel members was fair to good on all characteristics mentioned in the questionnaire. In this study, the reliability of the answers was further enhanced by taking the average rating of all five panel members. The
Cronbach alpha coefficient exceeded 0.6 for all characteristics except pigmentation (0.4).

Previous assessment procedures for laser treatment results of portwine stains have relied mainly on the assessment of color. A method often used to judge the amount of lightening of the portwine stain after treatment was introduced by Garden et al. In their system pretreatment and posttreatment slides are compared and lightening of the portwine stain is quantified on a percentage scale of 0 to 100 percent. Other methods classify a portwine stain by comparing it with a standard, either a color strip, e.g., according to the Munsell standard, or a standardized set of portwine stains. A visual analog score can then be calculated from the difference between pretreatment and posttreatment classifications.

Our study shows that the results of flashlamp pumped pulsed dye laser therapy go beyond a mere alleviation of color. Repeated treatment also resulted in a less sharply defined boundary of the portwine stain and a decrease of its size. Although these changes are likely to be associated with lightening of the stain, they do not necessarily occur in all patients at the same rate. These potential differences should be considered in the assessment of treatment results. However, an evaluation system based on individual ratings holds the risk of subjectivity. With the questionnaire we tried to achieve objectivity by using a series of standard questions. The reliability statistics prove that the answers indeed exceed subjectivity. Reliability of the rating system could probably be improved even further by implementing a series of reference photographs for each portwine stain characteristic, simplifying the rating procedure to a matching task.

A potential disadvantage of the questionnaire is found in the many aspects of a portwine stain it addresses, making the rating of a series of photographs a rather time-consuming task. For practical purposes a few simplifications could be considered. The characteristics showing the largest changes, and consequently seeming to be the most important parameters in evaluating treatment results, were color (the combination of hue and lightness), the sharpness of the boundary and the size of the portwine stain. Although the change in pigmentation is small it should not be disregarded, being one of the few adverse effects of treatment.
Patchiness and shape are probably best combined in one parameter, because panel members reported having difficulties judging these items separately. The item surface structure is indispensible in the rare case of scar tissue formation, although it did not show any significant changes with treatment in our study population. Hypertrophy was not altered significantly by treatment, and inclusion of this characteristic in the questionnaire is therefore debatable. A more compressed list would then consist of the six items color, boundary size, pigmentation, patchiness, shape (e.g., flatness), and surface structure.

In conclusion, setting standards on treatment and communicating results between treatment centers calls for a unified and objective approach in assessing treatment results. Our approach does not require additional investments in expensive measurement systems and may be applicable in most centers. It could be used in clinical trials designed to compare the effectiveness of new lasers with that of existing ones, or to evaluate different treatment modalities using existing laser systems. In this study, we used standardized photographs. It is as yet unknown whether similar results with respect to reliability and sensitivity to changes in the portwine stain can be obtained by judging patients in real life, before and after treatment. Also, future research should document the applicability of this rating system in monitoring progress in the treatment of individual patients.
References


Chapter 2