Morphometry of dermal collagen orientation by Fourier analysis is superior to multi-observer assessment.
Abstract

In human dermis, collagen bundle architecture is randomly organised, whereas in pathological conditions, such as scar tissue and connective tissue disorders, collagen bundle architecture is arranged in a more parallel fashion. Histological examination by one or two observers using polarised light is the most common method to determine collagen orientation, and is therefore considered the ‘gold standard’.

We hypothesised that an image analysis technique, Fourier analysis, would be more accurate than observer ratings. Fourier analysis was applied to 271 images of scar tissue and normal skin that were acquired by confocal laser-scanning microscopy. Observers rated the same areas using polarised light and they also rated the confocal microscopy images. Computer images consisting of different types of ellipses were generated with a fixed ‘true’ orientation. Observers and Fourier analysis evaluated the orientation of these images to establish their accuracy. The inter-observer reliability was acceptable when at least three observers rated polarised light images ($r \geq 0.69$) whereas already two observers were sufficient for rating confocal microscopy images ($r \geq 0.71$). Fourier analysis correlated better with the observer ratings of confocal microscopy images ($r = 0.69$) than with the observer ratings of polarised light microscopy images ($r = 0.42$). Fourier analysis was more accurate than four observers for the evaluation of the ‘true’ orientation for almost all types of computer generated images.

For the first time it is shown that Fourier image analysis is suitable for morphometry of the dermal collagen orientation and leads to a superior measurement of collagen orientation compared to subjective histologic evaluation by several experts. However, if observers perform an evaluation by conventional light microscopy, at least three observers are required to attain an acceptable inter-observer reliability.
Introduction

The mechanical qualities of skin provide a unique combination of elasticity and strength. Its pliability is not hampered by the strength that is required to resist physical stress to which the human integument is exposed. These functional qualities of the dermis are predominantly due to the dermal collagen network and, to a lesser extent, to elastin and extracellular matrix substances. Microscopic views of normal skin reveal that collagen bundles show a basket weave-like pattern and appear randomly organised.

The importance of the dermal architecture is illustrated by pathological disorders of connective tissue, such as Ehlers-Danlos Syndrome, Marfan Syndrome and cutis laxa. These disorders are typically associated with changes in mechanical qualities of the integument.

A different collagen architecture is also present in scar tissue, as human skin lacks the ability to regenerate after wounding. Tissue restoration takes place by a repair mechanism that often results in a tough scar which is less pliable and less resistant to sheer forces than native skin even when matured. Histopathological studies of mature scar tissue show collagen bundle orientation parallel to the epidermis rather than a more randomly organised pattern found in normal dermis.

Collagen structure is an integral part of wound healing research. Analysis of collagen orientation is usually performed by one or two observers using conventional light microscopy in combination with polarised light. This technique of light polarisation makes use of the birefringence characteristics of collagen. Non-birefringent structures become almost invisible and the collagen structure is easily identi-
The optimal assessment technique of collagen orientation, however, has not been established. Observer scores are currently the 'gold standard' for the assessment of collagen orientation. We evaluated a technique that is available in a number of image analysis software programs for application on personal computers and may potentially replace the observers' judgement: Fourier analysis. This mathematical algorithm is designed to quantify the spectrum over the whole range of frequencies in data. This frequency based method in itself is not new and has already offered benefits in many scientific fields such as one-dimensional time varying signals in electric engineering. It has also been introduced in the biomedical research in combination with infrared spectrometry to study the intermolecular structure of human tissues. The Fourier transformation is suited for analysis of two-dimensional images. In this way, structure organisation within an image is calculated. The technique has been performed for histologic images of ligaments and sclerodermal lesions.

The suitability of Fourier image analysis for the morphometry of the collagen structure in comparison to histologic assessment by experts has not been studied to date. We therefore conducted a study to evaluate the reliability and accuracy of Fourier analysis and observer ratings. We hypothesised that Fourier analysis is applicable for measurement of the collagen structure in normal dermis and scar tissue and that it is more accurate and reliable than results obtained by experts.

Firstly, the inter-observer reliability was established of collagen orientation assessment that was performed by one, two, three or four observers to study the minimal number of observers that are required for a reliable analysis of collagen orientation. All observers rated collagen orientation with conventional light microscopy and also with confocal microscopy images. This circumvented possible confounding influences by type of imaging (conventional light microscopy and confocal microscopy). The observer ratings of light and confocal microscopy images were then correlated with the Fourier analysis results. However, as there is no 'gold standard' we could not establish the accuracy of the different techniques, as we did not know the 'true' collagen orientation of the image. The accuracy...
Morphometry of dermal collagen orientation by Fourier analysis is superior to multi-observer assessment

racy, or validity", was accomplished using computer generated model images with a fixed ‘true’ orientation, which were also evaluated by observers and Fourier image analysis.

Material and methods

Histological evaluation of the slides was performed on mature scar tissue and normal dermis. The biopsies were harvested from burn scars after split skin grafting. The ethics committee from the Red Cross Hospital approved the protocol before the study started. Biopsies were harvested of 65 patients, who gave a signed informed consent before enrolment into the study (36 males and 29 females). The mean age of the population was 32.3 years with a standard deviation (SD) of 18.6 years.

Harvesting of biopsies and histopathological staining

We analysed 144 sections taken from scar tissue approximately one year after surgery. Eight sections of normal skin within the same group of patients were obtained and added to the analysis. After disinfecting the area and local infiltration with lidocaine (1% solution with epinephrine), a three millimetre punch biopsy was obtained. The biopsy was placed in formalin and processed into histological slides of approximately 5 μm. Sections were then stained with hematoxylin-eosin (HE).

Confocal laser-scanning microscopy

Images of HE sections were scanned with a confocal laser-scanning microscope (Leica Micro Optics, Heidelberg, Germany) with a 6.3 x 0.2 objective and a 50μm pinhole making use of the fluorescent properties of eosin (excitation: 488nm, detection: LP 610nm). In this way the interference of the hematoxylin stained nuclei in standard brightfield image was avoided. Images were adapted to the full dynamic range of the system (8 bit) to standardise the contrast/brightness ratio. The epidermis was always orientated parallel to the X-axis of the image. Images of 794μm x 794μm were created of the superficial dermal layer and the layer just underneath the superficial layer (also a 794μm x 794μm image), the so-called deeper dermis.
Fourier analysis

Image analysis was done with the Fast Fourier Transform module of the Qwin Pro image analysis software (version 2.2, Leica Imaging Systems LTD, Cambridge, UK). The collagen bundle orientation was estimated by calculating the width/height ratio of the zeroth-order maximum in the generated power plot of the image. This was termed the collagen orientation index, which may range between zero and one. Parallel collagen orientation results in an elongated power spectrum that leads to a small orientation index. Perfectly random tissue results in an orientation index that approximates one.

Observer ratings using polarised light and confocal microscope images

Four trained observers assessed the collagen orientation of the superficial and deep dermal layer in two ways. Firstly, the superficial and deeper dermal layers of all sections were analysed under a light microscope using polarised light. Secondly, scanned confocal microscope images of the superficial and deeper dermal layers, which were also used for the Fourier analysis, were rated. Examples of the appearance of the evaluated areas for normal skin and scar tissue are shown in Figure 1. In this figure the histological view of the section is displayed together with the same area after applying polarised light, and after scanning the section by confocal microscopy. The sections and images were blinded to prevent bias through prescience. A five-step scale was used for the scoring: extremely parallel, predominantly parallel, mixed organisation, predominantly random, extremely random. The ratings were averaged for both polarised light and confocal microscopy.

Computer generated images with established orientation

Images were composed which mimicked the eosin fluorescent images of skin tissue. The shape of an ellipse was used as a base element, which follows a two dimensional Gaussian distribution with a fixed length of ten pixels. By varying the width to one, two, five or ten pixels, four types of ellipses were created and put into the computer model. Small elliptic shapes with a 1:10 width/length ratio appeared as small lines that resembled small collagen bundles of scar tissue. The less elongated, and thicker, ellipses mimicked normal skin with
thicker collagen bundles that were frequently cut in a perpendicular fashion. The circle (the length and width being ten pixels) was put into the model as an extreme, merely theoretical example.

In the 512 x 512 pixels sized image, 2000 ellipses were used to get comparable filling of the images. The image was made transparent because overlapping parts of the ellipses were added. The final image was normalised to get a maximum intensity of 255, comparable to the fluorescent images scanned by confocal laser-scanning microscopy. Before ellipses were added to the image they were rotated over an angle. The variance of the angle was based on an orientation vector that was composed of a normally distributed random number in the horizontal and vertical plane. An example of an image with an orientation index of one is shown in Figure 2.

The computer designed twenty-one images with an orientation ranging from zero (perfect parallel) to one (perfectly random) with intervals of 0.05 for all three different types of ellipses and the circle. The 84 images were randomised, blinded and evaluated by the observers and Fourier analysis as mentioned above.

Statistical analysis

The data were analysed by the statistical program SPSS for Windows 8.0 (SPSS Inc. Chicago, IL., USA). The intraclass correlation coef-
ficient (ICC)\textsuperscript{12,13} with its 95% confidence interval (CI) was calculated to assess the inter-observer reliability for the group and the reliability of one observer, also known as the average measure ICC and the single measure ICC, respectively. For the calculations of the ICC, the two-way-random effect model was selected and calculated for absolute agreement of the scores. It is not possible to calculate the ICC between the different pairs of two and three observers in the same way, therefore these ICC's were calculated for each combination of two and three observers. The range of ICC's for all combinations of two and three observers is given.

Scatter plots were made for the combinations of average scores for both observer categories and the results of the Fourier analysis. The Spearman's Rho correlation coefficient was used for the correlation between Fourier analysis and observer ratings. Fourier calculations of images are based on a mathematical model and consequently give the same outcome. Therefore, no ICC measurements were required for this method.

**Results**

We analysed 144 sections taken from scar tissue approximately one year after surgery together with eight sections of normal skin that were obtained of the same patient population. Both the superficial and deeper dermis were considered for evaluations allowing 271 test areas.

**Inter-observer reliability for confocal and light microscopy images**

The single measurement reliability and the inter-observer reliability for combinations of two, three and four observers were established with the ICC and listed in Table 1. Overall, a higher inter-observer reliability was found for the ratings by confocal microscopy images compared to the ratings by means of conventional light microscopy. Table 1 also shows that conventional light microscopic evaluation by one or two observers has only a questionable reliability (r<0.70). The rating of collagen orientation by two observers of confocal microscopy images is good (r≥0.71). For both categories a good reliability was found using four observers: 0.80 for conventional light microscopy and 0.88 for confocal microscopy.
Correlation of Fourier analysis and observer's ratings

Fourier analysis correlated better with average ratings for confocal microscopy images (Spearman’s Rho=0.687) compared to conventional light microscopy (r=0.423). Figure 3a and 3b show the scatter plots of these data. An unexpected finding was the poor correlation between the average ratings of the four observers for light and confocal microscopy (r=0.472), which is illustrated by the wide scattering of the data in Figure 3c.

Accuracy of Fourier analysis and observer's ratings

Accuracy was evaluated by means of computer generated images with an established orientation of elliptic shapes, which is considered the ‘true’ orientation. Both the Fourier analysis and the observers evaluated the different types of shapes per category. The correlations are listed in Table 2. Overall, it is noted that both the average observer rating, which was an average rating of four observers, and the Fourier analysis showed a good to excellent correlation with the ‘true’ orientation of the different types of ellipses, except for the images that consisted of circles. The orientation assessment by Fourier analysis showed a higher correlation with the ‘true’ orientation compared to the average rating of 4 observers in general. The reliability increased when the ellipses were more elongated,

<table>
<thead>
<tr>
<th>Number of Observers</th>
<th>ICC*</th>
<th>Light Microscopy</th>
<th>Confocal Microscopy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (95%CI)</td>
<td>0.49 (0.43-0.56)</td>
<td>0.64 (0.58-0.70)</td>
</tr>
<tr>
<td>2</td>
<td>Range</td>
<td>0.58 - 0.74</td>
<td>0.71 - 0.86</td>
</tr>
<tr>
<td>3</td>
<td>Range</td>
<td>0.69 - 0.77</td>
<td>0.82 - 0.89</td>
</tr>
<tr>
<td>4</td>
<td>Mean (95%CI)</td>
<td>0.80 (0.75-0.83)</td>
<td>0.88 (0.85-0.90)</td>
</tr>
</tbody>
</table>

Table 1 Inter-observer reliability for collagen orientation assessment by observers using light microscopy and confocal microscopy images

*The Intraclass Correlation (ICC) together with the 95% confidence interval (CI) is given for the single measurement and four observers, however, for two and three observers separate ICC calculations were required for each combination, then the range of the ICC is given.
Perpectives on burn scar evaluation and artificial skin

Section 1: Tools for scar evaluation

Average observer score of confocal microscope images

Collagen orientation index by Fourier analysis

Average observer score using polarised light

3a

3b
approaching a nearly perfect assessment for the most elongated ellipses with a width/length ratio of 1:10 ($r=0.996$ for Fourier analysis). Figure 4a and b clearly illustrate the relation between the 'true' orientation and the average observer rating (Figure 4a) as well as Fourier analysis (Figure 4b) for the types of ellipses and the circle.

**Discussion**

Although the importance of collagen for the strength and pliability of skin is widely acknowledged, still no uniform evaluation method is available for collagen orientation. For research that focuses on wound healing and scar formation, polarised light is the most prevailing method for the evaluation of collagen structure\textsuperscript{14,15}. It is also frequently applied to the analysis of fibrotic lesions in other specialties, for example after myocardial infarction\textsuperscript{16,17}. The rating of collagen orientation by one or several observers has been widely used and accepted in the field. We therefore considered utilising this
method and comparing it with Fourier analysis to evaluate the accurateness of the measurements.

In the present study we first established the Intraclass correlation for inter-observer reliability of conventional light microscopy ratings. We clearly demonstrated that polarised light assessment should be performed by at least three observers in order to have considerable level of reliability ($r \geq 0.70$). We found that scoring by one or two observers' results in unacceptably low reliability.

The observers had to rate collagen orientation from the same areas after the sections were scanned by confocal laser-scanning microscopy. A higher Intraclass correlation coefficient was established for the confocal microscope image assessment by four observers compared to the polarised light assessment (0.88 and 0.80, respectively). This might be explained by the higher resolution of confocal microscope images, where out-of-focus information is rejected, compared to conventional light microscopy. No inter-observer reliability was needed for the Fourier analysis since this method is based on mathematical algorithms with no variability. The Fourier analysis has therefore a perfect reproducibility.

How does the collagen orientation assessment of the Fourier analysis relate to the observer ratings? The collagen orientation index correlated reasonably well with average confocal microscopy image ratings and poorly with the average polarised light microscopy ratings. The lower correlation between the Fourier analysis and the average observer ratings of polarised light is related to the differences in image processing between conventional light microscopy and polarised light microscopy.

<table>
<thead>
<tr>
<th>Correlation Coefficient (Spearman's Rho)</th>
<th>Fourier analysis versus 'true' orientation</th>
<th>Average rating of four observers versus 'true' orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ellipse 1:10</td>
<td>0.996</td>
<td>0.961</td>
</tr>
<tr>
<td>Ellipse 1:5</td>
<td>0.988</td>
<td>0.816</td>
</tr>
<tr>
<td>Ellipse 1:2</td>
<td>0.885</td>
<td>0.831</td>
</tr>
<tr>
<td>Perfect circle</td>
<td>-0.103</td>
<td>0.398</td>
</tr>
</tbody>
</table>

*Table 2: Accuracy of Fourier analysis and four observers with respect to orientation assessment*
Morphometry of dermal collagen orientation by Fourier analysis is superior to multi-observer assessment

Figure 4 Scatterplots of observer scores and Fourier analysis for computer generated images
and confocal microscopy. Interestingly, there was a poor correlation between observer ratings of the conventional light microscopy images and of the confocal microscopy images ($r=0.47$). This poor correlation, together with the superior inter-observer reliability for collagen orientation assessment by confocal microscopy images compared to polarised light images, may be explained by two factors: 1) the quality of the image; 2) the variance of the measured areas for polarised light evaluation.

The structure of the dermis is reproduced in a different and more distinct manner compared to polarised light by using confocal microscopy. This is mainly due to the fluorescent properties of eosin, which was used to display collagen. In this way the interference of other structures was avoided. The normal fluorescent microscope, like confocal microscopy, might therefore also give more reliable results than conventional light microscopy with polarised light. In light microscopy images using polarised light, the interpretation of collagen orientation is frequently complicated by stripes perpendicular to the long axis of the fibres, also described as cross-striations, which is illustrated in Figure 1e. Cross-striation of the collagen fibres are due to periodic changes in the orientation of the fibres and are seen as stripes perpendicular to the true orientation. Alternatively, immature collagen may be omitted from the analysis as it is hardly birefringent. In contrast, confocal microscopy shows immature and mature collagen fibres thus explaining the difference in representation of the collagen. We feel that both mature and immature fibres should be analysed for assessment of dermal architecture. The lack of moderate representation of immature fibres by polarised light might therefore cause confounding in the interpretation of the results.

Another explanation for the different ratings between images of confocal microscopy and conventional light microscopy is the consequent rating of exactly the same area for the images scanned by confocal microscopy compared to conventional light microscopy. We kept the image size constant through the study. In contrast, the polarised light evaluation was done with a conventional light microscope where the observers had to select the area of interest. The ability of both methods to assess the ‘true’ orientation was eval-
Morphometry of dermal collagen orientation by Fourier analysis is superior to multi-observer assessment because no samples of scar tissue or normal skin are available with a known 'true' orientation. Therefore, computer images were developed that roughly mimic dermal architecture. The simplicity of this model made it possible to show that assessment of the orientation present in the images is not only related to the orientation itself but also depends on the width/length ratio of the structures that are present. Some images contain perfect circles. In that case both Fourier analysis and observers will always identify a random orientation, despite the 'true' orientation assigned by the computer model. This phenomenon is illustrated in Figure 4a and b. On the other hand, the assessment of the orientation of elongated ellipses (1:10 width/length ratio) was nearly perfect: \( r=0.996 \) for Fourier analysis and 0.961 for the average observer rating. Overall, the Fourier analysis showed a superior correlation with the 'true' orientation compared to observers' ratings and is therefore more accurate.

Besides the Fourier analysis other technical methods, such as X-ray diffraction\(^1\) and laser scattering\(^2\) have been applied to analyse histopathological sections. For these applications, X-rays and laser light were used to obtain a diffraction pattern caused by the orientation of dermal architecture. In correspondence with Fourier analysis, a predominant parallel orientation of dermal collagen resulted in an elliptical scatter pattern, whereas a more circular scattering was caused by randomly organised tissue. An orientation index for tissue organisation was calculated from the scatter pattern by dividing its length and width\(^3\). If all parameters are well controlled it is our belief that the results of both techniques correspond to the collagen orientation index obtained by the Fourier analysis. In a pilot study, we tested laser scattering technique and Fourier analysis and compared the results with observer assessment\(^4\). We then established a poor correlation between the laser scattering results with the observer assessment and the Fourier analysis, which we related to the inability to control the direction of laser beam through the same field of interest. Others\(^5\) seem to have overcome this problem by creating fine vector plots for histological slides, which may increase the reliability of this technique.

Image analysis and morphometry of histopathologic sections gain
increasing popularity in the biomedical research. It allows (sometimes full automatic) analysis of biopsies in a short time. It remains nevertheless crucial to establish the reliability and accuracy of such image analysis techniques before they can be integrated into routine histopathologic applications. Although Fourier image analysis is not a new technique, we showed for the first time that this method is more suitable (i.e. in terms of reliability and accuracy) for morphometry of the dermal collagen orientation than multi-observer assessment. Collagen orientation assessment by observers who use conventional light microscopic methods therefore can not be considered as a ‘gold standard’ anymore. We advocate the use of Fourier analysis, which has become easily accessible by image analysis software that can run on personal computers, as it objectively quantifies the collagen orientation reliably and accurately.

Acknowledgements

We gratefully acknowledge Dr Riekie de Vet en Dr Wim Tuinebreijer for their advice concerning statistical matters, Dr Adam Angeles for his comments on the manuscript and Dr Henk van Veen and Joris Ruurda for their technical assistance.

This study was supported by the Technology Foundation (Stichting Technische Wetenschappen), the Dutch Burns Foundation and the Netherlands Organisation for Scientific Research.

References


