The influence of measurement location on reliability of quantitative nailfold videocapillaroscopy in patients with SSc

Andrea K. Murray¹, Andy Vail², Tonia L. Moore¹, Joanne B. Manning³, Christopher J. Taylor⁴ and Ariane L. Herrick¹

Abstract

Objectives. Nailfold videocapillaroscopy is being increasingly used as a marker of SSc-related microvascular disease, including in response to treatment. However, it requires further validation. Our aim was to assess the inter-observer, intra-observer and test-retest variability of semi-automated measurement of capillary features as well as of a manual density measurement.

Methods. All capillary apexes in images from 58 patients with SSc were marked up independently by two trained observers (inter-observer variability). The first observer then re-marked the images (intra-observer variability), and finally, the first observer marked up a second image of the same nailfold (test-retest). Mark-up of capillaries was carried out on cropped mosaic images (cropped independently by the observers to a fixed width, to allow the same length of nail bed to be studied for each patient) and on whole mosaic images (examining the whole nail bed).

Results. Reproducibility of independently cropped mosaic images was poor and was due to the variation in the positioning of the cropped area. However, quantification of whole mosaic images was highly reproducible, e.g. for inter-capillary distance, the intra-class correlation coefficient for inter-observer, intra-observer and test-retest reliability was 0.95, 0.98 and 0.90 (compared with 0.88, 0.79 and 0.89 for cropped mosaic images), respectively. Intra-observer limits of agreement for whole mosaic images were better than inter-observer reproducibility.

Conclusion. Quantitative assessment of SSc-related change in nailfold capillaries is unreliable if examination of the same set of capillaries cannot be guaranteed. Conversely a wide-field, high-magnification system that allows visualization of the whole nail bed offers a highly reproducible approach for quantitative assessment and therefore has potential as an outcome measure.

Key words: systemic sclerosis, reproducibility, nailfold videocapillaroscopy, capillary.

Introduction

Nailfold capillaroscopy is an integral part of the initial assessment of patients presenting with suspected or established SSc-spectrum disease, in particular in those presenting with RP. Normal capillaries are reassuring, but abnormal capillaries suggest underlying CTD [1]. High-magnification video microscopy, an extension of the original wide-field technique, allows measurement of capillary density and dimensions and therefore has the potential of being a non-invasive biomarker for SSc-related microvascular disease. This is because of the distinctive and identifiable changes that capillaries undergo early on in the SSc disease process.
Several investigators have described different methods to progress capillaroscopic analysis from the purely descriptive to the semi-quantitative [2–6]. The next step is to develop capillaroscopy systems/measurement methods that allow accurate, quantitative assessments. This is because for videocapillaroscopy to be a biomarker, measurements must be reliable both within and between observers. In order to facilitate quantitative assessment, we have previously developed software that allows collection and comparison of high-magnification (×300) panoramic images, then measurement of manual capillary width and density. These images capture capillaries across the whole nail bed, combining the advantage of a wide-field view with high resolution [7]. Using this software, we have performed manual measurements of capillaries [8]. However, quantifying abnormality brings enormous challenges: (i) in SSc, capillaries across any one nail bed are highly heterogeneous—many may have very abnormal architecture alongside others that look normal; (ii) measurement has a degree of subjectivity, i.e. which capillaries should be measured, and where in each capillary loop; and (iii) current measurement methods are very time-consuming, meaning that in practice only a small number of capillaries in any one nail fold is selected for study and these capillaries may be unrepresentative of the whole nailfold. To help overcome these challenges, we have developed additional software that allows rapid semi-automated quantification of inter-capillary distance and dimensions, with results comparable with manual measurements [9]. While this is likely to be a major advance, in order for the automated software to identify the region of interest containing the capillaries that are to be measured it is necessary, at present, for the observer to highlight (mark up) the apex of each capillary loop in the image. This is potentially a key issue, because the heterogeneity of SSc-spectrum nailfolds means that selection of different regions of interest could give very different results, with major implications for longitudinal studies examining change over time. This study aimed to assess the reproducibility of this user-dependent mark-up. Specifically, we assessed reproducibility of the semi-automated system in three ways (Fig. 1): inter-observer variability (different observers with the same image); intra-observer variability (same observer at different times, same image); and test-retest reliability (system variability, same subject imaged twice in quick succession, assessed by the same observer).

Due to the way the study was designed, with the initial intention to rely on cropped mosaic images only (examining a segment of the nailfold as described below), the study also provided information regarding the importance of which area/length of nail bed is chosen for analysis.

Patients and methods

Fifty-eight patients with SSc participated in the study: median age (range) 54 (18–84) years; 47 (81%) female; 41 (71%) lcSSc, 17 (29%) dcSSc [10]; 22 (38%) smokers, 7 (12%) ex-smokers. Disease duration since the first non-Raynaud’s clinical feature of disease was 9 (0–31) years; duration of RP 14 (0–64) years.

Images were obtained following 20 min acclimatization at 23°C in a low-lit, temperature-controlled laboratory. All patients had been asked to refrain from caffeine and smoking for 4 h before imaging. The study was approved by the Salford and Trafford Research Ethics Committee and all patients gave written consent.

Two panoramic capillaroscopy images (Fig. 1A and B) were obtained from the non-dominant ring finger during a single visit. Following the acquisition of Fig. 1A, the finger was taken out from under the microscope (×300 optical system magnification; KK technology, Honiton) and placed back under for Fig. 1B, therefore possibly introducing differences in the images due to the different position of the finger. Image mark-up (Fig. 1) was carried out on images that were

(i) cropped to a standard size (3 mm, in order for the same length of nailfold to be analysed for each subject) and
(ii) the whole mosaic (an image representing the whole nailfold).

Whole mosaic images were assessed because (as detailed in the ‘Results’ section) positioning of the cropped area, and therefore the location of the mark-up in the cropped mosaics, influenced reproducibility.

Cropped mosaic images

Since nailfolds vary in size according to anatomical variation, each nailfold image is of different size, representing a different nail width and including a variable number of capillaries. For the purposes of this study of reproducibility, it was predicted that it might enhance consistency to study the same length of capillary nail bed on each occasion. Therefore images were cropped to the same size, representing 3 mm of the nail bed. Each image was cropped independently by the two blinded observers before mark-up, i.e. neither observer knew the other’s cropped area and the first observer, marking up each image for a second time, did not know the original cropped area (i.e. each image was cropped once for each mark-up, as described below, also shown in Fig. 1). This was to reflect what would happen in clinical practice.

Panoramic images are composed of a series of computer screen dimension images. Cropping was carried out using automated software. The software cropped images to a length of 3.5 screen widths, representing 3 mm of the nail bed. The cropped region was selected on the basis that within that region there was at least one capillary that was likely to be trackable in the initial and subsequent mark-ups. The rationale was that this would facilitate locating the same capillary at a longitudinal repeat visit if assessment is only made over a small area of the nailfold rather than the whole length. The following steps were then performed on images

(i) The two observers blindly and independently marked the apex of each visible capillary in
image A (Fig. 1A). These data were used to assess inter-observer variability.

(ii) One of the blinded observers then performed a repeat mark-up of image A (after re-cropping); these data were used to assess intra-observer variability.

(iii) The same blinded observer also cropped and marked image B to allow comparison with those measurements made on image A; these data were used to assess test-retest reliability.

Whole mosaic images (whole nail bed) were also performed as per each step above.

Capillary feature extraction

Once images had been marked up, the semi-automated software was used to extract the following capillary features: inter-capillary distance, capillary width, tortuosity and derangement (tortuosity being a measure of the non-linearity of individual capillaries and derangement being a measure of the variance of capillary direction across the nailfold). A comprehensive description of the automation process is given elsewhere [11]. In addition, the program performed a calculation from the manual capillary mark-up to provide capillary density per millimetre (referred to, from this point onwards, as manual density). Values for inter-capillary distance, width, tortuosity and derangement are at present arbitrary, due to the way they are calculated within the program, whereas manual measurements are calibrated and therefore in real units.

Statistical analysis

Reproducibility for intra-observer, inter-observer and test-retest reliability was assessed using the Bland-Altman approach [12]. Essentially this considers the size of difference between pairs of measurements across the range of observed values. Where distributions were positively skewed, log transformation was applied. The effect
**TABLE 1** Reproducibility of measurements in cropped mosaic and whole mosaic images

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Original measurement (cropped mosaic, mean [S.D.])</th>
<th>Cropped or whole mosaic image</th>
<th>Inter-observer, Bland-Altman limits of agreement (ICC)</th>
<th>Intra-observer, Bland-Altman limits of agreement (ICC)</th>
<th>Test-retest, Bland-Altman limits of agreement (ICC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log(inter-capillary distance)</td>
<td>10.1 (0.6)</td>
<td>Cropped mosaic</td>
<td>−0.62 to 0.37 (0.88)</td>
<td>−0.85 to 0.46 (0.79)</td>
<td>−0.59 to 0.48 (0.89)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Whole mosaic</td>
<td>−0.39 to 0.27 (0.95)</td>
<td>−0.20 to 0.22 (0.98)</td>
<td>−0.50 to 0.40 (0.90)</td>
</tr>
<tr>
<td>Width</td>
<td>17 (3.9)</td>
<td>Cropped mosaic</td>
<td>−3.00 to 4.00 (0.88)</td>
<td>−2.70 to 5.10 (0.82)</td>
<td>−2.10 to 2.50 (0.96)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Whole mosaic</td>
<td>−2.10 to 2.20 (0.96)</td>
<td>−1.70 to 2.20 (0.96)</td>
<td>−2.30 to 3.00 (0.93)</td>
</tr>
<tr>
<td>Tortuosity</td>
<td>3.3 (0.1)</td>
<td>Cropped mosaic</td>
<td>−0.11 to 0.12 (0.88)</td>
<td>−0.13 to 0.15 (0.83)</td>
<td>−0.11 to 0.11 (0.85)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Whole mosaic</td>
<td>−0.08 to 0.07 (0.93)</td>
<td>−0.06 to 0.06 (0.96)</td>
<td>−0.09 to 0.09 (0.89)</td>
</tr>
<tr>
<td>Derangement</td>
<td>10.9 (5.5)</td>
<td>Cropped mosaic</td>
<td>−9.30 to 7.50 (0.64)</td>
<td>−12.80 to 8.80 (0.46)</td>
<td>−7.10 to 7.70 (0.74)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Whole mosaic</td>
<td>−5.10 to 4.60 (0.88)</td>
<td>−3.70 to 4.40 (0.91)</td>
<td>−7.00 to 6.90 (0.72)</td>
</tr>
<tr>
<td>Manual density</td>
<td>6.5 (2.1)</td>
<td>Cropped mosaic</td>
<td>−2.20 to 2.30 (0.85)</td>
<td>−1.40 to 2.00 (0.89)</td>
<td>−3.00 to 3.00 (0.80)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Whole mosaic</td>
<td>−2.00 to 1.00 (0.92)</td>
<td>−1.10 to 0.80 (0.98)</td>
<td>−1.70 to 1.50 (0.93)</td>
</tr>
</tbody>
</table>

Bland-Altman limits of agreement (ICC) for semi-automated measurements (inter-capillary distance, width, tortuosity and derangement), measured in arbitrary units and manually measured density, measured per millimetre.

of this is to consider the percentage difference between pairs of measurements across the range.

**Results**

The reproducibility of independently cropped mosaic images was poor (Table 1). Further investigation indicated that this was due to the positioning of the cropped area, and therefore the location of mark-up (Fig. 2). Overlaying the capillary mark-ups onto the images showed that the same area was not being selected, i.e. the identifiable/trackable capillary was not the same for each observer or for the same observer in the same image or in the repeat images. Therefore only a portion of the same capillaries were being measured. When images were categorized it became clear that those with least agreement had a low degree of overlap.

Following mark-up of whole mosaic images (Fig. 3), quantification of whole mosaic images was more reproducible than that of cropped mosaic images (Table 1): limits of agreement were tighter and intra-class correlation coefficients (ICCs) higher than for the cropped mosaic images for all inter- and intra-observer measurements. For test-retest, all limits of agreement were tighter for whole mosaic images except for width. ICCs were higher for all but width and derangement.

In whole mosaic images, for semi-automated measurements, intra-observer limits of agreement were narrower than those for inter-observer (with ICC being better or the same), and both (intra- and inter-observer limits of agreement and ICC) were better than test-retest reproducibility. For manual density, reproducibility was again best for intra-observer data, with test-retest and inter-observer measurements giving very similar results.

**Discussion**

The discovery that studying different areas of the nailfold gave very different quantitative results was not our primary objective. Rather this was an unexpected finding subsequent to our investigation of poor reproducibility with cropped mosaic images. Using cropped mosaic images, we were unable to select consistently under blinded conditions the same capillary or set of capillaries in different images of the same nail bed and therefore to standardize the area of mark up. As a result, the only way to proceed was to mark up the whole nail bed, which was possible with our software and is a valid approach as confirmed by our results.

This finding—that quantitative assessment of SSc-related change in nailfold capillaries is reliable only if examination of the same set of capillaries can be guaranteed—is a key message. Without this guarantee, any reports of improvement or deterioration in capillary appearances over time have to be treated with caution. Conversely, our results with whole mosaic images suggest that if the whole nailfold is examined, then quantitative nailfold videocapillaroscopy is highly reliable for all parameters other than derangement. Test-retest reliability was lower, demonstrating that the acquisition of a new image does introduce some variability (system variability), and although this is minimal, this does need to be taken into account in longitudinal studies that involve taking repeated images over time. By extrapolation, it will also be important in qualitative/descriptive studies to ensure that
the same section of nail bed is being examined in longitudinal studies.

This study builds upon previous studies in which we examined the reproducibility of manual measurements using initially single images [13], and more recently the panoramic mosaic [8]. When all of the capillaries in a 3 mm width were measured using Capiflow software to give a measure of density and width [13], the lower and upper values for intra-observer limits of agreement (for total width) were $-39/+/49\%$ and the corresponding values for inter-observer variability were $-29/+/38\%$. When we manually measured the total width of the first five capillaries on the left of the image [8], the lower and upper values for intra-observer limits of agreement (again for total width) improved from $-20/+/18\%$ to $-7/+/6\%$ when the same capillaries were measured. Similarly, lower and upper values for inter-observer limits of agreement improved from $-13/+/32\%$ to $-7/+/9\%$ [8], demonstrating the importance of measuring the same capillaries in the image. Our new study builds upon these previous studies by demonstrating that with a semi-automated measurement system, quantitative measurement of capillaries across the whole nail bed gives improved reproducibility as compared with over a subsection of nail bed by removing the possibility of measuring different capillaries. More recently, we have examined the reproducibility of a combination of quantitative and qualitative parameters with our manual measurement system [14], as it seems likely that these are to some extent complementary and that the ideal scoring system will incorporate both: six blinded observers from three centres found that qualitative parameters such as tortuosity and haemorrhages were moderately to substantially reliable (weighted $\kappa > 0.68$), whereas quantitative parameters (derived from the five largest capillaries) such as capillary density and width were highly reliable (ICC $> 0.92$). Other investigators have examined the reproducibility of capillaroscopy data over time or with different observers using different methods [6, 15, 16]. Different sample sizes and statistical analysis make it difficult to compare studies. However, we can place this study into the context of our previous cross-sectional and longitudinal measurements. Our previous cross-sectional study including 49 patients with SSc using the semi-automated software [9] indicated values of, for example, width: mean 13.4 (S.D. 2.5) for healthy controls, 16.9 (4.6) for primary RP and 20.0 (5.7) for SSc (arbitrary units). Our findings suggest that reproducibility values for intra- and inter-observation and for test–retest data would allow distinction between healthy controls and patients with SSc. Longitudinal studies are ongoing and measurement of changes with time will indicate whether these limits of agreement are close enough to allow measurements sensitive enough to detect longitudinal changes.

Our study did not investigate (i) the influence of two observers acquiring images or (ii) how much capillary features vary between all 10 digits. This will be the subject of future research.
Fig. 3 Limits of agreement for whole mosaic images. (A) Automated inter-capillary distance (logged scale), (B) tortuosity, (C) width, (D) derangement and (E) manual density per millimetre.
In conclusion, this study further demonstrates the heterogeneity of nailfold architecture in patients with SSc and the importance of ensuring that quantitative measurements of nailfold capillaries are carried out on the same microvessels in longitudinal studies. The novel ability to make rapid, semi-automated measurements across the whole nail bed at high magnification and to compare current and previous visits is a significant advantage, as this ensures the same capillaries are measured at each visit. Capillaroscopy is already being used to document changes in response to drug treatment and other therapies [17, 18]. However, as shown in this study, it is mandatory that the same section of the nail bed is examined. The high reproducibility of our measurements, when the whole nailfold is examined, suggests that this fast, reliable method of quantifying microvascular disease has the potential to be an outcome measure in future longitudinal studies, including those of treatment response.

Acknowledgements

Funding: This work was supported by the Raynaud’s and Scleroderma Association. A.M. was funded by a University of Manchester Stepping Stone Award.

Disclosure statement: The authors have declared no conflicts of interest.

References


