QUANTITATIVE OUTCOME MEASURES FOR SYSTEMIC SCLEROSIS-RELATED MICROANGIOPATHY –

RELIABILITY OF IMAGE ACQUISITION IN NAILFOLD CAPILLAROSCOPY

Graham Dinsdale[1], Tonia Moore[2], Neil O'Leary[3], Michael Berks[4], Christopher Roberts[3], Joanne Manning[2], John Allen[5], Marina Anderson[6], Maurizio Cutolo[7], Roger Hesselstrand[8], Kevin Howell[9], Carmen Pizzorni[7], Vanessa Smith[10], Alberto Sulli[7], Marie Wildt[8], Christopher Taylor[4], Andrea Murray[1], Ariane L Herrick[1,11]

1. Division of Musculoskeletal & Dermatological Sciences, University of Manchester, Salford Royal Hospital NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

2. Salford Royal Hospital NHS Foundation Trust, Salford, UK

3. Centre for Biostatistics, Division of Population Health, Health Services Research & Primary Care, University of Manchester, Manchester, UK

4. Centre for Imaging Sciences, Division of Informatics, Imaging & Data Sciences, University of Manchester, Manchester, UK

5. Microvascular Diagnostics, Northern Medical Physics and Clinical Engineering, Freeman Hospital, Newcastle upon Tyne, UK

6. Institute of Ageing and Chronic Disease, University of Liverpool, Liverpool, UK

7. Research Laboratory and Academic Division of Clinical Rheumatology, Dept. Internal Medicine, University of Genova, Italy

8. Department of Clinical Sciences, Section of Rheumatology, Lund University, Lund, Sweden

9. Institute of Immunity and Transplantation, University College London, Royal Free Campus, London, UK

10. Department of Rheumatology, Ghent University Hospital, Faculty of Internal Medicine, Ghent University, Ghent, Belgium

11. NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK.

Corresponding author: Dr Graham Dinsdale, C214 Clinical Sciences Building, Salford Royal Hospital, Stott Lane, Salford, M6 8HD, UK. Email: graham.dinsdale@manchester.ac.uk. Telephone: (+44) 161 206 2935

Word count (excluding title page, abstract, references, figures and tables): 2,124

Competing interests: The authors declare no conflict of interests.

Abstract

Background

Nailfold capillaroscopic parameters hold increasing promise as outcome measures for clinical trials in systemic sclerosis (SSc). Their inclusion as outcomes would often naturally require capillaroscopy images to be captured at several time points during any one study. Our objective was to assess repeatability of image *acquisition* (which has been little studied), as well as of measurement.

Method

41 patients (26 with SSc, 15 with primary Raynaud's phenomenon) and 10 healthy controls returned for repeat high-magnification (300x) videocapillaroscopy mosaic imaging of 10 digits one week after initial imaging (as part of a larger study of reliability). Images were assessed in a random order by an expert blinded observer and 4 outcome measures extracted: (1) overall image grade and then (where possible) distal vessel locations were marked, allowing (2) vessel density (across the whole nailfold) to be calculated (3) apex width measurement and (4) giant vessel count. Intra-rater, intra-visit and intra-rater inter-visit (baseline vs. 1 week) reliability were examined in 475 and 392 images respectively. A linear, mixed-effects model was used to estimate variance components, from which intra-class correlation coefficients (ICCs) were determined.

Results

Intra-visit and inter-visit reliability estimates (ICCs) were (respectively): overall image grade, 0.97 and 0.90; vessel density, 0.92 and 0.65; mean vessel width, 0.91 and 0.79; presence of giant capillary, 0.68 and 0.56. These estimates were conditional on each parameter being measurable.

Conclusion

Within-operator image analysis and acquisition are reproducible. Quantitative nailfold capillaroscopy, at least with a single observer, provides reliable outcome measures for clinical studies including randomised controlled trials.

Introduction

Nailfold capillaroscopy provides a non-invasive window into the digital microcirculation, and is a technique which is likely to see increasing use by rheumatologists with the inclusion of abnormal nailfold capillaroscopy into the 2013 American College of Rheumatology/European League Against Rheumatism criteria for systemic sclerosis (SSc) [1,2]. In clinical practice, its main application is in the identification of an underlying SSc-spectrum disorder in patients presenting with Raynaud's phenomenon or other clinical feature possibly suggestive of an underlying connective tissue disease. In research, there is increasing interest in using capillaroscopy parameters as outcome measures in clinical studies (including randomised controlled trials [RCTs]) in patients with SSc-spectrum disorders. Image annotation software allows tracking of nailfold capillaroscopic changes over time and we therefore now have the real possibility of non-invasive, capillaroscopy-derived quantitative outcome measures for SSc-related microangiopathy.

An outcome measure must be reliable. We and others have previously assessed reliability of nailfold capillaroscopy image interpretation/analysis [3-15]. However, in a real world setting (for example in an RCT when capillaroscopy images are acquired at several time points), there are factors in addition to image interpretation which can influence overall reliability. One of these is the reliability of image acquisition. In other words, how much variability between consecutive measures might be due to issues with repeatability (i.e. test re-test reliability)? Our objective was to assess repeatability (a prerequisite for outcome measure adoption) of image acquisition and measurement. This was part of a larger programme of work examining different aspects of nailfold capillaroscopy reliability (intra- and inter-observer reliability of image analysis has already been reported [15]).

Patients and methods

Patients

A total of 123 patients (101 with SSc and 22 with PRP), and 50 healthy controls (HC), were recruited into the study at Salford Royal NHS Foundation Trust, a tertiary referral centre for SSc, and imaged at a baseline visit [15]. One week post-baseline 41 patients (26 SSc, 15 PRP), and 10 HC returned for repeat imaging. All imaging sessions were carried out after 20 minutes acclimatisation at 23°C in a temperature and humidity-controlled laboratory. Participants were asked to refrain from caffeine and smoking for 4 hours before each study visit. All participants gave informed, written consent. The study was approved by the Greater Manchester East Research Ethics committee (reference: 11/NW/0444).

Image acquisition

At each study visit, panoramic nailfold videocapillaroscopy images were captured using a microscope with 300x magnification (KK Technology, Honiton, UK). Green LED illumination was used to ensure maximum capillary contrast. Images were recorded from all 10 digits, including the thumbs. Custom capture software allowed grayscale images to be recorded as panoramic mosaics across the entire nailfold (to avoid potential issues relating to any within-nailfold heterogeneity) by registering and then digitally stitching individual camera frames [3,16].

Image analysis and grading

Images from baseline and 1 week visits were assessed in a random order by a single expert observer (TM), who was blinded as to the image's origin (i.e. patient, finger, or visit). An example baseline-1 week image pair is shown in Figure 1. A custom software tool, written specifically for this study, was used to assess the images in 4 different ways: (1) the image was graded overall (normal, early, active, late[17], non-specific or two different ungradeable categories) and then, where judged possible by the observer, distal vessel locations were marked, allowing (2) vessel density (across the

whole nailfold) to be calculated. Each marked vessel was subsequently characterised, allowing (3) the apex width for each vessel to be measured (using a click-and-drag tool), and (4) the number of giant vessels across the entire nailfold to be counted. It should be noted that the observer had free choice in whether to assess the images completely or not. For example, if there were no clear vessels visible, the observer could choose to grade the image overall but elect not to try and identify or further measure any vessels.
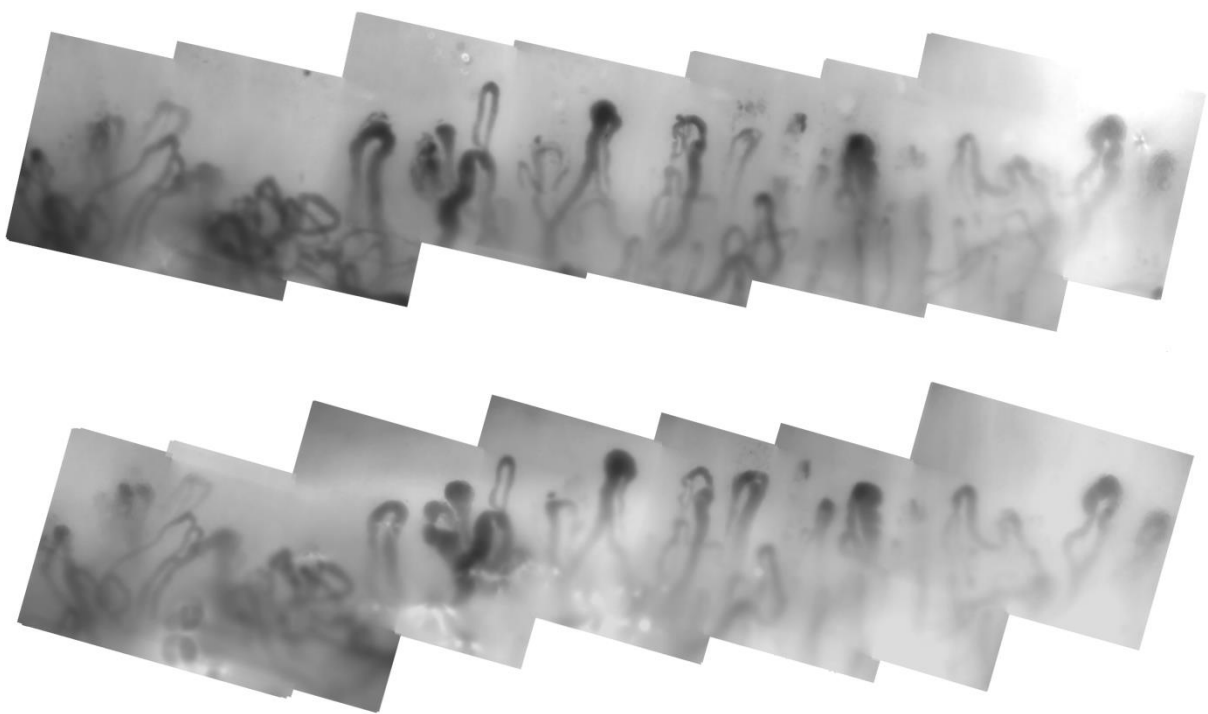

Figure 1. Example nailfold image pair from the left ring finger of a patient with SSc. Upper image taken at baseline, lower at 1 week time point.

Outcome measures and measure requirements

From the returned dataset, four image-level outcome measures were extracted or derived to allow further analysis: (1) the observer's overall grade (hereafter grade) of the image on a 4-point severity scale ranging from "normal" (assigned 1), through "early"(assigned 2) and "active" (assigned 3), to "late" (assigned 4). Images graded as "non-specific", or in either of the "ungradable categories" were not considered; (2) capillary density (hereafter density; vessels/mm) was calculated as the total number of distal vessels divided by the Euclidean distance between the leftmost and rightmost vessels marked in the image. Density was only calculated when 2 or more vessels were identified within an image; (3) mean apex width (hereafter width) was defined as the arithmetic mean of the measured apex width of all vessels measured in an image, conditional on at least one distal vessel being identified and measured; (4) presence of giant capillaries (hereafter giants) was defined as there being one or more vessels identified and subsequently categorically marked as having size/shape in the "Giant" category. The widely-accepted definition of giant vessels was used during this study: a giant capillary should have a normal shape but be uniformly enlarged along its length, with a width of greater than 50 μm.

Statistical analysis

We examined (a) the intra-rater intra-visit reliability to better understand the variability in results introduced solely by the observer, and (b) the intra-rater inter-visit (test re-test) reliability as a proxy for the reproducibility of the combination of the image acquisition and measurement process (the assumption being that within one week a participant's microvasculature will not change).

A linear, mixed-effects model was used to model each outcome measure. All models included between-patient variation with random -effects of finger nested within patient (3 level hierarchical model). All models included covariates for finger, gender and age as fixed-effects to account for systematic non-pathological differences. Reliability was assessed with intra-class correlation

coefficients (ICCs) from estimated model variance components and on a scale of 0-1. Including fixed effects of hand and fingers means these ICCs are the average reliability stratified by finger and hand. Outcome measures with reliability closer to 1 indicate a measure which has high correlation between repeat assessments (either within-visit or between-visit).

Reliability of measures can only be assessed were data is available. These analyses were therefore based on conditional models, i.e. including only images with a usable value being provided. For each outcome measure, we quantified the crude proportion of all evaluations in which the observer could provide a usable value. A high proportion indicates a good measure and reduced risk of bias in reliability estimates.

Analyses were performed using R [18] and the package lme4 [19] (R Project for Statistical Computing, Vienna, Austria).

Results

A total of 475 nailfold images from the same visit were evaluated twice by the observer; 239 of these were baseline images and 236 were 1-week images. Separately, 392 nailfolds from 51 patients were imaged and subsequently evaluated at both baseline *and* 1-week time-points. The numbers of evaluated nailfolds by the observer for this study do not directly map to patient numbers (e.g. 10 images/participant) at each time point due to the random assignment scheme used in the larger observer reliability study of which this work was a part [15].

Usable values (used for reliability estimates) were available from 53.6%, 98.4%, 91.6%, and 98.4% of image evaluations for (1) overall image grade, (2) mean capillary density, (3) mean apex width, and (4) presence of giant capillaries respectively.

The intra-rater reliability estimates for intra-rater intra-visit and intra-rater, inter-visit (test re-test) are detailed in Table 1 and were (respectively) as follows (ICCs): overall image grade, 0.97 and 0.90;

vessel density, 0.92 and 0.65; mean vessel width, 0.91 and 0.79; presence of giant capillary, 0.68 and

0.56. These estimates were conditional on each parameter being measurable.

Table 1. Estimates of conditional intra-rater reliability (intra-visit, same image twice; inter-visit, same finger imaged twice, 1 week apart) for each image measure, from variance component estimates of mixed-effects linear regression. Estimates are conditional on each parameter being measurable.

Measure (conditional on availability)	Intra-rater, intra-visit, n=475*	Intra-rater, inter-visit, n=392[†]
Overall image grade	0.97	0.90
Mean distal vessel width	0.91	0.79
Distal vessel density	0.92	0.65
Presence of giant capillary	0.68	0.56
All reliability estimated, within a randomly chosen finger. [*]Total includes number of evaluated nailfolds with repeat baseline evaluations (n=239) and repeat 1 week evaluations (n =236). [†]Number of evaluated nailfolds with both baseline *and* 1 week evaluations.		

Discussion

Reliability in this single observer study was greater than 0.90 for 3 out of 4 measures when re-

assessing the same image (overall grade, vessel density and width), and remained good for 2 out of 4

(overall grade and width) when comparing 1-week image outcomes to those from baseline images

(test re-test reliability). This suggests that both image analysis and acquisition are reproducible, and

that nailfold capillaroscopy, at least with a single observer, provides reliable outcome measures for

clinical studies including trials of treatment response. The relatively lower reliability for the presence

of giant capillaries may be explained by the difficulty observers have in classifying vessels explicitly

as giant (normal shape, uniform enlargement >50 µm), due to subjectivity in confirming the

uniformity aspect of the definition; hence the vessel has a high chance of being alternatively-

categorised on a subsequent reading of the same image. Examining the intra-rater, intra-visit (same

image) reliability allowed us to understand the variability introduced solely by the observer, whereas examining the intra-rater inter-visit (test re-test) reliability served as a proxy for the reproducibility of the image acquisition/measurement process itself (making the reasonable assumption that within one week the structure of the microvasculature does not markedly change).

Previous studies of reliability [3-15] have tended to concentrate on image *interpretation*. Reliability of image *acquisition* has been less studied, and yet this is clearly an important element of reliability when multiple, repeat images are recorded over time. For example, in a clinical trial investigators might wish to examine the effects of a therapy on nailfold capillary morphology, acquiring images at (say) 3-monthly intervals. To assess the evidence of whether true change has occurred, we need to know what the potential for variation is between measurements/assessments at different visits, independent of any pathological changes. Inter-observer reliability of image acquisition has still to be tested, although, in a clinical trial setting, it is very likely that the same observer will be acquiring all the images at each visit. There has been previously-published work exploring 'test-re-test' reproducibility in a study of 58 patients with SSc, examining the non-dominant ring finger [9]. Only patients with SSc were included and, although manual density was measured (as in the current study), other measurements were semi-automated via software (inter-capillary distance, width, tortuosity, derangement). Additionally, the second nailfold examination in the previous study immediately followed the first (the patient removed his/her finger from beneath the microscope then placed it back beneath). In the 2012 Murray et al study [9], the ICC (test re-test) for manually-measured capillary density (whole nailfold) was high at 0.93 compared to 0.65 in the present study, but was (as in the present study) lower than the intra-observer reliability, suggesting the presence of some 'system variability', albeit minimal. The higher apparent system variability in our present study might relate to the longer time period (one week) between images. It is possible that although on each occasion the patient was fully acclimatised, other difficult-to-control factors might influence a

second image taken one week (as opposed to minutes) after the first. For example the finger is less likely to be positioned in exactly the same way, potentially leading to different margins on the mosaic. Other factors which could potentially have an effect on image similarity over a 1 week time period include (changes in) medication, hydration and outdoor temperature. Sekiyama et al [10] examined intra- and inter-observer reliability for both widefield microscopy (stereomicroscope 10-25x magnification) and for videocapillaroscopy (200x magnification) in a study which involved 20 subjects being examined on two separate occasions by two observers: ICCs for intra- and inter-observer reliability for number of capillaries per mm were 0.99 and 0.95 respectively for videocapillaroscopy, and 0.99 and 0.95 respectively for widefield capillaroscopy [10]. These intra-observer reliabilities (which therefore included 'system variability) were higher than in our study, although the numbers of subjects examined was lower (and the diagnoses of the 20 subjects were not stated). Hudson et al in an earlier study [6] examined test-retest reliability of nailfold capillary density as measured by widefield microscopy (50x magnification, 5 patients with SSc): each of the two observers photographed the patients once in the morning and again in the afternoon. Interestingly, intra- and inter-rater reliabilities were much higher when the observers examined the same photographs or photographs taken at the same time of day (ICCs 0.73-0.74), than when agreements between (same rater) morning and afternoon assessments were examined (ICCs 0.30-0.34). An important point was that patients were not repositioned between the two morning assessments by the two observers, or between the two afternoon assessments, but were repositioned between morning and afternoon assessments. Capillaries are themselves invisible: what are imaged during capillaroscopy are columns of red blood cells, which can sometimes 'appear' and 'disappear' over short time periods depending on local perfusion state. This adds to the challenges of both qualitative and quantitative assessment.

A limitation of our study was that test-retest reliability was examined with one observer only, although conversely the number of images examined was high. Our findings should be placed in the context of the wider reliability study of which this formed a part [15]. Also, our findings are specific to the microscope system/magnification (300x) used: further work is required to assess relative reliabilities of different magnifications.

In conclusion, our results suggest that although there is a degree of system variability inherent in repeat measurement, both image analysis and acquisition are reproducible, at least with a single observer, and that quantitative nailfold capillaroscopy can therefore provide reliable outcome measures for clinical studies. Increased usage of automated image analysis, and its application to capillaroscopy [20], may help to ensure consistent reliability within and between studies, as well as having the advantage of removing the time-consuming burden of such analysis from busy clinicians.

REFERENCES

1. Van den Hoogen F, Khanna D, Fransen J, et al. 2013 Classification criteria for systemic sclerosis. Arthritis Rheum 2013; 65: 2737-47.

2. Van den Hoogen F, Khanna D, Fransen J, et al. 2013 Classification criteria for systemic sclerosis. Ann Rheum Dis 2013; 72: 1747-55.

3. Anderson ME, Allen PD, Moore T, et al. Computerized nailfold video capillaroscopy - a new tool for assessment of Raynaud's phenomenon. J Rheumatology 2005; 32: 841-8.

4. Ingegnoli F, Gualtierotti R, Lubatti C, Zahalkova L, Meani L, Boracchi B, et al. Feasibility of different capillaroscopic measures for identifying nailfold microvascular alterations. Semin Arthritis Rheum 2009; 38: 289-95.

5. Smith V, Pizzorni C, De Keyser F, et al. Reliability of the qualitative and semiquantitiave nailfold videocapillaroscopy assessment in a systemic sclerosis cohort: a two-centre study. Ann Rheum Dis 2010; 69: 1092-6.

6. Hudson M, Masetto A, Steele R, et al. Reliability of widefield capillary microscopy to measure nailfold capillary density in systemic sclerosis. Clin Exp Rheumatol 2010; 28 (5 Suppl 62): S36-S41.

7. Hofstee HMA, Serne EH, Roberts C, et al. A multicentre study on the reliability of qualitative and quantitative nail-fold videocapillaroscopy assessment. Rheumatol 2012; 51, 749–755.

8. Gutierrez M, Bertolazzi C, Tardella M, et al. Interreader reliability in assessment of nailfold capillary abnormalities by beginners: pilot study of an intensive videocapillaroscopy training programme. J Rheumatol 2012; 39: 1248-55.

9. Murray AK, Vail A, Moore TL, et al. The influence of measurement location on reliability of quantitative videocapillaroscopy in patients with SSc. Rheumatol 2012; 51 : 1323-30.

10. Sekiyama JY, Camargo CZ, Eduardo L, et al. Reliability of widefield nailfold capillaroscopy and videocapillaroscopy in the assessment of patients with Raynaud's phenomenon. Arthritis Care Res 2013; 65: 1853-61.

11. Overbury R, Murtaugh MA, Fisher A, et al. Primary care assessment of capillary abnormalities in patients with Raynaud's phenomenon. Clin Rheumatol 2015; 34: 2135-40.

12. Cheng C, Lee CW, Deskalakis C. A reproducible computerized method for quantitation of capillary density using nailfold capillaroscopy. J Vis Exp 2015; 105: doi: 10.3791/53088.

13. Hughes M, Moore T, O'Leary N, et al. A study comparing videocapillaroscopy and dermoscopy in the assessment of nailfold capillaries in patients with systemic sclerosis-spectrum disorders. Rheumatology 2015; 54: 1435-42.

14. Smith V, Beeckman S, Herrick AL, et al. An EULAR study group pilot study on reliability of simple capillaroscopic definitions to describe capillary morphology in rheumatic diseases. Rheumatol 2016; 55: 883-90.

15. Dinsdale G, Moore T, O'Leary N, et al. Intra-and inter-observer reliability of nailfold videocapillaroscopy - a possible outcome measure for systemic sclerosis-related microangiopathy. Microvascular Res 2017; 112: 1-6 [Epub ahead of print].

16. Allen PD, Taylor CJ, Herrick AL, et al. Image analysis of nailfold capillary patterns from video sequences. Proceedings of Medical Image Computing and Computer-Assisted Intervention, Springer, 1999; 698-705.

17. Cutolo M, Sulli A, Smith V. How to perform and interpret capillaroscopy. Best Practice Res Clin Rheumatol 2013; 27: 237-48.

18. R Core Team. R: A Language and Environment for Statistical Computing [computer program]. Vienna, Austria: Foundation for Statistical Computing, 2013. http://www.r-project.org/.

19. Douglas Bates, Martin Mächler, Ben Bolker, Steve Walker. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 2015; 67(1): 1-48.

20. Berks M, Tresadern P, Dinsdale G, et al. An automated system for detecting and measuring nailfold capillaries. Medical Image Computing & Computer-Assisted Intervention: MICCAI. 2014; 17: 658-65.