Infrared thermography (IRT) for the assessment of microvascular skin blood flow in a specialist connective tissue disease unit.

Kevin J. Howell

A submission presented in partial fulfilment of the requirements of the University of Glamorgan / Prifysgol Morgannwg for the degree of Doctor of Philosophy.

October 2009
# Table of contents.

Certificate of Research...................................................................................................7
Declaration......................................................................................................................8
Acknowledgements.........................................................................................................9
Abstract.........................................................................................................................11
Glossary and abbreviations............................................................................................14
List of figures..................................................................................................................16
List of tables...................................................................................................................17
1: Introduction..................................................................................................................18
   1.1 Aim.......................................................................................................................20
   1.2 Contribution to knowledge...................................................................................22
      1.2.1 A thermographic service for RP assessment..............................................22
      1.2.2 IRT for assessment of disease activity in LS..............................................22
      1.2.3 Developing good practice of IRT through appropriate imager procurement
           and quality assurance. .......................................................................................23
2: A thermographic service for Raynaud’s phenomenon assessment..........................24
   2.1 Introduction...........................................................................................................24
      2.1.1 Background. ...............................................................................................24
      2.1.2 Aim of the project .......................................................................................26
   2.2 Literature review....................................................................................................27
   2.3 Projects..................................................................................................................31
      2.3.1 IRT with cold challenge for the detection of RP of the feet in females.....31
      2.3.2 IRT with hand cold challenge to assess vasodilator therapy for RP in clinical
           trials.....................................................................................................................32
      2.3.3 An infrared thermometer technique, validated against IRT, for the assessment
           of hand cold challenge in a twin study of the heritability of RP.................34

3.1 Introduction.

3.1.1 Background.

3.1.2 Aim of the project.

3.2 Literature review.

3.3 Projects.

3.3.1 Simple inspection of thermograms for the detection of clinically active LS.

3.3.2 A protocol for assessment of skin temperature and microvascular blood flow in LS.

3.3.3 IRT and LDF for the assessment of disease activity in LS.

4: Developing good practice of IRT through appropriate imager procurement and quality assurance.

4.1 Introduction.

4.1.1 Background.

4.1.2 Aim of the project.

4.2 Literature review.

4.3 Projects.

4.3.1 Specifying and validating a portable in field of view (IFOV) temperature source for the calibration of medical IRT.

4.3.2 Guidelines for procuring and commissioning a medical thermal imager.

5: Discussion.

5.1 Assessment of contribution to knowledge of projects.

5.1.1 A thermographic service for Raynaud’s phenomenon.

5.1.2 Thermographic assessment of disease activity in LS.

5.1.3 Appropriate imager procurement and quality assurance.

5.2 Significance of contribution to knowledge.
6: Conclusions..............................................................................................................69

6.1 Fulfilling the thesis aims. ...................................................................................69

6.2 Future work. .......................................................................................................71

7 : Bibliography ...........................................................................................................73

Appendices:


Certificate of Research.

This is to certify that, except where specific reference is made, the work presented in this thesis is the result of investigation undertaken by the candidate.

Candidate .............................................

Director of Studies ......................................
Declaration.

This is to certify that neither this thesis nor any part of it has been presented or is being currently submitted in candidature for any other degree other than the degree of Doctor of Philosophy at the University of Glamorgan.

Candidate ........................................
Acknowledgements.

I am indebted to Dame Prof. Carol Black and Prof. Chris Denton, Consultant Rheumatologists at the Royal Free Scleroderma Unit, for their steadfast support of the Royal Free thermography service since I took over its management in 1992. Equally, I express my gratitude to Dr. Roy Smith, Head of Royal Free Medical Electronics, and his staff. The thermography lab at the Royal Free could not function without their support and extensive expertise on all medical equipment matters. The Scleroderma Unit and Medical Electronics Department contributed equal financial support to fund my doctoral studies. Special thanks are also due to Prof. Francis Ring at the University of Glamorgan. There is nobody in the world with more experience of the discipline of medical infrared thermography than Francis, and I have been fortunate to have the benefit of his support and advice from the beginning of my career.

I thank my Director of Studies Dr. Peter Plassmann for his infectious enthusiasm for my thesis, his unfailing good humour, and his thorough knowledge of what one must demonstrate to be worthy of the title “Doctor.”

The attached appendices also highlight how privileged I have been to work with such a diverse group of medical researchers. None of the appendices has been a solo effort, and I ask the reader to note the significant contributions of these talented people to my portfolio of publications.

Louise Parker and Dr. Svetlana Nihtyanova at the Royal Free have been great supporters of the thermography service for Raynaud’s phenomenon assessment.

I am grateful to Dr. Geoff Boxer (UCL Cancer) for sharing with me his experiences of the submission of a PhD. “by portfolio,” and to Dr John Allen (Freeman Hospital) for many helpful comments on the manuscript.
A word or two also for my mother, who has been a wholehearted supporter of every academic pursuit I have undertaken. Her home has been an environment conducive to clear thinking and coherent writing. I hope that the reader feels able to agree.

Thank you to the many friends who have taken a close interest in my progress during my studies, and provided support and motivation. Evie Lieberherr and Katherine Horrell are worthy of special mention.

Finally, I thank the patients who have given their time freely and willingly to contribute to the research which is presented herein.
Abstract.

**Aim.** To establish standardised infrared thermography (IRT) within a specialist connective tissue disease unit, assessing its utility:

- for the evaluation of Raynaud’s Phenomenon (RP) in clinical rheumatology and research
- for the detection of active localised scleroderma (LS) lesions in paediatric patients

and to develop improvements in IRT quality assurance for these medical applications.

**Methods.** For the evaluation of RP, a protocol for cold challenge of the feet was developed and validated. IRT was applied with hand cold challenge for the assessment of response to oral vasodilator therapies in two large randomised pilot studies. An infrared thermometer technique was developed, validated against IRT, and subsequently used for the assessment of peripheral vasospasm in a twin study into the heritability of RP.

The utility of inspection of thermograms for detecting clinically active LS lesions was established. A protocol was developed incorporating photography, IRT and laser Doppler flowmetry (LDF) for LS assessment, and the normal range of temperature and LD blood flow across several body sites was established in adults and children. The utility of the protocol for assessing LS activity in children was investigated.

To develop quality assurance of thermography, the author contributed to the specification and validation of blackbody medical temperature reference sources, and published guidelines for procuring and commissioning a medical thermal imager.

**Results.** Healthy controls had a higher mean toe temperature than RP patients (at baseline 29.2 ± 1.5°C v 24.8 ± 1.5°C [mean ± SD], p<0.01; t-test). IRT demonstrated improved finger rewarming 10 minutes after cold challenge in primary RP patients.
treated with fluoxetine compared with those treated with nifedipine (58.8% v 43.1%, p=0.03; t-test). IRT showed no such improvement in finger rewarming over nifedipine in patients treated with losartan. In a hospital setting, an infrared thermometer technique performed similarly to IRT with cold challenge for the detection of RP: the sensitivity of IRT was 83%, whereas for the infrared thermometer it was 89%. The specificity of both instruments was 84%. In a population setting using the infrared thermometer both baseline finger temperature and rewarming after ten minutes were significantly lower for RP subjects than for controls (for baseline: 28.3ºC v 30.0ºC, p<0.01, t-test; for rewarming: 4.6ºC v 5.3ºC, p<0.05, t-test). Infrared thermometer measurements in monozygotic and dizygotic twin pairs revealed a heritability of 65% for baseline finger temperature, 35% for fall after cold challenge, and 24% for rewarming over ten minutes.

In the larger of two published studies on the inspection of thermograms for detecting clinically active LS, sensitivity was 92%, and specificity was 68%. In lesions imaged within 2 years of onset, sensitivity was 81% and specificity 88%. Validation of a protocol combining IRT and LDF measurements revealed that, in adult controls, the mean temperature difference between the two sides of the body was less than 0.5ºC at all body regions. Mean differences in contralateral LD flux were less than 40% at all body sites. Variability in LD and IRT readings due to experimental factors was acceptably small in comparison to the physiological differences recorded. Applying the protocol in children with LS, the median relative increase in LD blood flow in clinically active lesions (compared with blood flow in contralateral unaffected skin) was 89% (range -69% to +449%), whereas the median flow increase in clinically inactive lesions was 11% (range -46% to +302%), p<0.001. Using IRT, the median temperature difference between clinically active lesions and contralateral unaffected skin was 0.5ºC (range -0.1ºC to +4.1ºC), whereas the median temperature difference for clinically inactive plaques was 0.3ºC (range -1.9ºC to +2.7ºC), p=0.024.
In hand cold challenge measurements made at the Royal Free Hospital, application of the medical blackbody temperature reference sources reduced the overall uncertainty in temperature readings by a factor of about 4, from typically ±2°C to ±0.5°C.

**Conclusion.** IRT or infrared thermometer data on skin temperature before and after cold challenge affords RP studies an important element of objectivity. RP detected in a population setting exhibits milder vasospasm than RP recruited from hospital patients, and thus the results of research performed at specialist centres may not be translatable to community settings.

Inspection of thermograms is an effective method for the detection of clinically active LS, although LDF performed better than IRT using a protocol reliant on objective measurements from small regions of interest.

IRT and infrared thermometry were generally less effective at discriminating between healthy and diseased subjects in situations where the temperature difference between groups was small (<2°C). The introduction of temperature reference sources, which reduce uncertainty in radiometric measurements to the order of 0.5°C, would increase the utility of IRT in settings where the temperature change associated with disease is small.
Glossary and abbreviations.

ACE: Angiotensin converting enzyme
CGRP: Calcitonin gene-related peptide
CTD: Connective tissue disease
DTGS: Deuterated tryglycerine sulphate
DZ: Dizygotic
EAT: European Association of Thermology
FLPI: Full-field laser perfusion imaging
FOV: Field of view
FPA: Focal plane array

Heritability: *The proportion of phenotypic variation in a population that is attributable to genetic variation among individuals*

IFOV: In field of view, or instantaneous field of view

Inter-operator reproducibility: *the variation in measurements taken by different persons or instruments on the same item and under the same conditions*

Intra-operator reproducibility / repeatability: *the variation in measurements taken by single persons or instruments on the same item and under the same conditions*

IR: Infrared
IRT: Infrared thermography
ITS-90: International temperature scale of 1990
LD: Laser Doppler
LDF: Laser Doppler flowmetry
LDPI: Laser Doppler perfusion imaging
LS: Localised scleroderma (*morphoea*)
MTBF: Mean time between failures
MTF: Modulation transfer function
MZ: Monozygotic
NETD: Noise equivalent temperature difference
NHS: National Health Service
NPL: National Physical Laboratory
PEV: Pyroelectric vidicon
PGE\textsubscript{1}: Prostaglandin E\textsubscript{1}
PGI\textsubscript{2}: Prostaglandin I\textsubscript{2}
Prevalence: the total number of cases of disease in a statistical population at a given time
PTFE: Polytetrafluoroethylene
QA: Quality assurance
Radiance: radiometric measure of the amount of light emitted from a particular area, falling within a given solid angle in a specific direction. Units $\text{W} \text{sr}^{-1} \text{m}^{-2}$
ROC: Receiver operator characteristic
RP: Raynaud’s phenomenon
Sensitivity: the proportion of positives which are correctly identified
Specificity: the proportion of negatives which are correctly identified
SSc. : Systemic sclerosis (scleroderma)
SSRI: Selective serotonin reuptake inhibitor
TI: Thermographic index
Uncertainty: the region about an observed value of a physical quantity which is likely to enclose the true value of that quantity (specified according to the ISO Guide to the Expression of Uncertainty in Measurement [GUM])
List of figures

Fig. 1.1 Examples of medical infrared thermograms captured by an uncooled focal plane array (FPA) imager.

Fig. 1.2 Structure of the thesis.

Fig. 2.1 Colour changes in the fingers during an episode of Raynaud’s phenomenon.

Fig. 2.2 Thermogram of the hands in RP 10 minutes after cold challenge in water at 15°C for one minute. The fingers remain cold, indicating peripheral vasospasm.

Fig. 3.1 An early-stage active LS lesion with inflamed skin.

Fig. 3.2 Late-stage LS, with growth failure and contracture of the affected limb.

Fig. 3.3 Example of a “thermography positive” lesion at the neck, which is “thermography negative” after five months of treatment.

Fig. 4.1 The “standard protocol”: components for quality assured medical IRT.

Fig. 4.2 Mean finger temperature during a cold challenge procedure, along with actual and recorded IFOV source temperatures.

Fig. 4.3 Calibration curve for the A320G thermal imager using the IFOV sources at the time point prior to cold challenge.

Fig 4.4 Mean finger rewarming before and after calibration with IFOV sources.
List of tables

Table 4.1: Specifications for three thermal imagers used at the Royal Free Hospital.

Table 5.1 Mean and range of baseline finger temperature, temperature drop and rise after cold challenge for RP and non-RP subjects in hospital and population settings.

Table 5.2 Citations of the work included within the thesis.
1: Introduction.

Medical infrared thermography (IRT) is the imaging of human skin temperature distribution by detection of infrared radiation using a thermal camera (fig. 1.1). The application of IRT in a medical context can be fundamentally different from its use in engineering because

- measurements are often made at near-ambient temperatures across a narrow range (<20°C).
- the images are used to assist with patient diagnosis or in the evaluation of treatment.

Whilst IRT has been employed for investigating many aspects of thermal physiology since the 1960s, the technical limitations of IRT in medical use have been less widely considered.

![Fig.1.1: Examples of medical infrared thermograms captured by an uncooled focal plane array (FPA) imager.](image)
Modern focal plane array (FPA) thermal imagers are designed for engineering applications such as condition monitoring. Correctly applying such industrial technology for physiological measurement poses challenges which are not always appreciated by medical practitioners. The potential for misuse of IRT in medicine consequently remains high.

For example, the absolute temperature accuracy of medical thermal imagers is typically quoted as no better than ±2°C. This is approximately 10% of the temperature range encountered in most medical measurements (20 – 40°C). Imagers with uncooled detectors also exhibit a drift in temperature measurement accuracy after switching on [1]. The time for the detector to reach stability can range from a few minutes to more than an hour.

Such limitations reduce confidence in the reproducibility of thermographic studies across different laboratories. Early multi-centre trials involving thermography have indeed proved problematic and inconclusive [2], casting doubt on the validity of IRT in medicine.

IRT nonetheless has an established track record in rheumatology for the assessment of microvascular skin blood flow in connective tissue diseases (CTDs). Raynaud’s phenomenon (RP) is often the earliest manifestation of microvascular dysfunction in CTDs such as scleroderma (systemic sclerosis, SSc.). The thermographic assessment of the rewarming of fingers after a standardised “cold challenge” of the hands is a well-established test for RP, but its utility is still not fully understood. In particular:

- there are few reports on the use of IRT for the assessment of vasodilator therapies in RP. Published studies give conflicting accounts of the correlation between symptomatic improvement in patients and objective improvement in the response to cold challenge.
• vasospasm of the toes is often described by RP patients, but no published data exist on cold challenge of the feet.
• the utility of the cold challenge test in a population setting remains unclear.

An inflammatory component is common to all CTDs and, where this inflammation affects the dermis, IRT may be an effective tool for the assessment of disease activity. Nonetheless, only two brief reports [3,4] have been published to date on IRT in the evaluation of localised scleroderma (LS), a CTD where inflammation and fibrosis is limited to the skin and underlying superficial tissues.

1.1 **Aim.**

The aim of the work reported within the thesis was to establish standardised infrared thermography (IRT) within a specialist connective tissue disease unit, assessing its utility:

• for the evaluation of Raynaud’s Phenomenon (RP) in clinical rheumatology and research
• for the detection of active localised scleroderma (LS) lesions in paediatric patients

and to develop improvements in IRT quality assurance for these medical applications.

The structure of the thesis is outlined in figure 1.2.

The standardisation of IRT is discussed in chapter 4 and encompasses factors such as patient preparation, image capture protocols and image analysis, reporting and archiving. The particular emphasis within this thesis will be on the standardisation of thermal imager performance through quality assurance, which is a neglected area in the literature.
Ch.1: Introduction
Aims, contribution to knowledge

Ch.2: Thermography in Raynaud’s phenomenon

Ch.3: Thermography in localised scleroderma

Ch.4: Thermography quality assurance

Introduction: Background and aims

Literature review

Project 1: Raynaud’s phenomenon in feet

Project 1: Inspection of thermograms to detect active LS

Project 1: Validation of in field of view calibration sources

Project 2: Thermography for assessment of vasodilators

Project 2: A novel protocol to measure skin temperature and blood flow

Project 2: Guidelines for procuring and commissioning a thermal imager

Project 3: IR thermometer method in population studies

Project 3: Thermography and LDF combined to detect active LS

Ch. 5: Discussion. Assessment of contribution to knowledge of projects

Ch. 6: Conclusion. Fulfilling the thesis aims and future work

Fig. 1.2: Structure of the thesis.
1.2 Contribution to knowledge.

The thesis claims contributions to knowledge in the field of IRT outlined in sections 1.2.1, 1.2.2, and 1.2.3 below.

1.2.1 A thermographic service for RP assessment.

- Established a protocol for cold challenge of the feet, and demonstrated significant differences in toe temperature between healthy controls and RP subjects.
- Applied IRT with hand cold challenge for the assessment of response to novel oral vasodilator therapies in RP in two large randomised pilot studies.
- Developed and validated against IRT an inexpensive infrared thermometer technique, subsequently used for the assessment of peripheral vasospasm in a twin study into the heritability of RP.

1.2.2 IRT for assessment of disease activity in LS.

- Established the utility of IRT for the detection of clinically active LS based on simple inspection of thermograms.
- Developed a protocol for skin assessment incorporating photography, IRT and laser Doppler flowmetry (LDF).
- Established the normal range of readings for temperature and LD blood flow across several body sites in adults and children.
- Compared the utility of IRT and LDF for assessing disease activity in children with LS.
1.2.3 Developing good practice of IRT through appropriate imager procurement and quality assurance.

- Contributed to the specification of portable blackbody temperature sources for quality assurance of medical IRT. Validated the devices in a hospital “field trial.”

- Produced guidelines for procuring and commissioning medical thermal imagers, with emphasis on European health and safety legislation.
2: A thermographic service for Raynaud’s phenomenon assessment.

2.1 Introduction.

2.1.1 Background.

Maurice Raynaud first described episodic skin vasospasm of the digits on exposure to cold in 1862 [5]. Raynaud’s phenomenon (RP) manifests as colour change in fingers (fig. 2.1) or toes, accompanied by numbness and paraesthesia [6].

Fig. 2.1: Colour changes in the fingers during an episode of Raynaud’s phenomenon

RP occurs most commonly as an idiopathic condition, whereby it is termed primary RP. The prevalence of primary RP has been reported between 3% and 21% [7], with the highest prevalence in cool climates [8]. Females report RP more frequently than males [9].
Rarely, RP presents secondary to associated medical disorders, the most significant of which are the autoimmune connective tissue diseases. The prevalence of RP within these small CTD patient populations can be high: RP is present in more than 90% of systemic scleroderma cases, and in up to 45% of systemic lupus cases [8]. RP may develop in such patients a number of years before any connective tissue symptoms are evident. Antinuclear antibodies and abnormal nailfold capillaries are helpful prognostic indicators of progression to such diseases [10].

The pathophysiology of RP is poorly understood [11]. In primary RP the vasospasm is considered to be reversible, arising from an abnormal microvascular physiological response to cooling. In secondary RP anatomical vascular changes also play a major role, with irreversible vessel occlusion and damage to the capillary wall. Consequently secondary RP patients may exhibit not only the symptoms of poor thermoregulatory blood flow, but also signs of failure of peripheral skin nutritional circulation such as fingertip ulceration or necrosis.

A variety of environmental factors may also contribute to RP, and genetic susceptibility to cold intolerance could also play a role [12].

Treatment for RP may involve advice on lifestyle (dressing to keep warm, etc.), use of “over-the-counter” remedies (e.g. antioxidant therapies), or pharmacological intervention with vasodilators (inter alia calcium channel blockers, ACE inhibitors, prostaglandins, SSRIs) [13].

Given that the symptoms of RP are:

- subjective, but sometimes debilitating;
- rarely evident in a clinical setting (with diagnosis based primarily on patient recall);
- potentially concomitant with autoimmune diseases;
- manageable with well-tolerated drugs of variable efficacy;
quantitative methods to detect and assess the severity of peripheral vasospasm are essential in a specialist rheumatology setting. IRT can play an important role in documenting the changes in thermoregulatory skin blood flow present in RP.

2.1.2 Aim of the project.

The aim of this project was to evaluate thermographic criteria for the detection and assessment of RP across a range of clinical and research applications, namely:

- Detection of RP in the feet (Appendix 1).
- Assessment of hand cold challenge responses (fig 2.2) to two novel oral vasodilators (losartan and fluoxetine) in large, randomised, controlled pilot studies (Appendices 2 and 3).
- Detection of vasospasm in a population setting for a study of the heritability of RP in twins. An infrared thermometer was validated against IRT in a hospital setting (Appendix 4), applied for the detection of RP in a population setting (Appendix 5) and ultimately used in a twin study of RP (Appendix 6).
2.2 Literature review.

Published work pertaining to the use of IRT for RP assessment is reviewed herein, considering the literature chronologically.

The thermographic assessment of acral skin temperature began soon after thermal imagers became commercially available in the 1960s. Findings were initially restricted to qualitative descriptions of thermal patterns due to the limited image processing available.

In 1968 James [14] published upper limb thermograms recorded from control and RP subjects whilst room temperature was reduced progressively, but this form of whole-body cooling struggled to differentiate controls from RP. Chucker et al. [15] described the use of IRT and cold challenge of the hands in water for discrimination of healthy subjects from RP. They used iced water immersion for 60 seconds as the stimulus – a procedure unlikely to be well tolerated by any RP patient with significant vasospasm.
By the 1980s protocols for cold challenge had become generally more tolerable for patients with the realisation that induced temperature gradients across the hand did not need to be large for the test to be effective. Kyle et al. [16] reported qualitative thermographic differences between primary RP, secondary RP and controls (30 subjects in each group) after cold challenge of the right hand for one minute in water at 20°C. Although some mention was made of absolute finger temperatures in this study, no detail was given on calibration of the thermal imager to a reference source.

Rudimentary computational techniques were by now available to extract temperature data from thermograms, and Ring [17] described a “thermographic index (TI),” for ischaemia of the hand, calculated by summing isothermal areas relative to 24°C, and expressed as a proportion of the total skin surface area measured:

\[
TI = \frac{\sum (\Delta t \cdot a)}{A}
\]

\[ t = 24^\circ C, \ a = \text{area of isotherm (cm}^2\text{), } A = \text{total area measured (cm}^2\text{)}.
\]

The TI increases therefore as skin temperature rises, and would be negative for a mean total area temperature below 24°C, and positive for temperatures above 24°C. The TI was found to be an effective way of measuring small changes in response to cold challenge over time. The cold immersion used of 20°C for one minute was also more tolerable for patients, and produced quicker recovery times, compared to challenge at lower temperatures. Again no detail was provided on the standardisation of the thermal imager.

The late 1970s began a period of interest in vasoactive drugs and their potential as possible treatments for RP. Ring and Bacon [18] studied the effects of inositol nicotinate in 6 RP patients, measuring hand temperature without cold challenge. They showed that two patients who remained on treatment for nine months had significantly warmer hands than the remaining patients, who had only been treated for two months. Ring et al. revisited inositol nicotinate in 1981 [19] using cold challenge thermography
in 20 RP patients treated over 36 weeks. IRT showed an improvement in the thermographic response after cold challenge at 36 weeks compared to baseline. The prostaglandins PGE$_1$ and PGI$_2$ showed particular promise clinically. Martin et al. [20] performed a single-blind crossover study comparing 72 hours of infusion of saline and PGE$_1$ in 12 SSc. patients. Hand thermograms were recorded daily during the course of the infusions, and two weeks after the infusions. Cold challenge was also performed immediately after the infusions. TIs for hand dorsum and fingers rose significantly on PGE$_1$ compared to placebo, with the peak difference after 2 days of infusion, but the authors were unable to demonstrate improvements to the cold challenge response on PGE$_1$.

The same research group also monitored responses to PGI$_2$ infusion in 24 SSc. patients across two centres: at one with an AGA 680 thermal imager and at another with a KT41 infrared thermometer [21]. They found IRT was able to demonstrate increasing finger and hand temperature during infusion, but this effect was shown less markedly using the infrared thermometer. Whilst IRT was standardised by reference to a blackbody temperature source, no details of quality assurance for the infrared thermometer measurements were reported. Consistent with the findings from PGE$_1$, only 2/24 patients improved their response to hand cold challenge after PGI$_2$ infusion. Kyle et al. [22], whilst reporting some therapeutic benefits of PGI$_2$ infusion on a double-blind, placebo-controlled study, also failed to demonstrate any improvement to the response to hand cold challenge over placebo. Shawket et al. [23], in a double-blind crossover study comparing infusions of PGI$_2$ and calcitonin gene-related peptide (CGRP), concluded that CGRP improved the response to hand cold challenge three days after infusion, whereas PGI$_2$ did not.

Hawkins et al. [24] achieved one of the largest RP drug studies in 1986 when they compared nifedipine and placebo in a double-blind crossover trial with 57 patients.
Temperature measurements were limited to an infrared thermometer technique, however, and no significant changes in the thermal gradient between fingers and dorsum of the hand could be shown. Gush et al. [25] also failed to elicit differences in finger temperature with IRT on comparing the sub-lingual form of nifedipine to placebo in a trial with nine RP patients.

These early drug studies took differing approaches to RP assessment with IRT, but there were similar trends in many of the results: clinical improvement did not always correlate with increased finger temperature; improvements in baseline finger temperature on treatment were not always mirrored by improvement in the response to cold challenge and attempts to standardise thermal imager performance were rarely reported.

The 1990s brought dramatic advances in IRT, with the introduction of higher performance imagers, and powerful computer software packages for analysis. Darton and Black [26] used a low-cost pyroelectric vidicon (PEV) imager to show good reproducibility of rewarming after hand cold challenge in normal subjects. Subsequently Darton and Black [27] compared cold challenge in 58 RP patients and 14 healthy controls. They showed differences in temperature range across the hand between RP and controls both before and after cold challenge, but they were unable to differentiate between primary and secondary RP using such measurements.

Over a thirty year period, the use of IRT for RP assessment had advanced from cold challenge of the hands in iced water and visual inspection of the resulting thermograms, to a mild, tolerable cold challenge evaluated quantitatively from software analysis of digital images. This enabled objective evaluation of the patient response to treatment during clinical trials, but the best way to employ IRT in studies of vasodilators was still unclear.
Prior to the publication of Appendix 1 in 1997, nothing significant had been published on RP in the toes. In addition to the early studies on prostaglandins etc., a number of other compounds were reported as showing promise for the treatment of RP. It remained to be seen if low-cost uncooled IRT devices were effective for the measurement of cold challenge responses in controlled trials of these new drugs. Up to this point, thermographic studies of RP had been performed almost exclusively by expensive cooled scanning devices. Cold challenge assessment in a population setting (e.g. for the evaluation of prevalence or heritability of RP) was also entirely unexplored. The projects described in section 2.3 established toe temperature in healthy and RP females, employed low-cost PEV IRT (StarSight imager, Insight Vision Systems, 8 – 14 µm) as an objective assessment of vasospasm in randomised trials of two oral vasodilators, and validated an infrared thermometer technique for cold challenge assessment against IRT, prior to use in a large twin population study into the heritability of RP.

2.3 Projects.

2.3.1 IRT with cold challenge for the detection of RP of the feet in females.

Appendix 1 describes an investigation of the response to cold challenge of the feet of healthy females, and females with RP in the feet.

Method. Six primary RP patients and 8 healthy controls participated. Each subject removed footwear and sat for 15 minutes prior to measurements. A baseline thermal image of the dorsal surface of both feet was captured after 15 minutes of acclimatisation at an ambient temperature of 23 ± 1°C. The feet were then covered with plastic bags and immersed in water at 15°C for one minute. The feet were then removed from the water, and thermograms were recorded at one-minute intervals for ten minutes. Toe temperatures were extracted from the thermograms recorded at baseline and at 0, 5 and
10 minutes after cold challenge. The mean temperature of all ten toes (Θ), and an index for medial-lateral foot temperature difference (ΔT) was calculated for each image.

**Results.** At baseline, healthy controls had a higher Θ than RP patients (29.2 ± 1.5°C v 24.8 ± 1.5°C [mean ± SD], p<0.01; t-test). In patients, Θ after cold challenge correlated strongly with Θ at baseline (r=0.96 at 0 minutes, r=0.96 at 5 minutes, r=0.92 at 10 minutes), but the correlation was less strong for control subjects. ΔT was greater in control subjects than in patients at baseline (0.7 [2 – -0.4] ºC v 0.5 [0.4 – -3.6] ºC, median and range, p<0.02; Mann-Whitney). The difference in ΔT between the groups was increased by cold challenge.

**Discussion.** The toes of female RP patients were colder than those of healthy female control subjects, both before and after cold challenge. The rewarming rate of normal toes was slower than that reported for normal hands undergoing a comparable cold challenge.

In the patient group, the strong correlation between Θ at baseline and after cold challenge suggested that baseline mean toe temperature alone may be an adequate indicator of RP.

### 2.3.2 IRT with hand cold challenge to assess vasodilator therapy for RP in clinical trials.

**Appendix 2** reports on a fifteen-week randomised, parallel group, controlled trial of losartan therapy for RP and scleroderma.

**Method.** Patients with primary RP (n=25) and RP secondary to SSc. (n=27) were randomised to receive 12 weeks of therapy with either 50 mg/day of losartan (an angiotensin II type 1 receptor antagonist) or 40 mg/day of nifedipine. All subjects kept diaries recording the frequency and severity of RP symptoms.
Before and after treatment, response to cold challenge (both hands gloved and immersed in water at 15°C for one minute, room temperature 23 ± 1°C) was assessed using IRT. Mean fingertip temperature (excluding the thumbs [28]) was determined from thermograms of the dorsal hand surface recorded before cold challenge and immediately after, 5 minutes after, and 10 minutes after cold challenge.

Results. From symptom diaries there was a greater reduction in the severity of RP episodes with losartan than with nifedipine after 12 weeks of therapy (p<0.05, t-test). RP episode frequency diminished after therapy only in the losartan group (p<0.01 versus baseline, paired t-test).

There was no improvement in recovery 10 minutes after cold challenge in either treatment group as assessed by both thermography and LDF.

Discussion. The study suggested a clinical benefit from losartan therapy for the treatment of RP, on the basis of symptom diary data.

The lack of any improvement in recovery from cold challenge was interesting given the change in diary scores after losartan therapy. The results may indicate that increases in microvascular blood flow that are too small to detect reliably in the laboratory can nonetheless lead to significant clinical improvements. Microvascular measurements should be included in future clinical trials of vasodilators so that the validity of such techniques can be further assessed.

Appendix 3 describes a randomised, 16-week cross-over study of the selective serotonin reuptake inhibitor fluoxetine for the treatment of RP.

Method. Patients with primary RP (n=26) and secondary RP (n=27) were randomised to receive six weeks of therapy with either fluoxetine (20mg/day) or nifedipine (40mg/day). After a two-week washout period, each group was then crossed over to the other treatment arm. Patients recorded the severity and frequency of RP attacks by
means of a visual analogue scale. Thermography studies were performed at the start of
the trial and at the completion of each treatment arm, using the cold challenge protocol
described in appendix 2. Thermograms were recorded prior to, and immediately after
cold challenge, and ten minutes after cold stress.

Results. RP attack frequency and severity were significantly reduced after treatment in
the fluoxetine group (p=0.0002 for severity and p=0.003 for frequency, paired t-test)
with the greatest response seen in females and patients with primary RP on sub-group
analysis. Rewarming from cold challenge after therapy improved in both treatment
groups, but the improvement was not statistically significant when all patients were
included in the analysis. Sub-group analysis, however, showed a statistically significant
greater rewarming in females after treatment with fluoxetine (p=0.05), but not in males
(p=0.94). Patients with primary RP showed an improved rewarming after fluoxetine
treatment (p=0.03), whereas secondary RP patients did not (p=0.97).

Discussion. Our results suggested fluoxetine to be an effective treatment for RP,
particularly in females and primary RP patients. The improvement in symptom diary
scores for these two sub-groups was mirrored by significant changes in rewarming after
cold challenge. This study indicated thermography is an important objective adjunct to
subjective outcome measures, provided the response to treatment is large enough.

2.3.3 An infrared thermometer technique, validated against IRT, for the
assessment of hand cold challenge in a twin study of the heritability of RP.

Appendix 4 compares IRT (using the StarSight PEV imager [29]) to infrared
thermometry for the detection of RP.

Method. Eighteen female patients with primary RP and nineteen healthy female controls
underwent the cold challenge outlined in Appendix 2. Thermograms of the dorsum of
both hands were recorded at four time points: prior to cold challenge, immediately after cold challenge, and at 5 and 10 minutes after cold challenge. After each thermogram, fingertip temperature was recorded at 8 fingertips using a Land Cyclops infrared thermometer. Mean temperature for all 8 fingertips was calculated for both measurement techniques at all four time points ($T_{\text{pre}}$, $T_{\text{post}}$, $T_5$ and $T_{10}$). Three measures were derived from each subject for both instruments: baseline temperature ($T_{\text{pre}}$), temperature drop after immersion ($T_{\text{pre}} - T_{\text{post}}$), and temperature rise over ten minutes after immersion ($T_{10} - T_{\text{post}}$).

**Results.** There was good agreement between the two instruments: for all three measures the intraclass correlation between instruments was >80%. A logistic regression model containing all 3 measures yielded a sensitivity for detecting RP of 83% for thermography, and 89% for the infrared thermometer. The specificity of both instruments was 84%.

**Discussion.** Although infrared thermometry had some limitations (labour intensive, lack of a permanent image record), it provided an inexpensive alternative to IRT for RP detection.

**Appendix 5** applies the infrared thermometer protocol validated in appendix 4 to identify RP in a population setting.

**Method.** All subjects were females (175 classified RP and 404 classified control by questionnaire). Each participant underwent cold challenge along with infrared thermometer assessment of baseline mean finger temperature (B), fall in finger temperature after cold challenge (F) and finger rewarming over the ten minutes after cold challenge (R). Logistic regression was used to fit models to the data, whereby the outcome variable was classification of RP by questionnaire and the predictor variables
were the temperature measurements B, F and R. Fit was compared between the 3-variable, 2-variable and 1-variable models.

**Results.** Both B and R were significantly lower for RP subjects than for controls (for B: 28.3°C vs 30.0°C, p<0.01, t-test; for R: 4.6°C vs 5.3°C, p<0.05, t-test). No model incorporating baseline temperature (B, BF, and BR) showed any significant difference in fit to the full BFR model, whereas all models not including B (FR, R, and F) showed significantly worse fits than the full model. Subjects with B ≤ 24°C were nearly 3 times more likely to be RP positive than RP negative.

**Discussion.** Baseline finger temperature helped to identify RP subjects in a population setting, but since the B, F and R variables were strongly correlated, cold challenge provided little additional information. All 3 thermographic variables may nonetheless be helpful criteria for RP assessment in other settings, e.g. assessment of severe disease or monitoring of response to treatment.

**Appendix 6** applies the cold challenge test and infrared thermometer measurements to a sample of 288 female twin pairs drawn from a larger sample of 3043 individuals who responded to a questionnaire-based study.

**Method.** Of the 288 twin pairs, 129 were monozygotic (MZ) pairs, and 159 were dizygotic (DZ). On the basis of the questionnaire responses subjects were classified as normal, cold-sensitive, RP, or severe RP. The subset was weighted to balance the proportions of MZs and DZs across 3 groups of twin pairs: 1) those that were concordant for the presence of RP on questionnaire 2) those that were discordant for RP and 3) those concordant for the absence of RP. All attendees underwent the infrared thermometer cold challenge protocol outlined in appendix 4. Heritability for the three temperature criteria baseline, fall and rewarm were estimated using maximum
likelihood structural equation modelling. Heritability for questionnaire responses suggesting cold sensitivity, RP and severe RP was also calculated.

Results. Cold sensitivity, RP and severe RP all showed greater concordance among MZ twins than DZ twins. Modelling suggested a heritability of 53% for cold sensitivity, 55% for RP, and 53% for severe RP. Analysis also indicated a heritability of 65% for baseline skin temperature, 35% for fall after cold challenge, and 24% for rewarming over ten minutes.

Discussion. The study demonstrated a higher concordance for RP in MZ twins than in DZ twins, confirming the genetic basis of RP. The physical response to cold was measured for the first time in a genetic study of RP, and it was shown that the skin temperature variables measured were also heritable. Baseline skin temperature and rewarming are correlated with the report of RP symptoms (appendix 5), so these variables should be included in studies of the underlying mechanisms of the condition.

3.1 Introduction.

3.1.1 Background.

Localised scleroderma (LS) is a connective tissue disorder characterised by the fibrosis of skin, often distributed unilaterally across the body [30]. LS begins with an inflammatory “active” phase during which the lesion of indurated skin spreads (fig. 3.1). Lesions then progress to a stage wherein inflammation diminishes, the dermis thins, and no further spreading occurs. This stage can also be associated with atrophy of structures underlying the dermis: fat and muscle (fig. 3.2).

Fig. 3.1: An early-stage active LS lesion with inflamed skin.

Fig. 3.2: Late-stage LS, with growth failure and contracture of the affected limb.
Although rare, LS is the most common form of childhood scleroderma [31]. Left untreated, severe LS leads to deformity, flexion contractures of joints, and limb length discrepancies.

The treatment of choice for severe LS is pulsed corticosteroid followed by low-dose methotrexate [32,33]. Successful treatment is dependent upon reliable measures of disease activity to enable informed decisions to be made about when to start, modify, and stop therapy. At present, no established reliable laboratory indicators of LS disease activity exist.

3.1.2 Aim of the project.

The aim of this project was to establish a protocol for thermography (along with other modalities) for the assessment of LS activity. The objectives were:

- Establishing the utility of IRT for the detection of active LS based on the inspection of thermograms (Appendices 7, 8, and 9).

- Developing a protocol for skin assessment incorporating photography, IRT and laser Doppler flowmetry (LDF) and establishing the normal range of readings for temperature and LD blood flow across several body sites in adults and children (Appendix 10).

- Comparing the utility of IRT and LDF for the assessment of disease activity in children with LS (Appendix 11).

3.2 Literature review.

Published work pertaining to assessment of disease activity in LS is reviewed herein. Allen et al. [3] were first to describe the use of IRT for LS assessment. They reported a case of linear scleroderma in a 5-year old boy. The patient was treated with intravenous pulsed methylprednisolone. After an initial clinical improvement, D-penicillamine was commenced and continued long-term. Over an 18 month period IRT showed a
progressive reduction in the temperature difference between the two sides of the body, correlating well with response to treatment. The authors did not comment on any measures taken to ensure standardisation of the thermal imager performance over the 18 month measurement period.

Subsequently, Birdi et al. [4] reported the use of IRT in 11 LS children. Lesions were judged “thermography positive” if the skin was at least 0.5°C warmer than contralateral control skin on inspection of the thermogram (fig. 3.3). All three lesions that had recently spread were “thermography positive,” whereas three lesions that had been quiescent long-term were “thermography negative.” The authors did not investigate the intra- and inter-operator reproducibility of the subjective assessment of “thermography positive” lesions.

Fig. 3.3: Example of a “thermography positive” lesion at the neck (left), which is “thermography negative” after five months of treatment (right).

In 1984 Serup and Kristensen [34] studied blood flow in 30 LS lesions using LDF, first described in a form practicable for cutaneous measurements by Stern at al [35]. LDF allows non-invasive real-time measurements of skin blood flow and has been described in detail by Oberg [36].
Serup and Kristensen [34] showed that blood flow was increased in the sclerotic areas of all LS lesions. There was also a tendency for lesions of shorter duration to be higher in flow.

Klyscz et al. [37] used LDF to monitor the reduction in blood flow in an LS lesion in response to antibiotic treatment in a 38-year old male.

No subsequent studies have been published using either IRT or LDF to measure treatment outcomes in LS. Furthermore, the two techniques had never been used in combination in LS. The projects described in section 3.3 developed IRT and LDF into techniques ready for use in clinical practice, and compared their utility for LS activity detection.

### 3.3 Projects.

#### 3.3.1 Simple inspection of thermograms for the detection of clinically active LS.

**Appendix 7** describes a retrospective study performed using thermograms recorded from Royal Free Hospital patients between 1993 and 2000.

**Method.** Seventeen patients were included in the review. The mean age at first thermographic assessment was 11.5 years (range 2-29 years). All patients had an onset of LS in childhood (mean age 6.9 years, range 1-15 years). Twenty-five lesions were reviewed (seven patients having more than one LS plaque). Nine patients attended for more than one thermographic assessment, and so 48 thermograms were included in total.

Prior to thermography all patients had rested in the laboratory at a room temperature of 23 ± 1°C for 15 minutes, partially disrobed as appropriate. Thermography was performed by the StarSight PEV imager (8 - 14μm).
Copies of each thermogram were issued along with descriptions of the lesion position to two rheumatologists (KM and GM). Both independently assessed the thermographic activity of each lesion, classifying any lesions greater than 0.5°C warmer than the contralateral site or adjacent tissue as “thermography +ve,” as per the criteria of Birdi et al. [4]. Each lesion was then classified clinically “inactive” or “active” from the descriptions written at the time of thermographic assessment in the patients’ medical records.

Results. GM achieved a sensitivity for identifying clinically active LS lesions by inspection of thermograms of 92%, and a specificity of 80%, and KM achieved similar results. There was discordance between the two observers in only one of 48 lesions. All five clinically “inactive” lesions judged “thermography +ve” were associated in the notes with subcutaneous lipoatrophy.

Discussion. The inspection of infrared thermograms of LS lesions had a sensitivity for detecting clinically “active” lesions of approximately 90%, and a specificity of around 80%. Results were reproducible between observers. IRT was less useful in later-stage LS associated with lipoatrophy: in these lesions there may be an increased risk of “false-positive” thermograms due to heat transfer from deeper tissues through the thinned fat layer.

Appendix 8 extends the retrospective analysis of LS thermograms in appendix 7 to include children with juvenile LS managed at not only the Rheumatology Department at the Royal Free Hospital, but also the Rheumatology and Dermatology Departments at Great Ormond Street Hospital.

Method. Forty children (14 male, 26 female) with onset of LS before the age of 16 years were included in this updated review. Sixty-eight clinically separable lesions were identified. Fifty per cent of the lesions had been assessed by IRT more than once, resulting in 130 thermograms reviewed for the study.
Results. GM achieved a sensitivity for identifying clinically active LS lesions of 92%, and a specificity of 68%. KM again achieved similar results. In lesions imaged within 2 years of onset, sensitivity was 81% and specificity 87.5%

Discussion. The increased specificity in a sub-group of children assessed within two years of disease onset reinforced the hypothesis that lipoatrophy and deformity in later-stage LS lesions limited the utility of IRT by increasing the risk of “false-positive” thermograms.

Appendix 9 is a retrospective study of outcomes in 34 LS patients, all of whom were treated at Great Ormond Street Hospital with pulsed intravenous methylprednisolone followed by oral prednisolone on a reducing regimen and maintenance treatment with methotrexate.

Method. Thermography was performed in 33 of these patients prior to treatment, and all 26 patients that were “thermography +ve” were followed up with IRT for a minimum period of 2 years. Judgment regarding whether thermography was positive or negative from each thermogram was made by two experienced observers (KH and LW).

Results. There was agreement between the two observers in 32/33 patients prior to treatment, and in 22/26 on follow-up. Thermography had a sensitivity for detecting clinically active LS pre-treatment of 79%. (26/33). Thirteen of the 26 “thermography +ve” patients at baseline remained “thermography +ve” at follow-up, although 10 of these patients were considered to have clinically inactive LS by this stage.

Discussion. IRT proved to be a sensitive tool for detecting LS activity at baseline, but this study reinforced the limitations of IRT in older LS lesions. The 15% lack of agreement between observers in follow-up thermograms also suggested that older lesions are more difficult to interpret using IRT.
3.3.2 A protocol for assessment of skin temperature and microvascular blood flow in LS.

Appendix 10 describes a protocol suitable for the assessment of unilateral LS in children, and its validation by measuring the skin of healthy adult control subjects, and unaffected skin in a group of children with LS.

**Method.** In seven adult control subjects, skin temperature was assessed by IRT (FLIR SC500 uncooled FPA microbolometer imager, 7.5 - 13µm) and blood flow was assessed by LDF at pairs of contralateral skin sites in six body regions. Digital photographs of each body region were also captured, and these images were superimposed on the equivalent infrared thermograms to produce an “overlay” composite image.

The same protocol was used to assess skin temperature and blood flow on the unaffected side of the body in 34 paediatric patients diagnosed with unilateral LS.

The variability of LD flux readings over an area of forearm skin of 10cm² was investigated, along with the repeatability of 10 readings of 10 seconds duration recorded from precisely the same skin point at one-minute intervals. The intra- and inter-operator reproducibility of the IRT and LD analyses were also determined.

**Results.** In adults, skin sites at the head were the warmest, and the sites at the periphery were cooler. LD flux was similarly distributed. Mean differences in temperature between the two sides of the body were less than 0.5°C in all body regions. Mean differences in contralateral LD flux were less than 40% at all body sites. Mean skin temperature in children was significantly lower than that in adults at the forehead, posterior trunk and arm. LD flux in children differed significantly from that in adults only at the leg.

The spatial variability of LD flux across forearm skin was 13%. Repeatability of LD flux, based on 10 readings from a single point was 9%. The intra-operator reproducibility of the LD trace analysis was 1%, and the inter-operator reproducibility
was 5%. The intra-operator reproducibility of IRT analysis was 0.2°C, and the inter-operator reproducibility was 0.3°C.

**Discussion.** The study established control data for the expected differences in temperature and LD flux between the sides of the body. The mean differences in temperature at all body sites were consistent with the assumption by previous authors that an elevation in skin temperature of 0.5°C in LS lesions is abnormal [4]. The variability in LD and IRT readings due to experimental factors was acceptably small in comparison to the physiological differences recorded.

### 3.3.3 IRT and LDF for the assessment of disease activity in LS

**Appendix 11** reports the use of LDF for the detection of clinically active LS lesions in children, and compared the technique’s performance to that of IRT, using the protocol outlined in appendix 10.

**Method.** 111 LS lesions were assessed clinically, and by LDF and IRT on two separate occasions in 41 children. The specificity and sensitivity of LDF and IRT to detect clinically active LS lesions were compared.

**Results.** 34% of the lesions were clinically identified as active. The median relative increase in LD blood flow in these lesions (compared with blood flow in contralateral unaffected skin) was 89% (range -69% to +449%), whereas the median flow increase in clinically inactive lesions was 11% (range -46% to +302%), p<0.001.

Using IRT, the median temperature difference between clinically active lesions and contralateral unaffected skin was 0.5°C (range -0.1°C to +4.1°C), whereas the median temperature difference for clinically inactive plaques was 0.3°C (range -1.9°C to +2.7°C), p=0.024.
ROC analysis identified a cut-off rise in LD blood flow of 39% as the best value for discriminating clinically active from inactive lesions, with a sensitivity of 80% and a specificity of 77%. No useful cut-off rise in temperature could be identified using IRT.

Discussion. LDF proved to be a sensitive and specific technique for identifying clinically active LS lesions, and was more effective for this purpose than the IRT protocol described in appendix 10.
4: Developing good practice of IRT through appropriate imager procurement and quality assurance.

4.1 Introduction.

4.1.1 Background.

Clark and de Calcina-Goff [38] described the minimum specifications for a medical thermal imager in their paper of 1997. The key properties discussed were

- **Wavelength ($\lambda$)** at which the device operates: **2-5 µm or 8-13 µm**.
- **Spatial resolution** (often quoted for array systems as the “instantaneous field of view” [IFOV], the angle of view recorded by a single pixel for a given field of view [FOV]): **2.5 mrad at 20° FOV or better**.
- **Accuracy** of temperature measurement: **± 1°C or ± 1%**.
- **Thermal sensitivity** (often quoted as “noise equivalent temperature difference” [NETD], since noise within the image normally limits the ability to discriminate small temperature differences): **0.1°C or less**.
- **Dynamic range** of the camera (describes the number of discrete levels into which the temperature measurement range of the device is divided): **≥12 bit** (4096 levels).

Table 4.1 outlines the specifications (where quoted) for the three thermal imagers that have been used at the Royal Free Hospital: the Insight Vision Systems “StarSight” camera (from the early 1990s), the FLIR SC500 (commissioned in 2001), and the FLIR A320G (purchased in 2008). It is instructive to compare these specifications to the minimum requirements quoted by Clark and de Calcina-Goff in 1997.
### Table 4.1: Specifications for three thermal imagers used at the Royal Free Hospital.

<table>
<thead>
<tr>
<th></th>
<th>IVS STARSIGHT</th>
<th>FLIR SC500</th>
<th>FLIR A320G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detector</td>
<td>Pyroelectric vidicon (PEV)</td>
<td>FPA Uncooled microbolometer</td>
<td>FPA Uncooled microbolometer</td>
</tr>
<tr>
<td></td>
<td>(DTGS*)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>λ</td>
<td>8 – 14 µm</td>
<td>7.5 – 13 µm</td>
<td>7.5 – 13 µm</td>
</tr>
<tr>
<td>FOV</td>
<td>Unspecified</td>
<td>24° x 18°</td>
<td>25° x 19°</td>
</tr>
<tr>
<td>Resolution</td>
<td>128 x 128 pixels</td>
<td>320 x 240 pixels</td>
<td>320 x 240 pixels</td>
</tr>
<tr>
<td>Spatial resolution</td>
<td>3.3 mrad †</td>
<td>1.3 mrad</td>
<td>1.4 mrad</td>
</tr>
<tr>
<td>(IFOV)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>Unspecified</td>
<td>± 2°C or ± 2%</td>
<td>± 2°C or ± 2%</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>Unspecified</td>
<td>0.07 °C</td>
<td>0.05 °C</td>
</tr>
<tr>
<td>(NETD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic range</td>
<td>6 bit</td>
<td>16 bit</td>
<td>16 bit</td>
</tr>
</tbody>
</table>

*Deuterated triglycerine sulphate

†Unspecified: calculated from pixel resolution and assuming the FOV of the FLIR SC500

The StarSight PEV imager failed to meet the criteria of Clark and de Calcina-Goff on at least two counts: the spatial resolution was too low and the camera was only a 6 bit device. Thermal sensitivity, although unspecified, would at best have been limited by the dynamic range: a 6 bit imager measuring across a 20°C range implies a minimum detectable temperature difference across the image of around 0.3°C at best. This is 3 times the value required by Clark and de Calcina-Goff. Whilst the author has shown in appendices 1 – 4 and 7 – 9 that instructive medical measurements can be made with
PEV technology, camera performance is not high enough for many medical applications.

In contrast the FPA devices offered good spatial resolution, thermal sensitivity and dynamic range which were more than adequate for medical thermography. With these devices only one criterion remained unmet: that of the temperature accuracy of ±1°C or better. In fact very few modern thermal imagers (including the high-end cooled photon detector devices with excellent sensitivity) offer accuracy much better than ±2°C, and ensuring temperature measurements are indeed accurate to better than ±1°C is an issue that will be addressed in section 4.3.1.

Ring et al. [39] described the components that must come together to guarantee quality assurance in medical IRT. These elements are shown schematically in fig. 4.1.

![Fig. 4.1: The “standard protocol”: components for quality assured medical IRT.](image-url)
Not all of the factors in figure 4.1 have been given equal attention in the literature. Whilst much is reported on preparation of the patient i.e. the factors that influence the physiology under investigation (room temperature, acclimatisation times etc.), little is published on the more “technical” elements that go to make up quality assurance in IRT. The least discussed of all is standardisation and calibration of the imager.

There are three reasons for the lack of data on camera standardisation. The first and most important of these is the pace at which infrared technology has developed in the last twenty years. The introduction of uncooled FPA imagers has meant that many of the reports describing assessment of cooled scanning systems are now obsolete.

Secondly imager manufacturers have not stressed the importance of understanding imager performance to medical customers, perhaps because their products have not always met requirements for medical use without modification to standard operating procedures. An example of this is the necessity for FPA imagers to be used in medicine only after a period of “warm-up” to stabilise the detector. There is a need for education of medical thermographers in the field of imager quality assurance, and medical professionals may not be hearing the right message from manufacturers.

Thirdly, medical IRT is often performed under pressures to reduce operating costs after the initial expenditure of purchasing the imager. Further investment in equipment to ensure IRT quality assurance must demonstrate “value for money.” This means the equipment must be reasonably-priced and offer a calibration that genuinely validates IRT performed across the narrow medical temperature range. Such a device has not been available and so thermographers, hearing the over-confident assurances of manufacturers, have tended to develop thermography services with only limited QA in place. This attitude at best limits the credibility of IRT and at worst exposes patients to the risk of unreliable temperature measurements and misdiagnosis.
4.1.2 Aim of the project.

The aim of this project was to develop standardised practice for the QA of medical IRT as performed by a modern uncooled FPA thermal imager, and disseminate advice on IRT QA and correct imager procurement to the wider medical IRT community. The project objectives were identified as:

- Contributing to the specification of portable in field of view blackbody temperature sources for the calibration of medical thermal imagers, and validating the manufactured devices in a hospital “field trial” (Appendix 12).
- Publishing guidelines for procuring and commissioning a medical thermal imager, with particular emphasis on European health and safety legislation (Appendix 13).

4.2 Literature review.

Published work pertaining to the standardisation and QA of IRT is reviewed herein, considering the literature chronologically.

Prior to the 1980s, little was published on standardisation of medical IRT, although Clark [40] had considered the dependence of viewing angle on the effective emissivity of skin. In 1980 Vermeij et al. [41] suggested the modulation transfer function (MTF) as a means of assessing the spatial resolution of a scanning thermograph. Friedrich [42] described the criteria an effective medical IRT device should meet.

Pochaczevsky et al. [43] published “Technical Guidelines” for medical IRT in 1986, but considered only the preparation of the patient and image capture protocols.

As discussed above, Clark and de Calcina-Goff [38] published draft standard proposals for IRT in 1997. This extensive document defined the terms used in IRT, and discussed briefly all of the elements outlined in fig. 4.1. Later in the year Plassmann and Ring [44] outlined the computing hardware and software requirements for standardised image
capture, and Diakides [45] reviewed the emerging uncooled FPA technology that has since revolutionised IRT.

Ring and Dicks [46] compared the spatial resolution of FPA imagers and older scanning systems by use of a heated array with linear elements at different spacing. Murawski et al. [47] considered further the software requirements for image capture and analysis.

Ammer and Ring investigated the influence of image capture protocol on the repeatability of temperature readings [48] and measured the effect of varying the field of view [49]. Subsequently Ammer [50] analysed the repeatability of temperature readings from the forehead using a standard view.

It was still the case at the start of the 21st century that little had been published addressing the major limitation of all thermal imagers – that of limited temperature accuracy and the difficulties associated with ensuring traceability of measurements at temperatures in the medical range.

In 2006 Simpson et al. [51] finally addressed the importance of calibration in medical temperature measurement, and traceability to the International Temperature Scale of 1990 (ITS-90) [52].

Plassmann et al. [1] followed with a detailed QA protocol for medical IRT with particular relevance to the limitations of modern uncooled FPA detectors.

Despite this realisation of the importance of calibration, traceability of measurement, and QA of uncooled FPA imagers, no guidelines had been published up to this point on the procurement of a thermal imager within the constraints of European medical device legislation. Equally no portable, practical calibration device had been proposed to ensure traceability of medical IRT to ITS-90.

The projects described in section 4.3 include a comprehensive review of the issues surrounding procurement and QA of a thermal imager within a European healthcare
institution. A validation system for medical IRT, traceable to ITS-90, was specified, designed and constructed by collaborators at the National Physical Laboratory, and evaluated within a clinical setting at the Royal Free Hospital.

4.3 Projects.

4.3.1 Specifying and validating a portable in field of view (IFOV) temperature source for the calibration of medical IRT.

Appendix 12 describes the National Physical Laboratory project to design a portable calibration system for medical IRT QA. The authors provided

- input to the specification of the device to ensure it met the needs of medical thermographers;
- facilities for “field-testing” the prototype device in a hospital environment.

Method. The design requirements for the IFOV sources were agreed upon at a meeting of the research team. NPL decided upon three materials with fixed-points within the required temperature range: gallium-zinc eutectic, gallium and ethylene carbonate. The blackbody cavity design was a PTFE cylindro-cone of length 150mm, and aperture diameter 26mm. Calculated cavity emissivity over the 8-13µm waveband was 0.9983. The gallium-based fixed points were cooled by peltier coolers/heaters to ensure they were completely frozen, before being heated to just above their melting points. The source temperature while passing through the melting plateau was monitored using a Land Cyclops C300 infrared thermometer as the transfer radiation thermometer (referenced back to NPL’s ammonia heatpipe as the comparison blackbody). The ethylene-carbonate fixed point was initiated by immersing it into a Dewar of water at approximately 75°C (to ensure it was fully melted), followed by immersion in tap water at 20°C and then a physical shock (to take the source through its freezing plateau).
Stability, repeatability, radiance temperature and overall uncertainty were determined for the cells from several melt/freeze cycles.

The cells were then transported for “field trials” at clinical centres. Each of the three centres used the cells for a period of one month, and reported on the performance and ease-of-use of the devices. At the Royal Free Hospital the author supervised a trainee Clinical Scientist in the evaluation of the cells, and the experience gained in calibration and QA was included as part of the trainee’s “Physiological Measurement” portfolio.

**Results.** The average temperatures of the sources were: gallium–zinc 25.3°C, gallium 29.8°C and ethylene carbonate 35.9°C. The uncertainty attributed to each cell was ±0.44°C (k = 2). All the clinical centres found the cells to be quite easy to use and robust. It was commented that the operating time of 1-2 hours for the ethylene carbonate cell was insufficient for routine clinical use.

Fig. 4.2, 4.3 and 4.4 are unpublished data illustrating how the IFOV sources were used during the field trial at the Royal Free Hospital to validate a hand cold challenge procedure. Fig. 4.2 plots mean finger temperature in blue before cold challenge, and thereafter recorded at one-minute intervals. Also shown are the temperatures recorded for each source at every time point, plotted in red. The error bars on all these readings reflect the specified accuracy of the A320G Thermovision imager of ±2°C. The known temperature of each source is also plotted onto the graph as a dotted green line, along with error bars that confirm the uncertainty in these values as ±0.44°C.
For each time point, the recorded camera temperature of each source (x-axis) was plotted against the known source temperature (y-axis) and fitted to a straight line in Microsoft Excel. This produced a calibration curve and formula for modifying the observed mean finger temperature values at every time point. Fig. 4.3 shows the resulting calibration curve for the measurement made prior to cold challenge. At all time points, the fit to a straight line was excellent ($r^2=1$), and the maximum standard error on the y-value estimate was 0.20°C.
Fig. 4.3: Calibration curve for the A320G thermal imager using the IFOV sources at the time point prior to cold challenge.

Fig. 4.4 compares the uncorrected mean finger temperature curve in red with the curve corrected using the IFOV sources in blue. The error bars for the corrected curve combine the uncertainty in the source temperatures with the maximum error on the y-estimate of the calibration curve to give an uncertainty for each value of:

$$\sqrt{(0.44^2 + 0.20^2)} \approx 0.5^\circ C.$$

Calibration with the sources

- reduces the mean finger temperature values by up to 0.7°C at low temperatures but makes little change at higher temperatures, suggesting the imager exhibits a small gain error across the range.
- reduces the overall uncertainty in temperature readings by a factor of 4.

*Discussion.* The cells provided a practicable IFOV calibration system for medical IRT, reducing uncertainty in the temperature measurements from typically ±2°C to ±0.5°C.
Effect of calibration on finger rewarming curve

Fig. 4.4: Mean finger rewarming before and after calibration with IFOV sources.

4.3.2 Guidelines for procuring and commissioning a medical thermal imager.

Appendix 13 is a review article offering guidelines for the purchase, commissioning and ongoing QA of a modern thermal imager for medical use within a European healthcare institution. This is the first article published describing the management of medical IRT equipment across its entire life-cycle. Among the topics addressed are:

- **Procurement.**
  - Identifying a clinical need.
  - Specifying the imager.
    - Thermal performance.
      - Accuracy and thermal sensitivity, as discussed in section 4.1.1.
      - Cooled versus uncooled detectors. Whilst uncooled microbolometers are cheaper, thermal sensitivity is higher for cooled photon detector systems.
    - Optics.
      - Choice of lens for close-up or widefield imaging.
• Array sizes and spatial resolution of the imager.
  o Software.
    ▪ The importance of software written under a quality system.
    ▪ Ensuring the image analysis software meets the needs of medical thermography.

• Whole lifetime costs.
  o Budgeting for the purchase, operation, and maintenance of a medical thermal imager.
  o Likely imager life: the concept of mean time between failures (MTBF).

• The concept of risk, and risk assessment.
  o Definition of a Medical Device and the European Medical Devices Directive. Most thermal imagers do not comply with this directive, so risk assessment must identify device deficiencies and address how to overcome them.

• Acceptance testing.
  o How to take delivery of the device and prepare it for clinical use.
    ▪ Asset logging, visual and functional checks, tests for electrical safety.

• Calibration and Quality Assurance.
  o Temperature calibration, with reference to the IFOV system described in Appendix 13.
  o Other QA, with reference to Plassmann et al. [1].
    ▪ Simple tests for offset drift after switching on, long term offset drift, offset variation over camera temperature range.

• Service and maintenance.
  o Service contracts.
Certification and accreditation of calibration during maintenance.

The review concludes with a reference section and an example risk assessment for a thermal imager in medical use.

The article is now cited in the “Guidelines” section of the European Association of Thermology web site at www.europeanthermology.com
5 : Discussion.

5.1 Assessment of contribution to knowledge of projects.

5.1.1. A thermographic service for Raynaud’s phenomenon.

Prior to the publication of Appendix 1, no data were available on temperature of the toes in RP in females. Appendix 1 was a pilot study and involved only a small number of patients. The data showed a strong correlation between baseline toe temperature and the temperature recorded at time-points after cold stress – a correlation that would also become evident in our later work on RP of the hands. Cold challenge of the feet remains a standard investigation for lower limb RP at the Royal Free Hospital. We and others also find the protocol useful for the detection of erythromelalgia [53], a rare lower limb disorder now recognised as having a vasospastic component.

Appendices 2 and 3 report the use of IRT for the assessment of response of the hands to cold in two substantive pilot drug trials (n>50). At many centres IRT or similar laboratory techniques are simply not available. Hence trials of this size often rely on the evaluation of symptomatic improvement (typically using some form of “Visual Analogue Scale” [54] whereby each subject self-assesses the frequency and severity of RP episodes). Adjunctive laboratory data on skin temperature changes after cold challenge afforded our RP studies an important element of objectivity.

Whilst both studies reported symptomatic improvements over nifedipine in RP after treatment, only the fluoxetine trial detected any significant improvements in the thermographic response to cold challenge. Both the losartan and fluoxetine studies were open-label, raising the possibility of placebo response to the novel treatment in the subjective patient self-assessments. Thermographic response to cold challenge should be unaffected by bias arising from the unblinded nature of the studies.
An important difference between the two studies was that the losartan investigation was a parallel-group trial, whereas the fluoxetine study used a crossover design [55-57]. Despite their complexity, crossover studies may be more desirable for detecting drug effects because they compare the response to both drugs in each volunteer. This eliminates the possibility of biasing a parallel-group study by randomising into one treatment arm a majority of subjects who are not responsive to any form of oral vasodilator.

The recommendations arising from the work in clinical trials of vasodilators are therefore:

- Crossover studies are to be preferred to parallel-group trials, and;
- If the studies need to be open-label, some form of objective assessment of treatment outcome like IRT is desirable.

**Appendices 4, 5 and 6** fulfilled the aim of validating an infrared thermometer technique against IRT, and applying it for the detection of RP in a population-based setting and ultimately for the investigation of the heritability of RP in a twin study. Once again the rationale was to bring objective measurement of skin temperature to a study in which the outcome measures would otherwise have been based on subjective recall of symptoms.

The Cyclops infrared thermometer technique of appendix 4 undoubtedly introduced more error into the measurements than thermography. The infrared thermometer recorded from only a very small skin surface area and required each fingertip to be measured individually at a consistent distance from the skin. Nonetheless Cyclops was more than sensitive enough to detect the large temperature differences between the two subject groups (>6°C as measured by Cyclops for mean baseline temperature). Cyclops correctly classified 86% of the subjects in a hospital setting.
As expected, the detection of RP in a population-based setting proved more challenging for Cyclops. Table 5.1 compares the Cyclops mean values and range for baseline finger temperature, drop in temperature after cold challenge, and rise in temperature after cold challenge for RP and non-RP groups in both the hospital and population-based studies.

<table>
<thead>
<tr>
<th></th>
<th><strong>RP</strong> Mean</th>
<th><strong>Range</strong></th>
<th><strong>NON - RP</strong> Mean</th>
<th><strong>Range</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hospital</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>setting</strong></td>
<td></td>
<td>n = 18</td>
<td></td>
<td>n = 19</td>
</tr>
<tr>
<td>Baseline ºC</td>
<td>23.8</td>
<td>20.9 – 27.2</td>
<td>29.9</td>
<td>23.1 – 33.9</td>
</tr>
<tr>
<td>Drop ºC</td>
<td>3.9</td>
<td>2.6 – 6.1</td>
<td>7.5</td>
<td>4.0 – 10.5</td>
</tr>
<tr>
<td>Rise ºC</td>
<td>1.6</td>
<td>0.3 – 3.6</td>
<td>5.7</td>
<td>-0.4 – 10.3</td>
</tr>
<tr>
<td><strong>Population</strong></td>
<td></td>
<td>n = 175</td>
<td></td>
<td>n = 404</td>
</tr>
<tr>
<td><strong>setting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline ºC</td>
<td>28.3</td>
<td>20.3 – 35.0</td>
<td>30.0</td>
<td>22.0 – 35.3</td>
</tr>
<tr>
<td>Drop ºC</td>
<td>6.6</td>
<td>2.3 – 12.8</td>
<td>6.9</td>
<td>0.3 – 11.4</td>
</tr>
<tr>
<td>Rise ºC</td>
<td>4.6</td>
<td>-1.1 – 12.5</td>
<td>5.3</td>
<td>-1.6 – 12.0</td>
</tr>
</tbody>
</table>

Table 5.1: Mean and range of baseline finger temperature, temperature drop and rise after cold challenge for RP and non-RP subjects in hospital and population settings.

All measurements showed a greater range in the population–based study, reflecting the much larger number of subjects compared to the hospital–based study. The mean readings for baseline, drop and rise were similar in both studies for the non-RP groups. The mean readings for the RP groups, however, suggested milder vasospasm in the population group than the hospital group. Compared to the hospital group, baseline temperature was 4.5ºC warmer in the population–based RP group, and both the drop and rise in temperature after cold challenge were greater. Hence the conclusion of the population–based study was that baseline finger temperature had a low discriminatory power for detecting RP, and cold challenge added little additional value.

The data serve as a sharp reminder that RP cases referred to a specialist centre have, on average, much lower finger temperature than RP cases detected in a population setting. The results of research performed at specialist centres may therefore not be translatable
to community settings. In the community, the report of colour change in the fingers along with numbness or paraesthesia does not appear to correlate with particularly low finger temperature or poor rewarming from cold challenge. Leesmans et al [58] have cast doubt on whether measurements of skin blood flow or temperature correlate with the subjective symptoms of vasospasm.

The key difficulty facing any study into the heritability of RP is the lack of a gold standard definition of the condition [59]. Appendix 6 takes the previously validated infrared thermometer technique and applies it in a twin-based study to provide objective skin temperature data. This can be an adjunct to classifying RP and non-RP subjects on a clinical basis. The study is the first to show that RP, and low finger temperature before and after cold challenge, are all phenotypes that are heritable.

5.1.2. Thermographic assessment of disease activity in LS.

The treatment of LS is one of the biggest challenges in paediatrics, not least because reliable and objective measures of disease activity have not been developed. Without such measures, the true efficacy of proposed therapies [60-63] cannot be ascertained objectively. Appendices 7 and 8 developed the work of others [3,4], applying IRT for LS activity detection in a significant number of patients for the first time. These studies confirmed the utility of IRT for discriminating active from quiescent disease. The research, along with Appendix 9, also identified new limitations of IRT in LS – namely that skin temperature rises in areas overlying lipoatrophy and muscle bulk loss, creating “false-positive” thermograms. A drawback of this work is that it was based on the inspection of thermograms to detect temperature differences, so an element of “subjectivity” remained in the assessment (although there was excellent agreement in the assessment of thermograms between observers).

Appendix 10 proposed a new protocol that recorded skin temperature from carefully defined regions of interest to increase the objectivity of assessment. The protocol also
included LDF – a technique providing a direct measurement of blood flux and hence likely to be less influenced than IRT by changes to underlying skin structure. Appendix 10 also included a report on the reproducibility of the measurements and normal values at healthy skin sites in both adults and children with LS. Whilst other authors have investigated the distribution of temperature [64-66] and laser Doppler flux [67,68] across a variety of healthy skin sites, appendix 10 is the first study to combine the two techniques and include data from substantial numbers of children.

**Appendix 11** confirmed that LDF was a sensitive and specific technique for the detection of active LS sites, but the performance of IRT was disappointing. This is interesting when one considers the utility of IRT demonstrated in appendices 7 and 8. The IRT protocol used in appendix 11 was different from that of the earlier studies, however, and the reliance on temperature measurements from small regions of interest, rather than global visual assessments of each thermogram, may have limited the reliability of the readings. This was in contrast to the expectation that adding objective measurement to the IRT protocol would improve the utility of the technique.

The work in appendix 10 confirmed that the repeatability of temperature readings from thermograms was high when an ROI was placed next to a skin marker repeatedly. Nonetheless, this did not test the ability of the operator to place the markers accurately on the skin itself – a task which is not easy at sites such as the cheek where there are no anatomical landmarks for reference. Limitations in the ability to place the markers at exactly contralateral positions could have affected the IRT results in appendix 11, and drift in the placement of the markers on repeated visits might limit serial IRT measurements using this protocol. LDF may be rather less influenced by this limitation because the probe is repositioned onto each measurement site three times.

Many of the patients recruited for the study in appendix 11 had longstanding disease (>5 years). In these patients there would be an increased likelihood of skin atrophy and
other structural abnormalities. The work of appendices 7, 8, and 9 confirmed that elevated skin temperature can be expected at these sites. Raised skin temperature in longstanding but quiescent lesions in appendix 11 may well have elevated the median skin temperature in the quiescent group, with the result that there was only a difference of 0.2°C in the median contralateral asymmetry in skin temperature between active and quiescent lesions.

5.1.3. Appropriate imager procurement and quality assurance.

It is important to consider the results gained from the studies of RP and LS described in chapters 2 and 3 in the context of the accuracy of radiometric temperature measurement, which is typically stated by thermal camera and infrared thermometer manufacturers as ±2°C at skin temperatures. In practice this means that two identically specified thermal imagers capturing data at different hospitals could record a 4°C difference in the same skin temperature. In a multi-centre trial this would have a major impact on the ability of IRT to detect even the large absolute temperature differences between RP and non-RP subjects reported in appendices 1 and 4. Appendix 11 reported skin temperature differences between patient groups of just 0.2°C: even these measurements of relative difference in skin temperature will be affected by limited imager stability in cases where the difference is calculated from two separate images.

Medical IRT centres in the UK have been active in researching ways to increase traceability and reduce uncertainty in medical IRT. As probably the busiest centre in the UK, (and one of the few using IRT within an NHS framework) the Royal Free Hospital has played an important role in QA work.

Appendix 12 took the requirements for measurement uncertainty and ease of use specified by the Royal Free and other centres, and incorporated them into specifications for the first IFOV blackbody validation system designed specifically for medical IRT. The Royal Free data presented in section 4.3.1 also gave NPL compelling evidence “in-
the-field” that the IFOV system was practicable, and provided an example of how the system corrected the temperatures measured during a cold-challenge procedure, significantly reducing uncertainty.

Appendix 13 is an example of the role that experienced medical IRT centres can play in providing reference material. Prior to the publication of this appendix, little material was available of relevance to modern FPA imagers, and none had considered QA within the context of the UK NHS or European medical device legislation.

5.2 Significance of contribution to knowledge.

The impact of the contribution to knowledge of the projects was assessed by considering citations of the work in papers by other authors. Table 5.2 summarises the citation search: self-citations were excluded from the analysis.

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>AUTHORS</th>
<th>NO. OF CITATIONS</th>
<th>SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Howell et al.</td>
<td>2</td>
<td>Google Scholar</td>
</tr>
<tr>
<td>2</td>
<td>Dziadzio et al.</td>
<td>20</td>
<td>Embase</td>
</tr>
<tr>
<td>3</td>
<td>Coleiro et al.</td>
<td>20</td>
<td>Embase</td>
</tr>
<tr>
<td>4</td>
<td>Cherkas et al.</td>
<td>1</td>
<td>Embase</td>
</tr>
<tr>
<td>8</td>
<td>Martini et al.</td>
<td>5</td>
<td>Embase</td>
</tr>
<tr>
<td>9</td>
<td>Weibel et al.</td>
<td>4</td>
<td>Embase</td>
</tr>
<tr>
<td>11</td>
<td>Weibel et al.</td>
<td>2</td>
<td>Embase</td>
</tr>
</tbody>
</table>

Table 5.2: Citations of the work included within the thesis.
Embase was used as the primary search tool: if no citations were found for a paper, Google Scholar was then also searched.

Anbar [69] cited Appendix 1 as an example of IRT for the assessment of neurogenic vascular dysfunction. Cobb et al. [70] referred to the foot cold challenge protocol: they noted our care to keep the feet dry would be an improvement to their warm immersion protocol for diabetes studies.

Appendices 2 and 3 were both controlled studies of novel drugs for RP, and hence they attracted considerable interest from rheumatologists. Several review articles on available RP treatments cited one or both papers [11,13,71-80]. Both papers met the inclusion criteria of Henness and Wigley [81] for their evidence-based review of secondary RP treatments. Thompson et al. [82] included appendix 2 in their meta-analysis of the effects of calcium-channel blockers in RP. They noted that none of the studies they reviewed included any published temperature data. It might have been better therefore to include the temperature data we collected in our published paper, even though we found no significant drug effects on rewarming. Foerster et al. [83] assessed RP severity using IRT and a VAS questionnaire. They noted that the conclusion of appendix 2 was consistent with their findings that VAS assessment and IRT-monitored rewarming do not always correlate. Merkel et al. [84] commented that a variety of outcome measures are required to assess scleroderma, noting that appendix 2 adopted this approach. Rey et al. [85] commented on appendix 3, and reported their own findings that SSRI treatment can sometimes induce the symptoms of erythromelalgia.

Appendix 4 was cited by Varju et al [86] as an example of the use of IRT for assessment of hand rheumatic diseases.
Appendices 8, 9, and 11 attracted citations in a variety of dermatology journals. Moore et al. [87] used IRT and laser Doppler perfusion imaging (LDPI) for the assessment of LS. They found the highest skin temperature in the oldest LS lesion, and cited appendix 8 as supporting evidence that skin structure changes throughout the disease course. The finding from appendix 11 that LDF is more sensitive than IRT for active LS detection was also cited as justification for the introduction of LDPI. Kowalewski et al. [88] found evidence of variable angiogenesis in LS lesions: even some old lesions demonstrated new vessel generation. The raised skin temperature in some inactive lesions reported in appendix 8 was consistent with their findings. Kreuter and Gambichler [89] noted IRT as described in appendix 8 was an important alternative to assessment of LS with ultrasound. Li et al [90] also noted the IRT work in appendix 9 as an alternative to ultrasonic techniques, whilst commenting that we found IRT of limited value in older lesions. Crespo et al. [91] lamented the lack of availability of IRT at their centre, citing appendix 9 as evidence of the utility of the technique. Kreuter et al. [92] commented that appendix 9 constituted strong evidence for the use of methotrexate and corticosteroids in severe LS. They cautioned that alternative therapies may be more appropriate in mild LS cases. Finally, Kukkonen et al. [93] used IRT in the study of sexual arousal, citing appendix 8 as evidence that IRT can be used to assess blood flow changes.

Taken together, the citations of these seven papers constitute a clear demonstration of the impact of the work and its contribution to knowledge. Appendices 12 and 13 (the work on QA of IRT) had attracted no citations by the time of the thesis submission. This may be in part because this is recent work, or it may reflect the publication of the work in quite low-impact journals with small readerships. The message of the importance of careful QA in clinical IRT is now published for those who wish to read it, but a legislative challenge remains in ensuring all practitioners adopt good practice.
6: Conclusions.

6.1 Fulfilling the thesis aims.

The aim of the thesis, stated in chapter 1, was “to establish standardised infrared thermography (IRT) within a specialist connective tissue disease unit, assessing its utility:

- for the evaluation of Raynaud’s Phenomenon (RP) in clinical rheumatology and research
- for the detection of active localised scleroderma (LS) lesions in paediatric patients

and to develop improvements in IRT quality assurance for these medical applications.”

Chapter 2 described the work with RP. A novel protocol for the evaluation of RP in the feet was developed and assessed. This is still used for RP lower limb assessment, and also has potential utility for the assessment of lower limb erythromelalgia.

IRT with cold challenge was also used for the objective evaluation of response to oral vasodilator therapies in two large pilot studies. Whilst both losartan and fluoxetine were more effective than nifedipine at reducing RP symptoms, only fluoxetine produced significant improvements in response to cold challenge over nifedipine in female primary RP patients. Fluoxetine may well be a more effective RP therapy than losartan in this group, but differences in the design of the two studies need to be taken into account. The fluoxetine study was a crossover design, which may have benefits over parallel group studies when looking for small changes in cold challenge responses. It would have been helpful to publish our temperature data showing no significant effects from losartan over nifedipine given the comment by Thompson et al. [82] on the paucity of published temperature data.
Comparison of IRT to an infrared thermometer showed that the two techniques were of equal utility for RP detection, although the latter was much more labour-intensive. The infrared thermometer method was found to be of some utility for the detection of RP in a population-based setting, and for confirming in a twin population that the vasospastic phenotype is heritable.

The difference in the infrared thermometer temperature data recorded from RP patients in the hospital and population settings was an important result. RP detected in a hospital setting was associated with much lower finger skin temperatures than RP detected in the general population. This might have been presupposed when one considers that RP cases referred to hospital are likely to suffer the more severe vasospasm, but the confirmation of this supposition is not only of scientific interest: it also impacts on the planning of clinical services.

Chapter 3 covered our extensive research into paediatric LS activity detection with IRT. Two papers showed that simply inspecting thermograms for asymmetry in temperature was sensitive for active LS. Specificity of IRT was limited, however, by “false-positive” thermograms of atrophic areas – a new finding.

Taking this limitation into consideration an additional technique, LDF, was employed in conjunction with IRT as a measure of skin blood flow. LDF was found to be more effective for active LS detection than IRT, albeit with the major limitation that LDF, being a point measurement, cannot map areas of skin blood flow.

The large number of patients (for such a rare disorder) included in our LS studies, coupled with our novel approach to objective measurement of disease activity, means that the work continues to be regularly cited by other authors.

Chapter 4 addressed the quality assurance of medical IRT: making thermographic measurements reliable and traceable. The Royal Free’s contribution to the design and evaluation of a portable IFOV calibration system was described. Practical
measurements with the IFOV system were presented to demonstrate that measurement uncertainty is reduced by a factor of about four. With this device in situ at each unit, reliable multi-centre trials are now achievable.

A review paper sought to reinforce the importance of QA to the medical IRT community. This paper is now cited as a key reference document by the EAT web site.

6.2 Future work.

In this section, future developments of the work described within this thesis are briefly considered.

As new Raynaud’s phenomenon therapies become available they should be evaluated with appropriately-designed, controlled studies that employ thermography as an objective outcome measure. Skin temperature measurement is instructive for RP evaluation in a population setting, but discriminating RP from normal subjects will be more challenging than in a hospital setting.

The Royal Free is now a recognised centre of expertise in the physiological measurement of localised scleroderma. The research into LS assessment highlights the challenge of identifying the “ideal” instrument for assessing microvascular blood flow. No such instrument currently exists, and so for the foreseeable future a variety of validated techniques will need to be employed in combination to provide reliable information about disease activity.

IRT continues to play a significant role in LS assessment: it gives an instantaneous, wide-field view of the inflammatory process, and is well tolerated by patients. The low specificity of IRT in older lesions is however an important limitation. IRT measures skin temperature, whereas the ideal imaging technique would measure skin blood flow. LDPI has shown promise for LS assessment [87] but its utility is limited by scan times of several minutes per frame. LDPI devices using dual or multiple wavelengths may
supply useful information about microvascular blood flow at different skin depths. Full-field laser perfusion imaging (FLPI) [94], a laser-speckle technique, has the advantage over LDPI of near-instantaneous frame rates, and has shown early promise at the Royal Free as a possible assessment tool for LS. A pilot trial of FLPI for LS is planned.

The work on quality assurance of IRT is a valuable achievement leading to both a practicable device for the traceable validation of medical IRT, and reference material that practitioners can access to learn about correct QA procedures. The IFOV device has still to be produced commercially, however, and will only be a viable product if significant numbers of users adopt it.

Connective tissue disorders remain poorly understood diseases of unknown aetiology. Evaluating the microvascular component of these conditions, and their impact on skin nutrition and thermoregulation, is vital to the research effort in rheumatology and dermatology. The need to measure human skin temperature radiometrically, and to have trust in the result, remains of paramount importance.
7 : Bibliography

References


5 Raynaud M: De l'asphyxie locale et de la gangrène symétrique des extrémités. 1862.


50 Ammer K: Repeatability of temperature measurements at the forehead in thermal images from the standard view "face". Thermology International 2006, 16:138-142.


60 Hulshof M, Bouwes Bavinck J, Bergman W, Masclee A, Heickendorff L, Breedveld F, Dijkmans B: Double-blind, placebo-controlled study of oral calcitriol


64 Uematsu S: Symmetry of skin temperature comparing one side of the body to the other. *Thermology* 1986, **1**:4-7.


Appendix 1

Statement of Support

I confirm the following contributions from Kevin Howell to the publication below:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Data collection (80%)</td>
</tr>
<tr>
<td></td>
<td>• Statistical analysis (50%)</td>
</tr>
<tr>
<td></td>
<td>• Preparation of manuscript (80%)</td>
</tr>
</tbody>
</table>

Dame Carol M BLACK

Professor of Rheumatology

Royal Free Hospital

London

24th August 2009
Temperature of the toes in Raynaud's phenomenon measured using infra-red thermography

K.J. Howell, L.F. Kennedy, R.E. Smith, C.M. Black

Academic Department of Rheumatology and Connective Tissue Diseases and + Medical Physics Department, Royal Free Hospital, London, UK.
*London Foot Hospital.

Summary

Aims of the study: To investigate the response to cold challenge of the feet of healthy individuals and patients with primary Raynaud's phenomenon.

Subjects: 6 primary Raynaud's phenomenon patients with symptoms in their feet (R) and 8 healthy controls (H). All were female.

Method: After adaptation to an ambient temperature of 23 ± 1 °C, a baseline thermal image of the dorsal surface of both feet was captured. Further images were taken at one minute intervals for 10 minutes after both feet were cold-challenged by immersion in water at 15 °C for 1 minute. Mean toe temperature, θ and the temperature difference between the hallux and the 5th toe, ΔT were calculated from the images taken at baseline, and at 0, 5 and 10 minutes after cold-challenge.

Results: The toes of H were significantly warmer than R at baseline (29.2 ± 1.5 °C v 24.8 ± 1.5 °C (mean ± SD), p<0.01; t-test). For R post challenge values were strongly correlated with baseline temperature (r=0.96 for 0 and 5 mins, r=0.92 for 10 mins). The correlation was less strong for H. At baseline, ΔT was significantly greater in H than R (0.7 [2 - -0.4]°C v -0.5 [0.4 - -3.6] °C, median and range, p<0.02; Mann-Whitney) and this was exaggerated 10 minutes post challenge (1.4 [3.9 - -1.9] °C v -1.5 [-0.5 - -2.5] °C, median and range, p<0.02; Mann-Whitney)

Conclusions: Mean toe temperature and medial-lateral temperature difference are important diagnostic indicators of Raynaud's phenomenon in the feet. Cold challenge may enhance the utility of these indicators.

Key words: Raynaud's phenomenon, foot, infra-red thermography, cold challenge.

Die Messung der Zehen temperatur mittels Infrarotthermographie bei Patienten mit Raynaud-Phänomen

Ziel der Studie war die Untersuchung der Reaktion der Füße auf eine Kältebelastung bei Gesunden und Patienten mit primärem Raynaudphänomen.

Untersucht wurden 6 Patientinnen mit primärem Raynaudphänomen (R), die Symptome auch an den Füßen zeigten, und 8 gesunde weibliche Kontrollpersonen (H).

Methoden: Nach Adaptierung an eine Umgebungstemperatur von 23 ± 1 °C wurden die Ausgangsbilder von der Dorsalseite beider Füße aufgezeichnet. Nach einer Kältebelastung durch ein 1 Minute dauerndes Wasserbad von 15°C wurden durch 10 Minuten hindurch in einminütigem Abstand weitere Bilder aufgezeichnet. Die mittlere Zehentemperatur θ und die Temperaturdifferenz zwischen Groß- und Kleinzehe, ΔT wurde an den Ausgangsbildern, unmittelbar, 5 und 10 Minuten nach der Kältebelastung errechnet. Ergebnisse: Die Zehen von H waren vor der Kältebelastung signifikant wärmer als bei R (29.2 ± 1.5 °C v 24.8 ± 1.5 °C (Mittelwert ± Standardabweichung), p<0.01; t-Test). Die Werte nach Kältebelastung von R korrelierten deutlich mit den Ausgangswerten (r=0.96 für 0 und 5 min, r=0.92 für 10 min). Diese Korrelation war für H weniger ausgeprägt. ΔT war signifikant größer bei H als bei R (0.7 [2 - -0.4]°C v -0.5 [0.4 - -3.6] °C, Median und Bereich, p<0.02; Mann-Whitney) und war am deutlichsten 10 Minuten
Introduction

Raynaud’s phenomenon is a common vascular disorder characterised by intermittent spasm of the peripheral blood vessels in response to cold stimulus or emotional stress, primarily affecting the fingers and toes (1). Sufferers experience phasic colour changes to the skin of the affected parts. Severity of symptoms varies between individuals, ranging from an occasional awareness of cold and white digits, to a persistent state of discoloration, numbness and pain. In the most severe cases tissue loss due to ulceration or gangrene may be experienced.

Primary Raynaud’s phenomenon describes a benign manifestation of symptoms which nevertheless can be severely disabling, whilst secondary Raynaud’s phenomenon comprises the same symptoms seen in association with another disease process. The pathology of Raynaud’s phenomenon is unknown, although a combination of factors have been implicated, such as the function of the vessel walls, composition of the blood, and neurological innervation of the vasculature. A role for sex hormones has also been suggested, since the condition predominantly affects women (2). Surprisingly, the prevalence and clinical importance of Raynaud’s phenomenon in feet has received little attention.

Infrared thermography is used as a clinical tool in the assessment of Raynaud’s phenomenon. Using skin surface temperature as an indirect appraisal of blood perfusion, it has been used to delineate the circulatory changes which occur in the hands in response to cold challenge (3). Characteristic temperature distribution and rewarming patterns have been demonstrated in the Raynaud’s phenomenon hand which can aid diagnosis and help to evaluate the efficacy of treatment. There is, however, a paucity of thermographic data available regarding the response of the Raynaud’s phenomenon foot to mild cold challenge. The objective of our study was to investigate the response to cold challenge of the feet of healthy individuals and patients with primary Raynaud’s phenomenon.

Method

Six primary Raynaud’s phenomenon patients with foot symptoms (R) and 8 healthy controls (H) took part in the investigation. All subjects were female, and all gave their informed consent prior to participation in the study. The median age of R was 32 years (range 18 to 56), while the median age of H was 25 years (range 22 to 56). The Raynaud’s patients were recruited from attendees at Rheumatology outpatient clinics at the Royal Free Hospital.

All thermographic studies were performed in the winter months at the Blood Flow Studies laboratory of the Academic Department of Rheumatology and Connective Tissue Diseases. Subjects were asked to refrain from smoking and drinking hot beverages or alcohol for a period of two hours prior to the measurements.

Upon entering the laboratory, subjects removed all footwear and hosiery and sat upright in a chair with the feet dependent and the soles resting on a wire base. This posture was maintained for 15 minutes to allow each subject time
to adapt to the ambient laboratory temperature of 23 ± 1 °C (figure 1).

The thermal imaging system used was the STARSight pyroelectric vidicon camera (Insight Vision Systems, Great Malvern, UK) (4). All images were digitised using a frame-grabber and stored on an IBM-compatible personal computer for later software analysis.

After equilibration a baseline thermal image of the dorsal surface of both feet was captured. The feet were then placed in thin plastic bags and immersed up to the ankles in water at 15 °C for one minute, with the subject remaining seated throughout. Care was taken to ensure the feet did not become wet. The feet were then removed from the water and bags, and returned to the same position maintained during equilibration. Further thermal images were then recorded at one minute intervals for a total of ten minutes.

Toe temperature analysis was performed on each subject using the baseline thermal images, and those recorded 0, 5 and 10 minutes after cold challenge. The temperature of each toe was deduced by specifying ten regions of interest in each image using the freehand drawing utility within the Thermosoft thermal image analysis software (EJC, Jenison, MI, USA).

Using a computer spreadsheet, the mean temperature of all ten toes (q) was calculated for each image. An index for medial-lateral temperature difference (ΔT) was also calculated, this being the mean temperature of both fifth toes subtracted from the mean temperature of both halluces.

**Results**

At baseline the healthy controls exhibited a significantly higher mean toe temperature (θ) than the Raynaud’s phenomenon patients (29.2 ± 1.5 °C v 24.8 ± 1.5 °C (mean ± SD), p< 0.01; t-test).

For patients, θ post cold challenge correlated strongly with θ at baseline (r= 0.96 at 0 minutes, r= 0.96 at 5 minutes, and r= 0.92 at 10 minutes) (figure. 2).

The correlation was less strong for healthy controls (r= 0.75 at 0 minutes, r= 0.66 at 5 minutes, and r= 0.52 at 10 minutes).

Medial-lateral temperature difference DT was significantly greater in the healthy control group than the Raynaud’s group at baseline (0.7 [2 -0.4] °C v -0.5 [0.4 -3.6] °C, median and range, p< 0.02; Mann-Whitney) (figure 3).

The difference in ΔT between the groups was slightly increased at 5 minutes post cold-challenge, and by ten minutes post cold-challenge had reached 2.9 °C (1.4 [3.9 -1.9] °C v -1.5 [-0.5 - -2.5] °C, median and range, p< 0.02; Mann-Whitney).

There were weak correlations between ΔT and mean toe temperature θ at baseline.

Figure 4 shows thermograms of a typical pair of healthy feet before cold challenge, and at 0, 5 and 10 minutes after cold challenge. Figure 5

![Figure 2](image1.png)

**Figure 2**
Mean toe temperature θ at 0, 5, and 10 minutes post cold challenge as a function of baseline temperature for healthy subjects and patients with Raynaud’s phenomenon

![Figure 3](image2.png)

**Figure 3**
Medial-lateral temperature difference (Dt) in the feet of healthy subjects (H) and patients with Raynaud’s phenomenon (R) before and after cold challenge
Figure 4
Thermograms of the feet of a typical healthy subject, taken before cold challenge and at 0, 5, and 10 minutes post cold challenge.

Figure 5
Thermograms of the feet of a typical Raynaud's phenomenon patient, taken before cold challenge and at 0, 5, and 10 minutes post cold challenge.
Discussion

Our data suggest that the toes of subjects with primary Raynaud’s phenomenon are colder than those of healthy individuals, both before and after a standardised cold challenge. The complete recovery of toe temperature was not achieved within ten minutes of cold challenge, even in the healthy control group. This observation is in contradistinction to the response of normal hands (5).

There is a strong correlation between mean toe temperature at baseline and at all times post cold-challenge in the Raynaud’s group. Equilibrated mean toe temperature (baseline temperature) may, by itself, be diagnostic of Raynaud’s phenomenon in the feet. Recent work by other authors suggests that this is not the case when considering finger temperature after cold-challenging the Raynaud’s phenomenon hand: although baseline finger temperature is higher in normal subjects than in Raynaud’s patients, cold challenge greatly improves diagnostic accuracy (6).

In comparing mean toe temperature at baseline with that at various time points post challenge, it is instructive to plot the data for both groups, H and R, on the same axes (fig. 2). The data points for Raynaud’s patients lie without exception to the left of the healthy group, since all subjects in R had a lower baseline mean toe temperature than the subjects in H. An initial re-warming between 0 and 5 minutes occurs in both groups, but little further heating of the toes occurs between 5 and 10 minutes in the Raynaud’s group, giving rise to a set of data points for R which is not noticeably elevated up the y-axis at 10 minutes in comparison to 5 minutes. In contrast, re-warming in healthy controls continues unabated between 5 and 10 minutes post cold-challenge.

Immediately post cold challenge (0 minutes), the data points for both groups can be considered to lie on the same straight line. However, the gradients of the straight lines describing toe temperature at 5 and 10 minutes post cold challenge as a function of baseline temperature appear to differ markedly between the two groups. There is a considerably greater spread of the data points about the straight lines in H, resulting in the poorer correlation between pre and post toe temperatures in that group. These differences between the groups (in the slopes of the plotted functions and in the spread of the data) suggest that the mechanism by which Raynaud’s affected toes rewarm is fundamentally different to that of toes in healthy subjects.

Medial-lateral temperature difference DT was significantly different in H and R at baseline and at all time-points post cold challenge (figure 3). The Raynaud’s group tended to have halluces colder than the 5th toes, resulting in a negative value for DT as defined above. In contrast, the trend amongst healthy individuals was for the halluces to be warmer than the 5th toes throughout the test. Any model seeking to describe the differing mechanisms of rearming in Raynaud’s and healthy subjects must also strive to explain the observed medial-lateral temperature differences.

We postulate that there are two heat flows into the toes, one by a conduction mechanism from the proximal part of the foot, and the other arising from the blood supply to individual toes. Radiographic evidence suggests that in the human foot the skin of the medial toes is more densely populated with thermoregulatory microvessels than the lateral side (7). Such vascular anatomy may tend to enhance the difference in DT between Raynaud’s and healthy subjects. In subjects where toe bloodflow is inherently high, it will always tend to be highest in the hallux, where the vascularisation is richer.

In the healthy foot, the flow of blood throughout the toes is the predominant process of heat transfer both before and soon after cold challenge. The microvessels of the toes are dilated and the blood supply good. In this situation the hallux might be expected to be the warmest toe since it has the smallest surface area to volume ratio.

In the Raynaud’s phenomenon foot, the conductive process of heat transfer to the toes predominates both before, and especially after cold challenge. The microvessels of the toes are constricted and the blood supply poor. In this situation the lateral toes might be expected to be the warmer digits, since they are closer to the proximal part of the foot, and will thus benefit most from the conduction of heat from this area.

We conclude that baseline mean toe temperature (θ) and medial-lateral toe temperature difference (ΔT) are good diagnostic indicators of
Raynaud’s phenomenon in the feet. In addition, cold challenge might enhance the utility of thermography, since θ recovers markedly better over a ten minute period in healthy individuals than in Raynaud’s patients.

References


Address of the author:
Mr. Kevin HOWELL
Royal Free Academic Unit of Rheumatology and Connective Tissue Diseases
Rowland Hill Street,NW3 2PF London
UNITED KINGDOM

(Manuscript received on 2.5.1997, accepted on 25.5.1997)
Appendix 2

Statement of Support

I confirm the following contributions from Kevin Howell to the publication below:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dziadzio M, Denton CP, Smith R, Howell K, Blann A, Bowers E, Black CM.</td>
<td>• Thermography and laser Doppler protocol design</td>
</tr>
<tr>
<td>Losartan therapy for Raynaud's phenomenon and scleroderma: clinical and</td>
<td>(100%)</td>
</tr>
<tr>
<td>biochemical findings in a fifteen-week, randomized, parallel-group,</td>
<td>• Thermography and laser Doppler data collection</td>
</tr>
<tr>
<td>controlled trial. Arthritis Rheum 1999; 42: 2646-2655</td>
<td>(100%)</td>
</tr>
<tr>
<td></td>
<td>• Thermography data analysis (100%)</td>
</tr>
<tr>
<td></td>
<td>• Preparation of manuscript (5%)</td>
</tr>
</tbody>
</table>

Dr M DZIADZIO

Royal Free Hospital

London

24th August 2009
LOSARTAN THERAPY FOR RAYNAUD’S PHENOMENON AND SCLERODERMA

Clinical and Biochemical Findings in a Fifteen-Week, Randomized, Parallel-Group, Controlled Trial

MAGDALENA DZIADZIO, CHRISTOPHER P. DENTON, ROY SMITH, KEVIN HOWELL, ANDREW BLANN, EMMA BOWERS, and CAROL M. BLACK

Objective. To compare the efficacy and tolerability of losartan, an antagonist of angiotensin II receptor type 1, with nifedipine for the treatment of primary and secondary Raynaud’s phenomenon (RP) in a pilot study.

Methods. In a randomized, parallel-group, controlled trial, patients with primary RP (n = 25) or RP secondary to systemic sclerosis (SSc [scleroderma]; n = 27) were allocated to receive 12 weeks’ treatment with either losartan (50 mg/day) or nifedipine (40 mg/day). Primary outcome variables were the severity and frequency of RP episodes and findings on vascular measurements, including thermography and laser Doppler flowmetry. Serum levels of soluble adhesion molecules, endothelin 1, fibrinogen, von Willebrand factor, and procollagen type I N-terminal propeptide (PINP) were also measured.

Results. There was a reduction in the severity of RP episodes following treatment with losartan and with nifedipine, but this effect was greater in the losartan arm of the study (P < 0.05): episode frequency was reduced only in the losartan group (P < 0.01 versus baseline). Symptomatic improvement was associated with a significant reduction in soluble vascular cell adhesion molecule 1 and PINP (P < 0.01). Subgroup analysis suggested that although these biochemical changes occurred mainly in SSc patients, the clinical benefit was greater in the primary RP group.

Conclusion. This study confirms the tolerability of short-term treatment of RP with losartan, and our data suggest its clinical benefit. Further evaluation of this drug as a long-term treatment for SSc-associated RP should be considered, since it may have additional disease-modifying potential.

Episodic digital ischemia provoked by cold and emotion was first described by Maurice Raynaud over 130 years ago (1). When it occurs in isolation, it is designated primary Raynaud’s phenomenon (RP) to distinguish it from those cases in which there is an underlying or associated pathology. It affects ~10% of the adult population, with a predilection for females, and up to 5% of patients presenting with this condition later develop an autoimmune rheumatic disorder, such as systemic sclerosis (SSc; scleroderma) (2). Successful treatment is often difficult, and although clinical trials suggest that vasodilators can be effective, the responses of individual patients to a particular agent are often idiosyncratic.

The pathogenesis of the vascular dysfunction underlying RP is incompletely understood, and relatively few well-controlled clinical trials with vasoactive drugs have been undertaken. Lack of distinction between primary and secondary RP in some of these studies makes their interpretation difficult. Calcium channel antagonists such as nifedipine have been shown to be effective (3–6), although striking differences in individual responses have been described (3,7). Moreover, efficacy is often particularly limited in patients with RP.
secondary to SSc (8), and side effects such as dizziness, hypotension, ankle edema, and headache are common (9,10). There is therefore a need to evaluate alternative treatments that might be more effective or have a better side-effect profile.

Losartan is a specific nonpeptide angiotensin II type 1 receptor antagonist (11) that has been successfully used to treat systemic hypertension (12). It does not exhibit the adverse respiratory side effects, such as cough, that are common to angiotensin-converting enzyme (ACE) inhibitors (13), which have also been shown to be of controversial benefit in RP (14). A small number of patients with primary RP have been treated with low doses of losartan, and a significant improvement in RP episodes was demonstrated in that study (15). The efficacy of losartan in secondary RP has not been studied, although 1 case report did not demonstrate its benefit in the treatment of established scleroderma renal crisis (16).

Here we report the results of a pilot study comparing the tolerability and efficacy of losartan with nifedipine in a group of patients with primary RP or RP secondary to SSc. The effect of these drugs on serum markers of vascular damage and connective tissue turnover was also investigated to evaluate their possible disease-modifying potential.

PATIENTS AND METHODS

Study design. This was a randomized, parallel-group study of 15 weeks’ duration. The study was performed over 1 winter to minimize seasonal effects on the severity and frequency of RP episodes. Patients stopped taking vasodilator drugs 3 weeks prior to entry into the trial (week 0) and underwent baseline medical assessment, including medical history and physical examination.

Graded nailfold capillaroscopy was used at baseline as a confirmatory test for primary RP; the abnormalities were quantified using a standard scale as grade I, II, or III (17). Vascular evaluation, including cold-challenge infrared thermography and laser Doppler flowmetry (LDF), were performed at baseline (week 3) and at the end of treatment (week 15). Serum and plasma samples were taken before and at the completion of therapy.

Patients were randomized to receive either nifedipine or losartan for 12 weeks. Nifedipine was formulated as 20-mg nifedipine retard tablets and were taken twice a day. Losartan was formulated as 50-mg tablets (Merck, Sharp and Dohme, Hertford, UK) and were taken once a day. The relatively low dosage of nifedipine was to minimize side effect–related withdrawals from the control group.

Patients. Sixty patients from the outpatient clinic at the Rheumatology Department, Royal Free Hospital, who had confirmed RP were screened for the study. Fifty-two patients were recruited into the study; 8 patients did not fulfil the inclusion criteria. The demographic characteristics of the study patients are summarized in Table 1.

RP was diagnosed on the basis of a history of episodic digital vasospasm with triphasic color changes. For inclusion in the study, patients had to have, on average, at least 6 episodes of RP per week. Absence of rheumatoid factor, antinuclear antibodies, or disease-specific antibodies (anticentromere, anti–Scl-70, anti–U3 RNP, and anti–polymerase I and III antibodies), as well as absence of significant nailfold capillaroscopic abnormalities were required for inclusion to the primary RP group.

Patients with features of undifferentiated connective tissue disease were not recruited. Patients under the age of 18 or over the age of 70 and women of childbearing age who were not using adequate contraception were excluded. Patients with a history of significant cardiorespiratory disease or renal impairment and those who were taking ACE inhibitors were also ineligible. Five of the patients who were screened had either positive autoantibodies (4 patients) or a capillary score of III (1 patient) in the absence of a defined connective tissue disease. Since these cases, which are sometimes designated “autoimmune RP,” may represent a distinct clinical group, these 5 patients were not included in the trial.

Of the patients entered into the study, 25 (10 males and 15 females; median age 49 years, range 19–63) had primary RP. The remaining 27 patients (3 males and 24 females; median age 54 years, range 30–68) had RP secondary to SSc, as classified by the preliminary criteria of the American College of Rheumatology (formerly, the American Rheuma-

Table 1. Baseline characteristics of losartan and nifedipine treatment groups

<table>
<thead>
<tr>
<th>Demographic variables</th>
<th>Losartan</th>
<th>Nifedipine</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of patients enrolled</td>
<td>26</td>
<td>26</td>
<td>52</td>
</tr>
<tr>
<td>No. of females:males</td>
<td>22:4</td>
<td>17:9</td>
<td>39:13</td>
</tr>
<tr>
<td>PRP</td>
<td>10:2</td>
<td>5:8</td>
<td>15:10</td>
</tr>
<tr>
<td>SSc</td>
<td>12:2</td>
<td>12:1</td>
<td>24:3</td>
</tr>
<tr>
<td>Median age, years (range)</td>
<td>51 (21–62)</td>
<td>51 (19–68)</td>
<td>51 (19–68)</td>
</tr>
<tr>
<td>PRP</td>
<td>49 (21–61)</td>
<td>49 (19–63)</td>
<td>49 (19–63)</td>
</tr>
<tr>
<td>SSc</td>
<td>55 (30–62)</td>
<td>53 (36–68)</td>
<td>54 (30–68)</td>
</tr>
<tr>
<td>No. of current smokers</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>No. of ex-smokers</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Clinical characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. with PRP</td>
<td>12</td>
<td>13</td>
<td>25</td>
</tr>
<tr>
<td>No. with SSc</td>
<td>14</td>
<td>13</td>
<td>27</td>
</tr>
<tr>
<td>lcSSc</td>
<td>5</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>dcSSc</td>
<td>9</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>No. of patients who have already tried nifedipine</td>
<td>6</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>No. of patients who have tried other vasodilators</td>
<td>10</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>No. of patients who have not tried any vasodilators</td>
<td>10</td>
<td>12</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 1. Baseline characteristics of losartan and nifedipine treatment groups

* PRP = primary Raynaud’s phenomenon; SSc = systemic sclerosis; lcSSc = limited cutaneous systemic sclerosis; dcSSc = diffuse cutaneous systemic sclerosis.
tism (ASS) (18). Eighteen patients had limited cutaneous SSc (lcSSc) and 9 had the diffuse cutaneous SSc (dcSSc).

Patients with primary and secondary RP were separately randomized to receive 1 of the 2 drugs to maintain equal proportions in the 2 treatment arms. Ethical approval for the study was obtained from the Royal Free Hospital Ethical Practices Committee, and informed consent was given by all patients.

Assessment of RP symptoms. Patients were given symptom diaries and asked to select 1 day each week for recording their symptoms throughout the 12 weeks of treatment. An episode of RP was defined as the occurrence of pallor followed by cyanosis, with or without associated pain. Patients recorded symptom severity using a visual analog scale ranging from 0–10, where 10 = the worst episode ever experienced and 0 = no episodes. The daily frequency of RP episodes was documented on the same day. Symptom diaries were collected at the last visit. Patients were asked to report any adverse events occurring during the study.

Noninvasive vascular studies. At baseline and after treatment, the response of each patient to a mild cold challenge of the hands in water was assessed using the techniques of LDF and infrared thermography. All participants were asked to avoid alcohol for 24 hours prior to the measurements, as well as hot caffeinated drinks and hot meals on the day of the vascular tests.

LDF refers to the measurement of the flux of moving red blood cells in a volume of skin beneath a measurement probe, and is derived from the Doppler shift of the frequency of laser light as it scatters from moving erythrocytes. LDF flux is measured in arbitrary units (AU). Infrared thermography estimates skin temperature using a thermal imaging camera.

At each visit, the subject sat comfortably in a temperature-controlled room (within 1°C of 23°C) for 15 minutes before measurements were taken. Laser Doppler skin probes were attached to the pulp of each middle finger. An initial thermal image of both hands was recorded using the Starsight Thermal Camera (Insight Vision Systems, Malvern, UK). Laser Doppler flux was then recorded continuously for a period of 5 minutes to ensure that skin flux had reached a stable level during the equilibration period prior to measurement (MBF3D Dual Channel Blood Flow Monitor; Moor Instruments, Devon, UK). This recording period was extended, where necessary, until stable flux readings from both fingers were observed for a period of 2 minutes.

Prior to mild cold challenge of the hands, a further thermal image was recorded. Both hands were then gloved, leaving the laser Doppler probes in place, and immersed in water at 15°C for 1 minute. After cold challenge, the hands were returned to their original positions, and thermal images were taken at 1-minute intervals for a 10-minute period. The laser Doppler trace was recorded from both fingers throughout this period.

Laser Doppler settings for all subjects were as follows: time constant 1 second and bandwidth 14.9 kHz. Laser Doppler traces were transferred to a personal computer for further analysis. The LDF mild cold challenge recovery curves from each hand were analyzed by calculating the average values of flux over a 10-second period at 3 time points: just prior to cold challenge and immediately after, 5 minutes after, and 10 minutes after cold challenge. The thermographic images were analyzed (Thermosoft image analysis software package; Moor Instruments) by computing the mean fingertip temperature (excluding the thumbs) at the same time points as for LDF.

Serum and plasma markers. Blood samples were collected immediately before and on completion of therapy. Serum and plasma were obtained by centrifugation of whole blood at 3,000 g for 10 minutes, and aliquots were stored at –20°C. To investigate changes in underlying disease mechanisms and to identify serologic factors for assay in future studies, biochemical markers previously suggested to be abnormal in patients with SSc or RP were measured. Thus, levels of circulating soluble markers of endothelial cell function, including soluble isoforms of intercellular adhesion molecule 1 (sICAM-1), vascular cell adhesion molecule 1 (sVCAM-1), E-selectin (E-selectin), von Willebrand factor (vWF), and fibrinogen, as well as procollagen type I N-terminal propeptide (PINP) were evaluated.

### Table 2. Effect of losartan or nifedipine therapy on clinical variables

<table>
<thead>
<tr>
<th>End point, group</th>
<th>Losartan</th>
<th>Nifedipine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Severity of RP episode</strong>&lt;br&gt;(0–10 scale)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole group</td>
<td>Week 3, mean ± SD</td>
<td>Week 15, mean ± SD</td>
</tr>
<tr>
<td>SSc patients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRP patients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P (SSc vs. PRP)</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>Frequency of RP episode&lt;br&gt;(no. of episodes/day)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole group</td>
<td>Week 3, mean ± SD</td>
<td>Week 15, mean ± SD</td>
</tr>
<tr>
<td>SSc patients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRP patients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P (SSc vs. PRP)</td>
<td>0.055</td>
<td></td>
</tr>
</tbody>
</table>

*RP = Raynaud’s phenomenon; SSc = systemic sclerosis; PRP = primary Raynaud’s phenomenon.
†Mean of individual patients, each expressed as a percentage of his or her own baseline.
‡P < 0.01, by paired t-test.
§P < 0.05, by paired t-test.
Commercially available enzyme-linked immunosorbent assay (ELISA) test kits were used for sICAM-1, sVCAM-1, and E-selectin (R&D Systems, Oxford, UK) and for ET-1 (Cozart Bioscience, Abingdon, UK). Serum concentrations of PINP were evaluated using a commercially available radioimmunoassay kit (Orion Diagnostica, Espoo, Finland). Fibrinogen was measured by a modified Clauss technique using bovine thrombin and a Thromboscreen T400C coagulometer (both from Pacific Hemostasis, Huntersville, NC). Levels of vWF were determined by ELISA according to a previously described assay (19).

**Statistical analysis.** Clinically significant improvement in RP episodes was defined as a 20% improvement in the variables documented in the symptom diaries at baseline. Severity and frequency scores for baseline (week 3) and for each of weeks 6, 9, 12, and 15 during treatment were assessed by analysis of variance (ANOVA). Similarly, severity and frequency scores normalized to 100% at baseline were also subject to ANOVA. Differences in the frequencies of adverse effects were tested by chi-square test. Pre- and posttreatment values of the biochemical variables were compared using Student’s paired t-test to assess the impact of therapy. The laser Doppler and thermographic data were analyzed by the use of ANOVA, followed, where appropriate, by Student’s t-test to compare groups. Where multiple testing was carried out, a more rigorous statistical significance was used (*P* < 0.01, rather than *P* < 0.05).

**RESULTS**

**Clinical variables and serum markers.** Baseline demographic and clinical variables for the 2 treatment arms are shown in Table 1. There were no statistically significant differences between the 2 treatment groups, save in the distribution of lcSSc and dcSSc (*P* < 0.006, by chi-square test). In particular, the preponderance of men in the nifedipine group was not statistically significant.

Comparison of the variables at baseline with those after the 12-week treatment period is shown in Table 2. These data demonstrate the mean change from baseline together with the mean percentage of change over the study period. Analysis of the symptom diaries revealed a significant mean reduction in RP severity (49%; *P* = 0.0003) and frequency of episodes (50%; *P* = 0.009) over the treatment period for patients receiving losartan (Figure 1). Although in the nifedipine group, there was a mean reduction of 18% in RP severity (Figure 2), this was not statistically significant (*P* = 0.49). There was also a nonsignificant increase in the frequency of RP episodes in this group (*P* = 0.52).

An idiosyncratic response of patients to nifedipine treatment was seen in this study. Overall, for nifedipine, 11 of 26 patients showed improvement in the severity score between weeks 3 and 15, compared with 21 of 26 for losartan (*P* < 0.04, by chi-square test). Similarly, for frequency of episodes, only 7 of 20 patients improved with nifedipine therapy, whereas 18 of 25 improved with losartan (*P* < 0.02, by chi-square test). When primary RP and SSc groups were considered separately, the only significant finding was for the frequency of RP episodes in the primary RP group: 9 of 12 patients taking losartan, but only 2 of 13 patients taking nifedipine, showed improvement (*P* < 0.03, by chi-square test).

Figures 3 and 4 are plots of normalized severity and frequency scores. These data were subjected to ANOVA, and significant effects (*P* ≤ 0.05) were investigated using the unpaired t-test. The improvement in the severity of the episodes in the losartan group was

![Figure 1](image1.png)

**Figure 1.** Effect of losartan on the raw severity score and frequency of episodes of Raynaud’s phenomenon. The data points for each patient are connected by a line. Some patients had identical responses; therefore, the number of lines is fewer than the number of patients in each group. Short horizontal lines show the group mean. *P* values determined by paired t-test.
observed after 3 weeks of treatment (week 6 of the study) and subsequently enhanced, with a significant reduction ($P < 0.002$, by $t$-test) in the score at week 15 (Figure 3). A reduction in the severity score in the nifedipine group was observed only after 9 weeks of treatment (week 12 of the study) (Figure 3), but did not reach significance by week 15. The difference between groups was significant at week 15 ($P < 0.05$, by $t$-test).

The improvement in the episode frequency, as mentioned above, was observed only in the losartan group. This occurred between weeks 9 and 12 and was maintained at the end of treatment (Figure 4). There was a statistically significant difference between treatments at week 15 ($P < 0.02$, by $t$-test). However, no steady state in the normalized severity and frequency scores at week 15 was reached in any of the groups (Figures 3 and 4).

This improvement in symptoms in patients receiving losartan was accompanied by a significant reduction in serum levels of circulating PINP ($P = 0.009$) and VCAM-1 ($P = 0.003$) (Figure 5 and Table 3). Similar trends were observed for ET-1 and sICAM-1 (18% and 11% reduction in posttreatment levels, respectively). Nifedipine significantly reduced serum concentrations of sVCAM-1 only ($P = 0.01$), with no effect on other biochemical variables. There was no reduction in the

---

**Figure 2.** Effect of nifedipine on the raw severity score and frequency of episodes of Raynaud's phenomenon. The data points for each patient are connected by a line. Some patients had identical responses; therefore, the number of lines is fewer than the number of patients in each group. Short horizontal lines show the group mean. $P$ values determined by paired $t$-test.

**Figure 3.** Effect of losartan and nifedipine on the severity of episodes of Raynaud’s phenomenon. The severity score was normalized to 100% at baseline. Bars show the mean and 1 SEM. $P$ values determined by unpaired $t$-test. $NS = not significant.$

**Figure 4.** Effect of losartan and nifedipine on the frequency of episodes of Raynaud’s phenomenon. The severity score was normalized to 100% at baseline. Bars show the mean and 1 SEM. $P$ values determined by unpaired $t$-test.
circulating levels of soluble E-selectin, vWF, and fibrinogen in either treatment arm. The magnitude of the effects of losartan and nifedipine in down-regulating biochemical variables was analyzed by unpaired t-test and showed significantly greater changes in sICAM-1 ($P = 0.01$) and PINP ($P = 0.01$) with losartan therapy.

The results from noninvasive vascular studies did not reflect changes in clinical and biochemical variables.

**Table 3.** Effect of losartan or nifedipine therapy on biochemical variables

<table>
<thead>
<tr>
<th>Variable, treatment</th>
<th>Week 3 (before tx), mean ± SD</th>
<th>Week 15 (after tx), mean ± SD</th>
<th>Week 3 to week 15, mean</th>
<th>Week 15 mean as % of week 3 mean</th>
<th>$P$, by paired t-test (within drug)</th>
<th>$P$, by unpaired t-test (between drugs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sICAM-1 (ng/ml)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Losartan</td>
<td>321.9 ± 141</td>
<td>285.4 ± 160</td>
<td>36.5</td>
<td>89</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Nifedipine</td>
<td>294.8 ± 276</td>
<td>313.5 ± 310</td>
<td>−18.6</td>
<td>106</td>
<td>0.13</td>
<td>0.01†</td>
</tr>
<tr>
<td>sVCAM-1 (ng/ml)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Losartan</td>
<td>701 ± 334.5</td>
<td>561 ± 314</td>
<td>139</td>
<td>80</td>
<td>0.003†</td>
<td></td>
</tr>
<tr>
<td>Nifedipine</td>
<td>822.3 ± 463</td>
<td>590.6 ± 354</td>
<td>231.6</td>
<td>72‡</td>
<td>0.01†</td>
<td>0.32</td>
</tr>
<tr>
<td>sE-selectin (ng/ml)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Losartan</td>
<td>39.06 ± 17</td>
<td>37.5 ± 16</td>
<td>1.55</td>
<td>96</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>Nifedipine</td>
<td>49.17 ± 29</td>
<td>49.25 ± 26</td>
<td>−0.79</td>
<td>100</td>
<td>0.97</td>
<td>0.59</td>
</tr>
<tr>
<td>Endothelin 1 (fmole/ml)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Losartan</td>
<td>0.516 ± 0.44</td>
<td>0.421 ± 0.37</td>
<td>0.097</td>
<td>82</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>Nifedipine</td>
<td>0.592 ± 0.44</td>
<td>0.573 ± 0.54</td>
<td>0.018</td>
<td>97</td>
<td>0.83</td>
<td>0.65</td>
</tr>
<tr>
<td>vWF (IU/dl)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Losartan</td>
<td>49.8 ± 29.04</td>
<td>52.9 ± 31.54</td>
<td>−3.12</td>
<td>106</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>Nifedipine</td>
<td>56.15 ± 33.7</td>
<td>50.1 ± 24.88</td>
<td>6.05</td>
<td>89</td>
<td>0.19</td>
<td>0.26</td>
</tr>
<tr>
<td>Fibrinogen (µmole/liter)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Losartan</td>
<td>1.44 ± 0.43</td>
<td>1.47 ± 0.52</td>
<td>−0.03</td>
<td>102</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>Nifedipine</td>
<td>1.35 ± 0.4</td>
<td>1.32 ± 0.43</td>
<td>0.03</td>
<td>98</td>
<td>0.76</td>
<td>0.62</td>
</tr>
<tr>
<td>PINP (µg/liter)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Losartan</td>
<td>62.45 ± 52.5</td>
<td>53.9 ± 19.5</td>
<td>8.48</td>
<td>86</td>
<td>0.009†</td>
<td></td>
</tr>
<tr>
<td>Nifedipine</td>
<td>54.2 ± 25.9</td>
<td>54.9 ± 26.9</td>
<td>−0.65</td>
<td>101</td>
<td>0.73</td>
<td>0.01†</td>
</tr>
</tbody>
</table>

* tx = treatment; sICAM-1 = soluble intercellular adhesion molecule 1; sVCAM-1 = soluble vascular cell adhesion molecule 1; sE-selectin = soluble E-selectin; vWF = von Willebrand factor; PINP = procollagen type I N-terminal propeptide.

† $P < 0.01$, by paired t-test.

‡ $\approx 20\%$ reduction from baseline.
Thus, there was no improvement in recovery 10 minutes after cold challenge, as assessed by thermography, nor any increase in LDF flux in any treatment group.

Adverse events. Adverse events were more common among the patients taking nifedipine compared with those taking losartan. Side effects of treatment were reported by 10 of 26 patients (39%) receiving nifedipine and 3 of 26 patients (12%) receiving losartan. This difference was statistically significant ($P < 0.005$).

Well-known side effects such as headache, flushing, nausea, and ankle swelling were reported by patients in the nifedipine group and led to the withdrawal of 4 patients (15%): 3 patients because of severe headaches, which did not dissipate in the first 4 days of treatment, and 1 patient because of persistent ankle swelling. These patients were subsequently treated with alternative vasodilators.

Occasional dizziness was reported by 3 of 26 patients (12%) in the losartan group. One SSc patient experienced pleuritic-type chest pain during the second week of treatment with losartan and was withdrawn from the study. Clinical examination, blood tests, electrocardiogram, and chest radiographs did not show any abnormality other than those related to SSc. Therefore, this adverse event was unlikely to have been caused by the study treatment.

Forty-five of the 52 patients (87%) completed the trial. The number of patients withdrawing from the study was significantly higher in the nifedipine group (6 of 26 patients; 23%) compared with the losartan group (1 of 26 patients; 4%) ($P < 0.02$, by chi-square test). Side effects in 4 patients taking nifedipine (15%) and 1 patient taking losartan (4%) led to premature discontinuation and study withdrawals. Two patients from the nifedipine group (8%) failed to complete the trial because of inefficacy of the treatment. All patients who withdrew discontinued treatment during the first 2 weeks of study without providing completed diaries.

Comparative responses for primary and secondary RP. Baseline variables were compared between patients with primary RP and SSc. These data are summarized in Tables 2 and 4. At baseline, SSc patients reported a greater frequency of RP episodes, which just failed to reach significance ($P < 0.06$). There was no significant difference in episode severity between the disease groups. Analysis of baseline biochemical variables revealed significantly higher concentrations of sVCAM-1 ($P = 0.007$) and PINP ($P = 0.02$) and the same trend for sICAM-1, E-selectin, and to a lesser extent, ET-1. No significant differences were observed for fibrinogen and vWF.

The baseline mean hand temperature was lower in the SSc group (mean ± SD 28.4 ± 0.6°C) than in the primary RP group (29.5 ± 0.8°C), but this difference was not significant. At baseline (week 3), laser Doppler flux before cold challenge was significantly lower in the SSc group (mean ± SD 137 ± 92 AU) than in the primary RP group (269 ± 187 AU) ($P = 0.009$, by unpaired $t$-test).

Subgroup analysis revealed some differences in the clinical and biochemical treatment effects (Tables 2 and 5). Generally, a better symptomatic response was observed at week 15 in patients with primary RP in both treatment arms. In the losartan-treated group, the difference was highly significant, with a decrease of 62% in the mean episode severity ($P = 0.002$) and decrease of 55% in frequency ($P = 0.005$). In the nifedipine group, the 30% decrease in mean episode severity in patients with primary RP was not significant ($P = 0.06$). There was a nonsignificant increase in episode frequency in this group (Table 2). In SSc patients at week 15, there were neither clinically nor statistically significant changes in

<table>
<thead>
<tr>
<th>Variable</th>
<th>PRP</th>
<th>SSc</th>
<th>$P$, by unpaired $t$-test (between diseases)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± SD</td>
<td>Range</td>
<td>Mean ± SD</td>
</tr>
<tr>
<td>sICAM-1 (ng/ml)</td>
<td>268.6 ± 111</td>
<td>147–589</td>
<td>352.6 ± 263</td>
</tr>
<tr>
<td>sVCAM-1 (ng/ml)</td>
<td>600.6 ± 463</td>
<td>251–1,428</td>
<td>891.4 ± 334</td>
</tr>
<tr>
<td>sE-selectin (ng/ml)</td>
<td>37.7 ± 16.5</td>
<td>14.9–80.2</td>
<td>48.3 ± 27</td>
</tr>
<tr>
<td>Endothelin 1 (fmoles/ml)</td>
<td>0.48 ± 0.41</td>
<td>0.14–1.48</td>
<td>0.61 ± 0.46</td>
</tr>
<tr>
<td>vWF (IU/dl)</td>
<td>50.2 ± 36.5</td>
<td>23–162</td>
<td>54.56 ± 26.5</td>
</tr>
<tr>
<td>Fibrinogen (μmoles/liter)</td>
<td>1.48 ± 0.32</td>
<td>0.99–2.63</td>
<td>1.33 ± 0.42</td>
</tr>
<tr>
<td>PINP (μg/liter)</td>
<td>49.5 ± 19.9</td>
<td>26.7–94</td>
<td>66.22 ± 27.8</td>
</tr>
</tbody>
</table>

* See Table 3 for definitions of abbreviations.
† $P \leq 0.01$, by unpaired $t$-test.
episode frequency or severity with nifedipine. In contrast, losartan produced clinically significant reductions in both frequency and severity (45% and 36% decrease, respectively), although these just failed to reach statistical significance (Table 2).

The only statistically significant changes in serum markers in the primary RP and SSc groups, when considered separately (Table 5), was a decrease in sICAM-1 in the primary RP group treated with losartan (P = 0.0005). However, there was a reduction in sVCAM-1 in all subsets, as well as a reduction in PINP in SSc patients treated with losartan (Table 5). No appreciable differences were observed in soluble E-selectin, fibrinogen, vWF, and ET-1 levels.

**DISCUSSION**

The overall improvement in the severity and frequency of episodes of RP in losartan-treated patients is encouraging. Although subjective, self-reporting of RP symptoms appears to be a reliable assessment tool, particularly when combined with longitudinal analysis of the data (20). The results for serum markers provide additional indirect evidence that losartan could benefit the underlying pathologic processes in individuals with secondary RP. A steady state did not appear to have been reached at the end of this study, and it is possible that a greater benefit might be achieved after more prolonged administration. Although most studies evaluating the effect of drugs on RP symptoms are also performed over 6–12-week periods, longer-term trials of losartan may be worthwhile.

An unblinded design is a significant limitation of our study, increasing the likelihood of a placebo response to a perceived novel treatment. However, reported placebo responses for clinical outcome variables in other RP studies is usually <20% from baseline (20), which is substantially less than the 49% reduction in severity and 50% reduction in frequency of RP episodes observed for losartan in the present study. That the clinical benefit of losartan was less evident in scleroderma patients is in keeping with the results from other studies which have suggested that treatment with vasodilators is less effective for secondary RP (8,21,22). This may reflect underlying changes in endothelium with its activation and damage, as well as coexistent inflammatory processes (23,24).

Although several studies have reported the effectiveness of nifedipine (4–7), different individual re-

<table>
<thead>
<tr>
<th>End point, group</th>
<th>Losartan</th>
<th></th>
<th></th>
<th></th>
<th>Nifedipine</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sICAM-1 (ng/ml)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSc</td>
<td>376 ± 159</td>
<td>362 ± 178</td>
<td>96</td>
<td>0.68</td>
<td>331 ± 335</td>
<td>356 ± 396</td>
<td>107</td>
<td>0.21</td>
</tr>
<tr>
<td>PRP</td>
<td>291 ± 132</td>
<td>236 ± 125</td>
<td>81</td>
<td>0.0005†</td>
<td>258 ± 71</td>
<td>242 ± 79</td>
<td>93</td>
<td>0.72</td>
</tr>
<tr>
<td>sVCAM-1 (ng/ml)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSc</td>
<td>837 ± 366</td>
<td>650 ± 384</td>
<td>78‡</td>
<td>0.03</td>
<td>950 ± 555</td>
<td>809 ± 399</td>
<td>85</td>
<td>0.03</td>
</tr>
<tr>
<td>PRP</td>
<td>554 ± 228</td>
<td>465 ± 186</td>
<td>84</td>
<td>0.05</td>
<td>671 ± 357</td>
<td>455 ± 271</td>
<td>67</td>
<td>0.03</td>
</tr>
<tr>
<td>sE-selectin (ng/ml)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSc</td>
<td>44.0 ± 20.0</td>
<td>43.4 ± 17.8</td>
<td>100</td>
<td>0.8</td>
<td>52.8 ± 33</td>
<td>53.5 ± 27</td>
<td>101</td>
<td>0.87</td>
</tr>
<tr>
<td>PRP</td>
<td>38.2 ± 19.5</td>
<td>43.9 ± 31.8</td>
<td>83</td>
<td>0.5</td>
<td>43.7 ± 22</td>
<td>42.9 ± 23</td>
<td>98</td>
<td>0.5</td>
</tr>
<tr>
<td>Endothelin 1 (fmole/ml)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSc</td>
<td>0.60 ± 0.50</td>
<td>0.51 ± 0.47</td>
<td>85</td>
<td>0.2</td>
<td>0.6 ± 0.42</td>
<td>0.56 ± 0.40</td>
<td>93</td>
<td>0.29</td>
</tr>
<tr>
<td>PRP</td>
<td>0.42 ± 0.36</td>
<td>0.32 ± 0.17</td>
<td>76§</td>
<td>0.24</td>
<td>0.57 ± 0.48</td>
<td>0.58 ± 0.73</td>
<td>101</td>
<td>0.96</td>
</tr>
<tr>
<td>vWF (IU/dl)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSc</td>
<td>50.1 ± 6.7</td>
<td>58.8 ± 22.2</td>
<td>117</td>
<td>0.17</td>
<td>49.4 ± 42.3</td>
<td>46.6 ± 39.3</td>
<td>94</td>
<td>0.82</td>
</tr>
<tr>
<td>PRP</td>
<td>59.3 ± 37.8</td>
<td>55.1 ± 27.7</td>
<td>93</td>
<td>0.54</td>
<td>51.4 ± 28.2</td>
<td>42.6 ± 19.2</td>
<td>83</td>
<td>0.15</td>
</tr>
<tr>
<td>Fibrinogen (µmoles/liter)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSc</td>
<td>1.35 ± 0.46</td>
<td>1.56 ± 0.58</td>
<td>115</td>
<td>0.13</td>
<td>1.31 ± 0.49</td>
<td>1.35 ± 0.39</td>
<td>103</td>
<td>0.74</td>
</tr>
<tr>
<td>PRP</td>
<td>1.52 ± 0.38</td>
<td>1.37 ± 0.45</td>
<td>90</td>
<td>0.25</td>
<td>1.42 ± 0.22</td>
<td>1.27 ± 0.50</td>
<td>89</td>
<td>0.39</td>
</tr>
<tr>
<td>PINP (µg/liter)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSc</td>
<td>68.9 ± 27</td>
<td>58.5 ± 20.7</td>
<td>85</td>
<td>0.03</td>
<td>63.3 ± 29.4</td>
<td>64.9 ± 30.2</td>
<td>102</td>
<td>0.55</td>
</tr>
<tr>
<td>PRP</td>
<td>55.4 ± 22.9</td>
<td>49.0 ± 17.6</td>
<td>88</td>
<td>0.17</td>
<td>40.6 ± 10.4</td>
<td>39.8 ± 10.2</td>
<td>98</td>
<td>0.76</td>
</tr>
</tbody>
</table>

*See Table 3 for definitions of abbreviations.
† Mean of individual patients, each expressed as a percentage of his or her own baseline.
‡ P ≤ 0.01, by paired t-test.
§ ≥20% reduction.
sponses to this drug have been observed (3,21). The 18% overall improvement observed in the nifedipine-treated arm of the study is below that reported in some other trials. A poor response to nifedipine in our study might be attributable to the relatively low dose (40 mg/day), which was selected to minimize side effect–related withdrawals from the control group. Maintenance doses employed for treating refractory RP often exceed 60 mg/day (4,6) but generally require a dose-escalating regimen (6,24). Furthermore, nifedipine has been found to be less effective (24) or of no benefit (8,22) in secondary RP, and this was also our finding.

The results for markers of endothelial cell activation, vascular damage, and extracellular matrix turnover are intriguing but must be interpreted cautiously. It is possible that some of these changes occurred by chance, especially since a number of potentially linked variables were measured. Also, although a number of these markers have been found to be increased in patients with SSC or primary RP (19,25,26), their reliability as disease indicators has not yet been unambiguously demonstrated (27,28). A wide range of values for these markers in this study and in others (19,27) makes the interpretation of changes in serum levels difficult. Disease subsets in the SSC group (29–31), baseline disease activity (27,28,31), and reported circadian variation in the serum levels of some markers (32,33) should be also taken into account.

Nevertheless, several studies have suggested that altered adhesion molecule expression and increased fibrinolysis may be involved in the pathogenesis of both SSC and primary RP (19,27–29). Elevated plasma levels of ET-1, a potent vasoconstrictor, have also been reported in patients with SSC (30,31) as well as in patients with RP (26), and may be a predictor of prognosis in SSC (26,28). In our study, we found elevated baseline levels of circulating ICAM-1, VCAM-1, E-selectin, and ET-1 in SSC, which is in keeping with findings of other groups of investigators (26,28). Losartan treatment was associated with a significant decrease in the levels of VCAM-1 and a large, although not significant, reduction in ICAM-1 and ET-1. This effect of losartan has been reported previously (34) in patients with systemic hypertension.

Profibrotic activity for angiotensin II has been widely described. In particular, angiotensin II can induce collagen and fibronectin synthesis (35,36) and stimulate transforming growth factor β gene expression in fibroblasts and endothelial cells (37,38), thereby promoting extracellular matrix deposition. Losartan has been shown to reduce lung fibroblast proliferation and collagen production induced by angiotensin II in vitro (39). The observed reduction in serum levels of PINP among patients in our study, reflecting the down-regulation of collagen synthesis, was statistically significant in the losartan group. Serum markers of collagen turnover may reflect fibrotic disease activity (40,41), raising the possibility that losartan might have additional disease-modifying potential in SSC, although such a conclusion would be premature from the current study.

Previous attempts to derive reliable, objective, noninvasive measures of blood flow for use in drug studies in RP have had limited success, and our data confirm this, despite significant changes in symptoms of vasospasm. In earlier studies, thermography has been used to quantify dynamic response to drug treatment (42), but its effectiveness has not been widely accepted because of the low reproducibility of the results (17,43). LDF has also been used in intervention studies, with variable outcome (17,44).

In conclusion, the results of this pilot study suggest that 12 weeks of treatment with losartan improves symptoms of vasospasm in RP patients and to a lesser extent in SSC patients. The ability of losartan to modify some serum markers of vascular damage and connective tissue turnover is consistent with the hypothesis that it might also favorably modulate some underlying processes in SSC and RP. We believe that further evaluation in larger studies is warranted, including perhaps examination of its effect on organ-based complications such as pulmonary hypertension and pulmonary fibrosis.

ACKNOWLEDGMENT

We thank Christopher Knight for valuable discussions and help in preparing the manuscript.

REFERENCES

6. Gjorup T, Kelbaek H, Hartling OJ, Nielsen SL. Controlled double-blinded trial of the clinical effect of nifedipine in the...


Appendix 3

Statement of Support

I confirm the following contributions from Kevin Howell to the publication below:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Thermography data collection (100%)</td>
</tr>
<tr>
<td></td>
<td>- Preparation of manuscript (5%)</td>
</tr>
</tbody>
</table>

Prof C P DENTON
Consultant Rheumatologist
Royal Free Hospital
London
28th September 2009
Treatment of Raynaud’s phenomenon with the selective serotonin reuptake inhibitor fluoxetine

B. Coleiro, S. E. Marshall¹, C. P. Denton, K. Howell, A. Blann², K. I. Welsh¹ and C. M. Black

Centre for Rheumatology (Royal Free Campus), Royal Free and University College School of Medicine, Rowland Hill Street, London, ¹Oxford Transplant Centre, Churchill Hospital, Oxford and ²Haemostasis, Thrombosis and Vascular Biology Unit, City Hospital, University Department of Medicine, Birmingham, UK

Abstract

Objective. To compare fluoxetine, a selective serotonin reuptake inhibitor, with nifedipine as treatment for primary or secondary Raynaud’s phenomenon.

Methods. Twenty-six patients with primary and 27 patients with secondary Raynaud’s phenomenon were assigned randomly to receive 6 weeks of treatment with fluoxetine (20 mg daily) or nifedipine (40 mg daily). Following a 2-week washout period, each group was crossed over to the other treatment arm. The primary outcome variable was the frequency of attacks of Raynaud’s phenomenon. Self-reported attack severity, thermographic recovery from cold challenge and plasma levels of von Willebrand factor and soluble P-selectin were also measured.

Results. There was a reduction in attack frequency and severity of Raynaud’s phenomenon in patients treated with either fluoxetine or nifedipine, but the effect was statistically significant only in the fluoxetine-treated group ($P = 0.0002$ for attack severity and $P = 0.003$ for attack frequency). Subgroup analysis showed that the greatest response was seen in females and in patients with primary Raynaud’s phenomenon. A significant improvement in the thermographic response to cold challenge was also seen in female patients with primary Raynaud’s phenomenon treated with fluoxetine but not in those treated with nifedipine. There was no significant treatment effect on von Willebrand factor or soluble P-selectin. No significant adverse effects occurred in the fluoxetine-treated group.

Conclusion. This pilot study confirms the tolerability of fluoxetine and suggests that it would be effective as a novel treatment for Raynaud’s phenomenon. Larger and placebo-controlled trials are warranted to assess fluoxetine’s therapeutic potential further in this vasospastic condition.

Key words: Raynaud’s phenomenon, Scleroderma, Fluoxetine, Nifedipine, Serotonin reuptake inhibitor.

Raynaud’s phenomenon is characterized by episodic vasospasm of the extremities precipitated by cold or emotional stress. It was first described by Maurice Raynaud over 100 years ago [1], and may occur as a primary phenomenon or may be secondary to other disorders, such as systemic sclerosis. It is not uncommon, affecting up to 10% of the adult population, and has a female preponderance [2]. The severity of Raynaud’s phenomenon varies from mild infrequent episodes to more severe daily attacks that interfere with everyday activities and may result in fingertip ulceration and even gangrene. Treatment may be offered in these more severe cases, usually in the form of a vasodilator drug. Although a variety of vasodilators are available, none is universally effective and the response to treatment is often idiosyncratic. Moreover, effective vasodilators such as nifedipine may be associated with severe, intolerable side-effects [3]. Hence, different classes of drugs have been assessed for use in the treatment of Raynaud’s phenomenon in order to broaden the therapeutic range available, thus increasing the
chance of finding a drug that is suitable for the patient in terms of both efficacy and tolerability. The most widely used agents are vasodilators, including a number of different calcium-channel antagonists and α-adrenergic blockers, such as prazosin. Most of these drugs have dose-dependent side-effects, such as headache, ankle oedema and postural hypotension. It seems likely that the most effective drugs will be those which directly target key mediators, and a search for novel agents has led to studies of angiotension receptor antagonists [4] and of the potent synthetic antioxidant probucol [5]. Both of these drugs were apparently superior to nifedipine in controlled clinical trials.

The pathogenesis of the altered vascular tone that underlies Raynaud’s phenomenon is incompletely understood and it is possible that several mechanisms are responsible. An increasing body of evidence suggests that serotonin may be involved. Serotonin is a selective vasoconstrictor in vivo: infusion of serotonin into the human brachial artery resulted in the characteristic sequential colour changes of Raynaud’s phenomenon [6, 7]. Also, ketanserin, a serotonin antagonist that acts by blocking serotonin 2 receptors, has been used successfully in the treatment of Raynaud’s phenomenon, improving digital arterial flow at all temperatures and reproducibly relieving cold-induced vasoconstriction [8]. It should be noted that a subsequent placebo-controlled trial was not positive [9]. There have been anecdotal reports suggesting that fluoxetine is beneficial in Raynaud’s phenomenon [10, 11].

Alteration in endothelial function and platelet activation may be responsible for some of the clinical aspects of Raynaud’s phenomenon and scleroderma [12]; increased levels of plasma markers of endothelial function and platelet activation in patients with connective tissue diseases are evidence of this involvement [13]. Indeed, high levels of von Willebrand factor, indicating severe endothelial damage, are a poor prognostic indicator in systemic sclerosis [14]. Furthermore, adenosine nucleotides and serotonin (possibly arising from platelets) stimulate the release of von Willebrand factor from endothelial cells in vitro [15].

We hypothesized that treatment of patients with primary or secondary Raynaud’s phenomenon with a selective serotonin reuptake inhibitor (SSRI) would lead to a reduction in symptoms, and the present study was conducted to assess the therapeutic potential of fluoxetine in a much larger cohort of well-characterized patients with primary or secondary Raynaud’s phenomenon. To assess its possible future use in clinical practice, we compared its effects with those of nifedipine, currently the most widely used vasoactive drug for Raynaud’s phenomenon, and we specifically compared the responses to these two agents in order to investigate our clinical suspicion that individual patients demonstrate significantly different responses to a variety of therapeutic interventions. Subgroup analysis was used to identify particular subgroups of patients who were more likely to derive benefit from this alternative therapeutic agent.

Methods

Study design

This was a prospective, randomized cross-over study conducted over a period of 16 weeks during one winter. The study was approved by the Royal Free Hospital Ethical Practices Committee. Following recruitment and informed consent, patients discontinued any vasodilator drugs and were advised to start keeping a symptom diary of the frequency and severity of their Raynaud’s attacks. After this 2-week washout period, thermography and nailfold capillaroscopy were performed and blood samples taken. Patients were randomized to receive either fluoxetine 20 mg daily or nifedipine LA 40 mg daily for 6 weeks, after which assessments were repeated. After a 2-week washout period, patients crossed over to receive 6 weeks of treatment with the other drug. This was followed by further blood sampling and thermographic assessment.

Patients

Patients were eligible if they were experiencing at least six attacks of Raynaud’s phenomenon per week and were aged between 18 and 75 yr. Significant cardiorespiratory and renal disease or epilepsy or any medical condition contraindicating the use of nifedipine or fluoxetine and the concurrent use of calcium channel blockers or SSRIs for other indications were also exclusion criteria. Patients were enrolled consecutively into the study according to these criteria, and comprised a cohort of individuals (mostly living within Greater London) with severe symptomatic Raynaud’s phenomenon and willing to participate. Fifty-three patients were recruited into the study, and their characteristics are shown in Table 1. Primary Raynaud’s phenomenon was identified by the absence of definite nailfold capillaroscopic abnormalities and negative antinuclear autoantibody reactivity by immunofluorescence on Hep2 substrate using serum diluted 1:100 [16].

Severity and frequency of attacks

Patients were asked to record, on one particular pre-selected day of every week, the number of attacks of Raynaud’s phenomenon occurring that day and to score the average severity of attack using a visual analogue scale on which 0 represented no attacks and 10 the most severe attack ever experienced.

Thermography

Thermography studies were performed before the start of the trial and at the completion of each treatment arm. An infrared thermal imaging camera (Starsight: Insight Vision Systems, Malvern, UK) was used to measure the skin temperature of the hands. All participants were asked to avoid alcohol for 24 h before the study and hot caffeinated drinks and hot meals on the day of the test. During the test, the patients sat comfortably in a temperature-controlled room (23 ± 1°C) for 15 min before the measurements commenced. A baseline thermal image was obtained, after which the hands were
immersed in water at 15°C for 1 min. Gloves were worn for the cold challenge to avoid problems of evaporative cooling, but were removed for rewarming and imaging. Thermal images were recorded immediately after the cold challenge and 10 min later. Rewarming was assessed using the Thermosoft programme (EIC, USA), averaging the temperatures of all fingers at baseline and after recovery.

**Vascular markers**

Venous blood was obtained after non-traumatic venipuncture into 0.11 M sodium citrate. Citrated plasma was withdrawn after centrifugation for 20 min at 1000 g and 4°C and was stored at −70°C until assayed. Von Willebrand factor was measured by an established enzyme-linked immunosorbent assay (ELISA) technique using commercial antisera from Dako (Ely, UK) and reference von Willebrand factor from NIBSC (Potters Bar, UK). Soluble P-selectin was measured by ELISA using commercial reagents (R&D Systems, Abingdon, UK). The intra-assay coefficient of variation (CV) of these ELISAs was <5% and the inter-assay CV <10%.

**Statistical analysis**

Pre- and post-treatment values of clinical variables (severity and frequency of Raynaud’s attacks) and serological tests were analysed by paired Student’s t-test.

**Baseline clinical variables**

Table 2 shows the baseline clinical variables in different subgroups before the start of treatment. Although differences existed between these subgroups, they were not statistically significant and were unlikely to account for differences in the treatment response.

**Results**

**Clinical variables**

Analysis of the symptom diaries showed that both fluoxetine and nifedipine produced a reduction in the severity and frequency of attacks of Raynaud’s phenomenon (Figs 1 and 2 respectively). The reduction in attack severity was statistically significant with fluoxetine (P = 0.0002) but not with nifedipine (P = 0.14). Likewise, it was only fluoxetine that produced a statistically significant reduction in attack frequency (P = 0.003 for fluoxetine compared with P = 0.22 for nifedipine).

Subgroup analysis compared the response to treatment between males and females and between patients with primary and secondary Raynaud’s phenomenon. The results showed that fluoxetine induced a reduction in attack severity and frequency in both males and females, but the effect was statistically significant only in females (P < 0.0002 for attack severity and P = 0.0004 for attack frequency). Nifedipine also induced a reduction in attack severity and frequency in both males and
females but none of these results were statistically significant. Fluoxetine also produced a statistically significant reduction in attack severity both in patients with primary Raynaud’s phenomenon and in patients with secondary Raynaud’s phenomenon ($P = 0.009$ and $0.01$ respectively). Although the former responded slightly better, the difference was not statistically significant ($P = 0.65$ for attack severity and $P = 0.78$ for attack frequency). Fluoxetine-induced reduction in attack frequency was significant only in patients with primary Raynaud’s phenomenon ($P = 0.003$). Nifedipine also resulted in a reduction in attack severity and frequency in both primary and secondary Raynaud’s phenomenon groups, but the reductions did not reach statistical significance.

**Infrared thermography**

More objective evidence of the severity of Raynaud’s phenomenon and the response to treatment was obtained from thermographic assessment of the patients at the start of the trial and after each treatment arm. Baseline thermographic data for this series of patients as a whole and for different subgroups are presented in Table 3 together with the percentage increase in hand temperature after a cold challenge. Baseline hand temperature was similar for males and females, and in primary compared with secondary Raynaud’s phenomenon. The degree of rewarming after a cold challenge was greater in males compared with females but this difference was not statistically significant ($P = 0.16$). Patients with primary Raynaud’s phenomenon also showed a slightly better response to the cold challenge, but again the effect was not statistically significant. Neither fluoxetine nor nifedipine produced any particular change in the baseline hand temperature of these patients ($P = 0.25$ and 0.37 respectively).

The degree of rewarming after cold challenge was assessed by measuring the increase in hand temperature from immediately after the cold water immersion to 10 min later, and this value was expressed as a percentage of the hand temperature difference immediately before and after the cold challenge. Overall, there was a greater extent of rewarming after a cold challenge after treatment with fluoxetine or nifedipine when compared with the pretrial value. Although the temperature rise was greater with fluoxetine than with nifedipine, neither increase was statistically significant ($P = 0.11$ and 0.63 respectively). However, subgroup analysis showed that the extent of rewarming was significantly greater after treatment with fluoxetine in females ($P = 0.05$) but not in males. This corresponds with the significant reduction in the severity and frequency of attacks of Raynaud’s phenomenon, as assessed by the symptom diaries, that occurred in the fluoxetine-treated females but not in the males. Neither sex showed a significant improvement in rewarming after treatment with nifedipine.

Further subgroup analysis showed that patients with primary Raynaud’s phenomenon demonstrated an improvement in rewarming after treatment with either fluoxetine or nifedipine, but this effect was statistically significant only in the fluoxetine-treated patients. Patients with secondary Raynaud’s phenomenon did not show any improvement in rewarming with either treatment. This corresponds with the reduction in both attack severity and frequency recorded in the symptom diaries, which was statistically significant in the fluoxetine-treated group with primary Raynaud’s phenomenon but not in the fluoxetine-treated group with secondary Raynaud’s phenomenon.

**Vascular markers**

Despite changes in symptom or thermographic responses, there was no significant difference in the
level of soluble P-selectin or von Willebrand factor after treatment with fluoxetine. Thus, for P-selectin the mean (SEM) levels were 127.1 (3.1) at baseline and 125 (4.6) after treatment. For von Willebrand factor, the level was 88.2 (4.0) at baseline and rose to 100.2 (8.6) after the fluoxetine treatment period. No significant changes occurred with nifedipine therapy (data not shown).

**Adverse events**

Although both treatment arms were well tolerated, side-effects were commoner with nifedipine, which led to a higher rate of withdrawal from the trial for this drug than for fluoxetine. Table 4 shows the number and percentage of patients who developed side-effects, which were classed as severe when they resulted in withdrawal from the trial, moderate when they necessitated a dose reduction and mild when they were reported but required no dose adjustment, implying that they were tolerable or transient. The commonest side-effects of nifedipine that led to withdrawal from the trial were severe headaches, nausea and palpitations. Other side-effects included facial flushes and swelling of the lower limbs. In the case of fluoxetine, it was apathy, lethargy and impaired concentration that most often led to discontinuation of treatment.

**Discussion**

Our pilot study assessed the clinical efficacy and tolerability of fluoxetine in a larger number of patients with primary and secondary Raynaud's phenomenon than has been described in the literature before, and compared its effect with that of nifedipine, a calcium channel blocker that is a well-established treatment for this disorder [17–20]. The results suggest that fluoxetine is an effective and well-tolerated form of treatment for Raynaud's phenomenon.

The absence of a placebo group is a significant weakness of this study, as substantial placebo effects have been observed previously in trials of treatments for Raynaud's phenomenon. However, the magnitude of the clinical effect in certain subgroups (e.g. a thermographic improvement of 54 and 76% in fluoxetine-treated females and primary Raynaud's patients respectively) suggests that this is more than a placebo response, because the placebo effect has been observed in other Raynaud's trials to be not more than 20%. Another potential limitation of this study is the relatively short duration of the washout period between treatment arms.

It is possible that this was too short for fluoxetine owing to its long half-life, but a longer washout period would have introduced additional problems when comparing the two treatment periods.

Another limitation of the study is that it was open. This introduced potential biases and confounders but was necessary for essentially practical reasons. Objective thermographic and serological assessments were selected as robust end-points to supplement more subjective, though clinically relevant, self-reported symptom diaries, because of the open-label nature of the study.

Although patients were not formally assessed for any underlying depression, the alleviation of which might have accounted for at least part of fluoxetine's clinical effect as assessed by the subjective symptom diary records, we feel that it is unlikely that this was a major factor, as objective evidence of response was obtained from analysis of the thermographic data: significant improvement occurred in the same subgroups (female patients and patients with primary Raynaud's phenomenon) as those showing symptomatic benefit. This remains an important consideration because an improved mental attitude resulting from successfully treated depression might well influence responses in self-reported diaries. Formal psychometric testing would be a valuable addition to any future study protocol.

This trial was conducted over one winter and, although there was a difference in ambient temperature between the beginning and the end of the study period, this potential bias was cancelled by the randomized cross-over design of the trial, which ensured that equal numbers of patients were assigned to fluoxetine and nifedipine at any particular period of the trial.

The calcium channel antagonist nifedipine was chosen in the cross-over arm of this trial because of its well-established role in the treatment of Raynaud's phenomenon, as shown by several studies. It was therefore surprising that in this trial there was no significant response, either symptomatic or thermographic, to nifedipine. This might have been partly due to the lower doses used (40 mg daily), as doses of 60 mg have been described in treating refractory Raynaud's phenomenon.

The relatively small number of patients evaluated in this pilot study means that subgroup analysis must be interpreted cautiously. However, it appears that patients with primary Raynaud's phenomenon responded better to fluoxetine, both symptomatically and thermographically, than the group of patients with an underlying connective tissue disorder. This might have resulted from the more advanced vasculopathy, with an element of irreversible structural damage, in patients with secondary Raynaud's phenomenon making them less amenable to pharmacological therapy.

One of the mechanisms by which fluoxetine may provide relief in Raynaud's phenomenon is by reducing the circulating level of serotonin, which is known to be a selective vasoconstrictor. Although platelets are a rich source of serotonin, they cannot synthesize it [21] but accumulate it throughout their physiological life.

**Table 4. Frequency of adverse effects**

<table>
<thead>
<tr>
<th></th>
<th>Fluoxetine</th>
<th>Nifedipine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe side-effects&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4 (8.2)</td>
<td>9 (17.6)</td>
</tr>
<tr>
<td>Moderate side-effects&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5 (10.2)</td>
<td>7 (13.7)</td>
</tr>
<tr>
<td>Mild side-effects</td>
<td>22 (44.9)</td>
<td>19 (37.3)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Headaches, nausea, palpitations, apathy, lethargy.
<sup>b</sup>Facial flushing, lower limb swelling.
Normal plasma serotonin levels are very low [22] but rise when platelets aggregate [23]. Fluoxetine, which is an SSRI, blocks the uptake of serotonin into platelets [24] and will thus decrease the amount of serotonin that is released during platelet activation/aggregation. Fluoxetine is known to deplete platelet serotonin by 95% [25].

A difficulty encountered in recruiting for this trial—one that might pose a problem in clinical practice—was reluctance on the part of the patients to take an anti-Raynaud’s drug that is widely used as an antidepressant. Despite this, fluoxetine may occupy a useful niche in the treatment of this vasospastic condition because, apart from expanding the choice of drugs available, it has a low incidence of haemodynamic side-effects, which are often associated with the use of other vasodilators, such as the calcium channel blockers.

In conclusion, this pilot study has shown that the SSRI fluoxetine is generally well tolerated and an effective agent in reducing the severity and frequency of attacks of Raynaud’s phenomenon. The response to treatment was variable and the greatest benefit was seen in female patients and patients with primary Raynaud’s phenomenon. Some of the variability in the response to treatment may also have been due to genetic differences in metabolic or signalling pathways related to serotonin, and this possibility may be addressed in future studies. Larger and placebo-controlled trials are now warranted to assess fluoxetine further for its clinical efficacy and tolerability; if the results were favourable an important agent would be added to the therapeutic armamentarium in the often difficult management of Raynaud’s phenomenon.

Acknowledgements

I would like to acknowledge the support given by Dr Carol Leighton, who assisted with randomization. Financial support was provided by the Arthritis Research Campaign (UK) and the Raynaud’s and Scleroderma Association (UK).

References

Appendix 4

Statement of Support

I confirm the following contributions from Kevin Howell to the publication below:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Contribution</th>
</tr>
</thead>
</table>
- Thermography protocol design (100%)  
- Data collection (50%)  
- Preparation of manuscript (25%) |

Dr L CHERKAS

Twin Research and Genetic Epidemiology Unit  
St Thomas' Hospital  
London  
24th August 2009
The use of portable radiometry to assess Raynaud's phenomenon: a practical alternative to thermal imaging

L. F. Cherkas, K. Howell¹, L. Carter, C. M. Black¹ and A. J. MacGregor

Twin Research and Genetic Epidemiology Unit, St Thomas' Hospital, London and Department of Rheumatology, Royal Free Hospital, London, UK

Abstract

Objectives. To compare the performance of a portable radiometer with thermal imaging and to assess the potential for radiometry to provide a practical alternative for assessing vascular responsiveness in Raynaud's phenomenon (RP).

Methods. Subjects comprised 18 patients with diagnosed RP and 19 non-RP subjects. A thermal imager (Starsight) and a portable radiometer (Cyclops) measured digital temperature at baseline and the subsequent drop and rise in temperature following a cold challenge test.

Results. The intra-class correlations between the two instruments for all three measures exceeded 80%. The overall performance of each instrument was almost the same, the Starsight thermal imager correctly classifying 84% of subjects as RP or non-RP and the Cyclops portable radiometer correctly classifying 86% of subjects. The sensitivity of the thermal imager was 83%, compared with 89% for the portable radiometer; the specificity of both instruments was 84%. The positive and negative predictive values of the thermal imager were 83 and 84% respectively, and those for the portable radiometer were 84 and 89%.

Conclusions. The two instruments performed equally well and the differences between them in their absolute measurements did not influence their ability to detect RP. Portable radiometry provides a practical, cheap, accurate and reliable alternative to thermal imaging and has the potential to be used in range of clinical and epidemiological settings.

Key words: Thermography, Radiometry, Raynaud's phenomenon, Classification, Cold challenge test.

The uses of thermography in medicine are wide-ranging [1] and include the detection of breast disease [2], lumbar disc herniation [3] and deep venous thrombosis [4]. In rheumatology, thermography can be used to quantify the degree of synovitis in patients with inflammatory joint disease [5]. Thermography also shows promise as a diagnostic tool for Raynaud's phenomenon (RP), in which disturbances of vasomotor control produce abnormal surface temperature patterns, when performed in conjunction with a cold challenge of the hands or feet [6, 7]. However, the technique has limited practical application outside specialist centres because of the requirement for costly thermal imaging equipment and its lack of portability.

Recent advances in technology have led to the development of inexpensive hand-held radiometers that might offer an alternative to thermography. However, whilst portable radiometers can be used outside the specialized setting of a vascular laboratory, it is unclear if the radiometric technique is an effective substitute for established thermographic methods. In this study we compared the performance of two instruments—a pyroelectric thermal imager and a portable radiometer—in discriminating the cold challenge responses of RP patients from those of normal subjects.

Method

Subjects

Measurements were taken from 18 female subjects classified as having RP and from 19 healthy female controls. The 18 RP subjects included 16 identified from a hospital database of primary RP and two identified from a healthy volunteer population (hospital workers). A nurse trained in the assessment of the disease classified RP following an interview. The interview questions...
related to the subject's history of sensitivity to cold and digital colour changes, including white. All RP subjects satisfied the clinical criteria for definite RP validated by Brennan et al. [8].

**Instruments**

The instruments used by independent trained operators were (i) the Starsight pyroelectric vidicon thermal imager (Insight Vision Systems, Great Malvern, UK) [9], with image capture and analysis software by Thermosoft (EIC, Jenison, MI, USA) and running on a 486-66 PC-compatible computer under Windows 3.11; and (ii) the Cyclops 330S Portable Radiometer (Land Instruments, Dronfield, UK). Both the Cyclops 330S hand-held thermometer and the Starsight thermal imager use the pyroelectric method to detect infrared radiation within a similar spectral range (7-15 µm). The Starsight uses a staring array detector to produce an image of the target, whereas the Cyclops simply records spot temperature. Consequently, in comparison with the Starsight, the Cyclops is a considerably less sophisticated instrument and costs less than one-tenth of the price.

**Procedure**

All subjects underwent a cold challenge test. This test has been used extensively (with only minor differences in protocol) for the thermographic assessment of RP [6, 7, 10, 11]. Hot or caffeinated drinks were avoided on the study day and subjects were clothed lightly. Prior to the cold challenge, there was an initial equilibration phase during which subjects were seated and exposed to an ambient temperature of 23°C for 15 min. This temperature was achieved by the use of a ceiling-mounted air-conditioning unit set to slow airflow. The cold challenge was conducted by immersing the subjects' gloved hands in water at 15°C for 60 s.

Two operators then took temperature measurements independently on the same subject. Subjects were seated and held their hands at chest height, palms forward. The Starsight thermal imager was mounted on a tripod and took an instantaneous image of both hands. The thermograms were analysed with the Thermosoft program by using the freeform shape facility to define regions of interest and measure their temperature. In this study, the operator measured the temperature of all fingertips (excluding thumbs) in an area bordered by the fingertip and the distal finger crease.

The Cyclops was held by an operator who sat close to and opposite the subject and took sequential measurements of all eight fingertips immediately after each thermal image had been taken. The Cyclops operator aimed to measure the temperature at the centre of the whorl visible on the palmar aspect of each fingertip.

Measurements were taken by both operators at baseline (T$_{pre}$) and at three further time points during a 10-min period after cold challenge: T$_{post}$ = immediately after immersion; T$_{6}$ = 5 min after immersion; T$_{10}$ = 10 min after immersion. This allowed computation of baseline (T$_{pre}$), drop (T$_{pre}$ - T$_{post}$) and rise (T$_{10}$ - T$_{post}$). The measurements were taken independently by the two operators, who were blinded to each other's observations.

**Analysis**

The analysis first assessed the level of agreement between the two instruments and, secondly, addressed the question of whether the differences between the two instruments affected their ability to distinguish between RP and non-RP subjects in the clinical setting.

The agreement between the instruments was assessed using the method of Bland and Altman [12]. Their ability to detect RP was assessed by using the data from the Starsight camera to construct a rule to discriminate between RP and non-RP subjects, which was then applied to the Cyclops data. This rule was derived from a logistic regression model that contained all three variables (baseline, drop and rise); patients were classified as having RP if the probability of disease on the basis of their measurements was >0.5. For the purpose of analysis, the average temperature of all eight digits at each time point was taken.

All analyses were carried out using STATA (StataCorp, College Station, TX, USA).

**Results**

**Subjects**

The mean age of the patients with RP was 6 yr older than that of the normal volunteers (Table 1). None of the patients with RP had associated connective disease.

<table>
<thead>
<tr>
<th></th>
<th>RP (n = 18)</th>
<th>Non-RP (n = 19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yr)</td>
<td>Mean 37</td>
<td>Mean 31</td>
</tr>
<tr>
<td>Starsight</td>
<td>Range 17-61</td>
<td>Range 18-53</td>
</tr>
<tr>
<td>Baseline</td>
<td>26.17</td>
<td>30.66</td>
</tr>
<tr>
<td>Drop (°C)</td>
<td>4.89</td>
<td>4.78</td>
</tr>
<tr>
<td>Rise (°C)</td>
<td>0.83</td>
<td>0.26 to 2.71</td>
</tr>
<tr>
<td>Cyclops</td>
<td>23.78</td>
<td>29.91</td>
</tr>
<tr>
<td>Baseline</td>
<td>3.93</td>
<td>23.06 to 33.88</td>
</tr>
<tr>
<td>Drop (°C)</td>
<td>1.55</td>
<td>7.46</td>
</tr>
<tr>
<td>Rise (°C)</td>
<td>0.32 to 3.58</td>
<td>3.95 to 10.46</td>
</tr>
</tbody>
</table>

Table 1. Characteristics of subjects

The use of portable radiometry to assess RP 1385
One was taking vasodilators at the time of the study, but still showed the overall pattern of lower baseline temperature, 'blunted' drop and slower rise following the cold challenge typical of the RP subjects compared with normal subjects (Table 1). All the normal volunteers were healthy.

**Agreement between instruments**

The intra-class correlations, mean difference, standard deviation of the difference between the instruments ($S_{diff}$, Starsight – Cyclops) and the range of values defining the 95% limits of agreement (mean difference ± 2$s_{diff}$) for the two instruments for each of the three key measures are given in Table 2.

All the intraclass correlations exceeded 80%, indicating good agreement between the instruments. There was no consistent bias of either instrument and the limits of agreement were acceptable. The variance of the difference was not affected by the mean. As expected, the limits of agreement were wider for baseline measurements (which were absolute measures and would have been influenced by calibration differences between the two instruments) compared with measurements of drop and rise (which were relative measures and were less influenced by calibration differences).

**Classification of RP**

The performance of Starsight in correctly classifying subjects as having RP or not and how well the Cyclops measures compared are shown in Table 3. The clinical definition of RP was used as the gold standard. The sensitivity of the Starsight thermal imager was 83% compared with 89% for the Cyclops portable radiometer. The specificity of both instruments was 84%. The positive and negative predictive values of the Starsight were 83 and 84% respectively, and those of the Cyclops radiometer were 84 and 89% respectively. The overall proportion of subjects correctly classified by each camera was almost the same; 84% for Starsight (31/37) and 86% for Cyclops (32/37).

**Discussion**

We have investigated the performance of two methods of skin temperature measurement in subjects with RP and in healthy individuals: a thermal imaging system, which has been in common use in clinical vascular assessment, and a portable radiometer, which is better suited to taking measurements in wider clinical settings. Our analysis shows that there was close agreement between the instruments in absolute terms and any differences between the instruments in their absolute measurements did not influence their ability to detect RP in a clinical setting. These findings indicate that a portable radiometer can reliably replace thermography for the assessment of RP in non-specialized settings.

Portable radiometry does have a number of limitations, and this technique has not been used widely in the assessment of RP. Unlike thermography, it does not take an instantaneous measurement of all digits at one point in time, as the digits can only be measured sequentially by the operator. Therefore, part of the variation in temperature between fingers will be due to the unavoidable time lag in taking the measurement. The technique is labour-intensive. Recognition of the pattern of skin temperature changes is also more difficult.

However, our results indicate that the lack of real-time measurement by the Cyclops is of no practical significance. Data sampling rates using the radiometer were sufficiently fast to keep pace with the rewarming process, even in the control subjects. The cold challenge test is sufficiently standardized for the assessment of RP to allow data collection to be restricted to a few key measures and time points that can be readily captured using the radiometer.

It should be stressed, however, that a thermally controlled environment, careful patient equilibration and the rigorous application of the test protocol are essential elements of any successful technique for the measurement of human body temperature [13]. We have demonstrated that these criteria are met when a trained nurse uses a radiometer in an air-conditioned, dedicated room.

This study was motivated by a desire to extend objective measurements of RP outside specialist settings for use in epidemiological surveys. It is therefore of importance to demonstrate similar accuracy in classifying RP subjects. We accept that neither the thermal imager nor the radiometer currently offers the most accurate method for skin surface temperature assessment. Modern focal plane array (FPA) imagers [14] now offer greater temperature sensitivity and image resolution than the Starsight pyroelectric system, although the accuracy of carefully calibrated FPA systems typically remains no better than ±1°C. Neither pyroelectric
vidicon thermography nor radiometry should be considered the tool of choice where high sensitivity to temperature is essential. FPA thermal imagers will continue to dominate the market for medical temperature measurement in specialist units. At such centres, the thermographic work carried out is often multidisciplinary, encompassing such fields as rheumatology, dermatology and neurology for both routine clinical practice and research. Much of this work will necessarily require high imaging performance, and for such applications the versatility of modern imagers justifies the cost.

The temperature changes provoked by cold challenge of the hands in studies of RP are large compared with the temperature changes known to be of diagnostic significance in other medical conditions [15]. Hence cold challenge studies are not particularly reliant on high thermographic imaging performance. Our results suggest that the radiometric technique has utility equal to that of more sensitive techniques designed to detect peripheral vasospasm. We believe that the technique can be transported to any dedicated temperature-controlled environment, and thus may be widely used in a range of clinical and epidemiological settings.

Acknowledgements

We should like to thank the subjects who volunteered for this study, the RP patients from the Royal Free Hospital and the hospital workers who were the healthy subjects. This work was supported by the Arthritis Research Campaign. AJM is an ARC Senior Fellow.

References

Appendix 5

Statement of Support

I confirm the following contributions from Kevin Howell to the publication below:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Contribution</th>
</tr>
</thead>
</table>
| Cherkas LF, Carter L, Spector TD, Howell KJ, Black CM, MacGregor AJ. Use of thermographic criteria to identify Raynaud's phenomenon in a population setting. J Rheumatol 2003; 30:720-722 | • Thermography protocol design (100%)  
  • Preparation of manuscript (10%) |

Dr L CHERKAS

Twin Research and Genetic Epidemiology Unit

St Thomas' Hospital

London

24th August 2009
Use of Thermographic Criteria to Identify Raynaud’s Phenomenon in a Population Setting

LYNN F. CHERKAS, LIISA CARTER, TIM D. SPECTOR, KEVIN J. HOWELL, CAROL M. BLACK, and ALEX J. MacGREGOR

ABSTRACT. Objective. To assess the value of thermographic measurements of digital skin temperature after cold challenge in classifying Raynaud’s phenomenon (RP) in a healthy population.

Methods. One hundred seventy-five patients with RP and 404 controls were subjected to a 15°C, 60 s cold challenge test. All participants were women. Digital temperature measurements were taken at baseline, immediately postimmersion, and 10 min after immersion using a portable radiometer.

Results. Baseline skin temperature was a significant predictor of RP; however, the fall in temperature on immersion and the subsequent rewarming rate provided no additional information.

Conclusion. Baseline skin temperature can help to predict the occurrence of RP in patients drawn from the general population, but has relatively low discriminatory power. The cold challenge test itself is of limited additional value for classification. Although objective temperature measurements show little power overall to discriminate between RP and non-RP patients, detecting low baseline digital temperature may be a useful adjunct to clinical history in classifying the disease.

Key Indexing Terms:
THERMOGRAPHY
COLD CHALLENGE TEST

Problems of disease definition present an obstacle for objective studies of Raynaud’s phenomenon (RP). It is rare to assess patients during an attack; classification is almost always reliant on the patient’s recall and is prone to bias. While physiological measurements (including thermography, laser-Doppler flowmetry, and finger systolic blood pressure, alone and in combination with a provocative test such as cold challenge) have shown promise in providing an objective assessment of RP1-4, all have been assessed in the clinic setting, using patients with either secondary RP or severe and established disease. The use of objective tests has not been examined in a population setting.

We used portable radiometry5,6 to assess the cold challenge test as an objective measure of RP in a sample from the healthy population.

From the Twin Research and Genetic Epidemiology Unit, St. Thomas’ Hospital, and the Department of Rheumatology, Royal Free Hospital, London, England.

Supported by the Arthritis Research Campaign (UK). A.J. MacGregor is an ARC Senior Fellow.

L.F. Cherkas, DPhil, Genetic Analyst; L. Carter, RN, Research Nurse; A.J. MacGregor, MD, ARC Senior Fellow, Twin Research and Genetic Epidemiology Unit, St. Thomas’ Hospital; T.D. Spector, MD, Consultant Rheumatologist; K.J. Howell, MSc, Clinical Scientist; C.M. Black, MD, Professor of Rheumatology, Department of Rheumatology, Royal Free Hospital.

Address reprint requests to Dr. A.J. MacGregor, Twin Research and Genetic Epidemiology Unit, St. Thomas’ Hospital, Lambeth Palace Road, London, SE1 7EH, UK.

Submitted March 7, 2002; revision accepted September 19, 2002.

RAYNAUD’S PHENOMENON CLASSIFICATION

MATERIALS AND METHODS

Study participants. These comprised 175 patients classified with RP and 404 controls identified by a questionnaire survey of 3652 women as part of a twin study investigating genetic influence on RP.

Classification of RP. Participants responded to a series of established screening questions: (a) Are your fingers unusually sensitive to the cold? (b) Do your fingers sometimes show unusual color changes? If yes, do they become white, blue, purple, or red7? RP was classified as present if patients reported a history of 2 or more color changes including white, based on accepted criteria8.

Thermographic assessment. All participants underwent a cold challenge test, which followed a standard protocol9,10. Hot or caffeinated drinks were avoided on the study day. Participants were lightly clothed with arms bare from the shoulder. Prior to the cold challenge they were exposed to an ambient temperature of 23°C for 15 min. After the equilibration phase, a single trained operator took sequential measurements of all 8 fingertips excluding thumbs, aiming to measure the temperature at the center of the whorl visible on the palmar aspect of the fingertips. This established the baseline skin temperature (B).

The participants’ gloved hands were then immersed in a bowl of water at 15°C for 60 s. Immediately after the hands were taken out of the water, the gloves were removed and measurements of the fingertips were taken (Tpost) and again at 10 min postimmersion (T10).

Portable radiometer. Digital temperature measurements were made using a validated10 Cyclops 330S portable radiometer (Land Instruments, Dronfield, UK).

Statistical analysis. The analysis investigated the discriminatory value of 3 variables in classifying RP: (1) baseline temperature (B); (2) fall (F) (B – Tpost); and (3) rewarm (R) (T10 – Tpost). These measurements were derived from the average temperature of all 8 digits at each time point. Logistic regression was used to fit models to the data in which the clinical classification of RP was included as the outcome variable and the temperature measurements (i.e., B, F, and R) as the predictor variables. Age was included as a confounder. The full set of 3 variable, 2 variable, and 1 variable models was examined and their fit compared. All analyses were carried out using Stata10.
RESULTS
Response characteristics of participants with and without RP. The mean age and the age range of the participants in both groups were similar (Table 1). RP patients had significantly lower baseline temperature and showed a significantly slower rewarming rate compared to non-RP participants.

Logistic regression models (Table 2). The set of models incorporating the baseline temperature variable (i.e., B, BF, and BR models) all showed no significant difference in fit compared to the full 3 variable model (BFR). Conversely, the set of models that did not include the baseline variable (B) (i.e., FR, R, and F models) all showed a significantly worse fit than for the full model. The importance of B is seen in the area under the curve (AUC) values, where there is little difference between the AUC of the B model and the AUC of the full model (BFR). Allowing for the possible confounding effect of age did not affect the results.

Subjects with very low baseline temperatures (≤ 24°C) were nearly 3 times more likely to be RP positive than RP negative (likelihood ratio = 2.89) (Table 3). The cutoff resulted in high specificity (96.0%) but low sensitivity (11.4%) as many patients reporting symptoms of RP did not have particularly cold hands.

DISCUSSION
Our results showed that baseline digital temperature can help predict the presence of RP in a population sample, but information derived from the cold challenge procedure is of little additional value. It is most informative at the lower end of the temperature range, where only 4% of controls were found to have baseline hand temperatures below 24°C. All participants with these low finger temperatures had a 3-fold increased likelihood of being classified with RP.

The majority of our patients reporting symptoms of RP did not have particularly cold hands. On its own, therefore, the baseline measure is not a good overall discriminator of RP in population studies. The most valuable contribution of an objective measure of baseline digital temperature for the purposes of classification of RP in a population setting might be to supplement existing clinical criteria.

In this study, we used a clinical definition of RP. When RP has been defined thermographically on the basis of rewarming in the cold challenge, baseline finger temperature alone may be of less predictive value. Our results do not preclude a role for the cold challenge test in other clinical settings, for example in assessing patients with more severe disease or monitoring an individual’s response to treatment. However, even in these circumstances, the precise contribution of serial measurements from a cold challenge test has yet to be determined.

Our findings highlight an individual’s baseline temperature as a stable physiological variable that might provide insight into the etiology of RP. From an epidemiological perspective it would be of interest to compare baseline temperatures in a suitably controlled environment across populations where the prevalence of RP differs. Baseline skin temperature may also prove to be an important phenotype in understanding the genetic basis of RP.

ACKNOWLEDGMENT
We thank the twins who volunteered for this study.

REFERENCES


10. Intercooled Stata for Windows 95, version 5.0. College Station, TX: Stata Corp.; 1997.


Appendix 6

Statement of Support

I confirm the following contributions from Kevin Howell to the publication below:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cherkas LF, Williams FMK, Carter L, Howell K, Black CM, Spector TD, MacGregor AJ.</td>
<td>• Cold challenge protocol design (100%)</td>
</tr>
<tr>
<td>Heritability of Raynaud's phenomenon and vascular responsiveness to cold: a study of adult female twins. Arthritis Rheum 2007;57:524-528</td>
<td></td>
</tr>
</tbody>
</table>

Dr L CHERKAS
Twin Research and Genetic Epidemiology Unit
St Thomas' Hospital
London
24th August 2009
Heritability of Raynaud’s Phenomenon and Vascular Responsiveness to Cold: A Study of Adult Female Twins

L. F. CHERKAS,1 F. M. K. WILLIAMS,1 L. CARTER,1 K. HOWELL,2 C. M. BLACK,2 T. D. SPECTOR,1 AND A. J. MACGREGOR3

Introduction
The episodic vasospasm that characterizes Raynaud’s phenomenon (RP) occurs in up to 15% of the adult population (1). Environmental trigger factors, including exposure to cold and to vibrating tools, make a well-recognized contribution to the clinical presentation. However, the extent to which the environment explains an individual’s underlying susceptibility to the condition is unclear. Climatic variation, for example, does not fully explain the variation in RP prevalence between different populations (1). Indeed, few specific environmental risk factors have been identified despite numerous population-based studies. A contribution from genetic factors has been suggested by the identification of pedigrees with multiple members affected by RP (2) and by reports of an increased risk among first-degree relatives of patients (3). However, these observations may equally be explained by a contribution from the shared family environment.

We report the results of a study examining the occurrence of RP among a sample of female twins enrolled in the national TwinsUK Registry. The twin study design allows the separation of the contribution of genetic factors from those of the shared family environment. In common with other epidemiologic studies of RP, our assessment was based on individual respondents’ recall of their symptom history and the presence or absence of RP was classified using standard criteria (4). In addition to this questionnaire-based assessment, we extended our evaluation by conducting a cold challenge test in a sample of respondents. In the past, cold challenge testing in RP has been confined to small studies of patients and required equipment impractical for use in a population setting. In developing the present study we validated a simple cold challenge protocol using a portable radiometer and demonstrated that it could be deployed reliably in the large-scale assessment of patients with RP (5). These thermographic measurements provide an additional objective assessment of the genetic and environmental contribution to the physiologic basis of RP that is less prone to potential bias than assessment based on recall alone.

Subjects and Methods

Design. The study followed a 2-stage protocol: 1) an initial screening questionnaire was sent to a sample of adult women enrolled in the St Thomas TwinsUK Adult Twin Registry and 2) a subset of respondents, weighted in favor of RP-affected pairs, were invited to undergo clinical assessment by interview and cold challenge test.

Questionnaire sample. The initial sample was drawn from subjects enrolled in the St Thomas UK Register. Further details are available at www.twinsuk.ac.uk. This cohort of monozygotic (MZ) and dizygotic (DZ) twins was assembled through successive media campaigns recruiting healthy twin volunteers who had agreed to take part in medical research. For historic reasons, most of the twins were female. Their zygosity was ascertained by a standard questionnaire and in cases of uncertainty was confirmed by multiplex DNA fingerprinting.

Questionnaires were sent to 3,652 female twins between ages 30 and 60 years comprising 911 MZ and 915 DZ twin pairs. Stratified sampling was used to ensure an MZ-to-DZ ratio of 1:1 and equal proportions of subjects in 10-year age groups. The questionnaire did not indicate the primary objective of the study. RP screening questions were included in a larger set of questions asking about lifestyle and other health issues.

The questions that related to RP were designed to allow classification using standard clinical criteria (4). Twins were asked 1) whether their fingers were unusually sensitive to the cold; 2) whether their fingers sometimes showed...
unusual color changes; 3) whether, if their fingers showed unusual color changes, these colors included white, blue, purple, or red; and 4) whether they experienced associated pain or numbness. A history of ≥2 color changes including white was considered positive for RP. Those that experienced pain and numbness were classified as having severe RP. Those who reported cold sensitivity without color changes, or color changes that did not include white, were classified as being cold sensitive.

Clinical assessment and cold challenge test. A sample of respondents, comprising 288 twin pairs (129 MZ and 159 DZ), were invited to undergo clinical assessment and cold challenge testing. The subset was weighted to balance the proportions of MZs and DZs among 3 groups of twin pairs: those that were 1) concordant for the presence of RP, 2) discordant for RP, and 3) concordant for the absence of RP. Because twin pairs in which 1 member reported RP (groups 1 and 2) would be most informative for analyses, a greater proportion of these pairs were invited for clinical assessment.

A study nurse interviewed twin pairs attending for assessment. The assessment included a reevaluation of their symptoms using the original set of questions. Subjects reporting color changes were asked to confirm these using standard color charts (6). Rheumatic symptoms were recorded, and blood samples were obtained for rheumatoid factor and antinuclear antibody. All attendees underwent a 15°C, 60-second cold challenge test following a standardized protocol (7).

Statistical analysis. Twin studies. An individual’s phenotype is the result of the effects of both genotype and environment. To study the source of individual differences (i.e., the variance) in a phenotype, genetically related subjects are required. MZ twins share the same genetic makeup and DZ twins share on average 50% of their segregating genes. It is assumed that both types of twins have been exposed to the same shared environments, so any greater similarity between MZ twins than DZ twins is due to genetic influences.

Concordance. Casewise concordance for RP was determined. This is the probability of the co-twin of an affected twin reporting RP. Under complete ascertainment, casewise concordance is calculated from the formula 2C/(2C + D), where C is the number of concordant pairs and D is the number of discordant pairs. Greater than expected concordance indicates familial occurrence, whereas an excess of MZ compared with DZ concordance indicates the familial occurrence is mediated by genetic factors (8).

Genetic modeling. Heritability is a measure of the proportion of variation in a trait that is attributable to genetic variation and was estimated through maximum likelihood structural equation modeling (9,10). This approach assumes that the presence of RP is determined by a normally distributed underlying liability that leads to expression of the trait when it exceeds a certain threshold value. For dichotomous traits, the correlation in liability among twins can be estimated from the frequencies of concordant and discordant pairs. Comparison of the observed correlation in liability between MZ and DZ twins allows separation of the variance into shared additive genetic factors (A; a correlation of 1 in MZ twins and 0.5 in DZ twins), dominance genetic factors (D; correlation of 1 in MZ twins and 0.25 in DZ twins), the shared environment (C; correlation of 1 in both MZ and DZ twins), and the unique environment (E; uncorrelated in MZ and DZ twins). By sequentially removing these variance components in a stepwise manner from the full model, the significance and size of their contribution to the variance can be assessed and the deterioration in the fit of each submodel can be tested by hierarchical chi-square tests. This process leads to a model that explains the variance with as few variables as possible (the best-fitting model). Preliminary analysis was carried out in Stata software (11) and modeling was performed using Mx software (12).

Analysis of the cold challenge data included 3 measurements: baseline skin temperature, temperature after cold immersion, and the rewarming rate (the difference in temperature immediately after cold immersion and at 10 minutes). The relative contribution of the variance components A, C, D, and E was estimated using DeFries-Fulker regression (13) weighted to take account of the sampling proportions of the pairs.

Results
Of the 3,652 individuals who received questionnaires, 3,043 individuals (83%) responded. These included 702 MZ pairs (77%) and 727 DZ pairs (79%). Three twins were

<table>
<thead>
<tr>
<th>Table 1. Characteristics of the sample*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MZ</strong></td>
</tr>
<tr>
<td>n = 1,400 (700 pairs)</td>
</tr>
<tr>
<td><strong>Age, median (range) years</strong></td>
</tr>
<tr>
<td><strong>Body mass index, mean (range) kg/m²</strong></td>
</tr>
<tr>
<td><strong>Alcohol consumption, mean (range) units/week</strong></td>
</tr>
<tr>
<td><strong>Ever smoked, %</strong></td>
</tr>
<tr>
<td><strong>Cold sensitive, no. (%)</strong></td>
</tr>
<tr>
<td><strong>Classified as RP positive, no. (%)</strong></td>
</tr>
<tr>
<td><strong>Classified as severe RP, no. (%)</strong></td>
</tr>
</tbody>
</table>

* MZ = monozygotic (identical) twin pairs; DZ = dizygotic (nonidentical) twin pairs; RP = Raynaud’s phenomenon.
identified in the sample as having autoantibodies suggestive of rheumatic disease (rheumatoid factor in 2 subjects, antinuclear antibody in 1). These twins and their co-twins were excluded from subsequent analyses. The characteristics of the final sample of 700 MZ and 726 DZ pairs are shown in Table 1. The 2 zygosity groups were matched for age, reported body mass index, smoking history, and alcohol history. They also reported a similar prevalence of cold sensitivity, RP, and severe RP (Table 1). The majority of subjects classified as having RP reported symptoms suggestive of severe disease.

Cold sensitivity, RP, and severe RP all showed a greater concordance among MZ twins than DZ twins, indicating a genetic contribution (Table 2). Variance components analysis indicated a heritability of 53%, 55%, and 53% for cold sensitivity, RP, and severe RP, respectively. A potential contribution from the shared environment was rejected for all 3 traits.

A total of 288 pairs (129 MZ and 159 DZ) underwent clinical assessment (Table 3). All pairs were selected at random from the respondents and none showed significant differences compared with the original sample (based on age, body mass index, smoking, and alcohol use). When subjects underwent clinical assessment the agreement with original questionnaire responses was fair: kappa statistics for cold sensitivity, primary RP, and severe primary RP were 0.64, 0.46, and 0.47, respectively. When attending for the clinical assessment, twins were no longer blinded to the study’s hypothesis. Subjective data from the clinical assessments, therefore, were not used to alter the original disease status classification.

The results of cold challenge testing demonstrated that MZ and DZ twins had similar mean values of baseline temperature, drop in temperature after immersion, and rate of rewarming (Table 3). For all 3 measures, a greater correlation in response was seen among MZ twins compared with DZ twins. Variance components analysis showed a significant contribution from additive genetic factors of 65% for baseline skin temperature, 35% for temperature drop after immersion, and 24% for rewarming rate.

### Discussion

RP is common and may affect as many as 1 in 4 adults (1). There is a well-documented association with autoimmune rheumatic disease developing before or after the onset of RP; other recognized causes of secondary RP include thoracic outlet syndrome, paraproteinemias, drugs, and chemicals (14). RP is also associated with a range of vascular conditions that have significant morbidity including stroke, migraine, and coronary artery disease. The etiology of RP remains poorly understood, although an underlying vascular defect has been proposed (15). It may be that a functional imbalance in vasoconstrictors and vasodilators observed in RP (14) is widespread, influencing a number of vascular beds.

Increased susceptibility to RP among relatives of affected probands has been demonstrated in population studies, suggesting a role for genetic factors. However, these studies have relied on probands’ recalled account of symptoms among their relatives and may be prone to bias. Only one study has evaluated RP directly among relatives (3). To date, no RP study has distinguished the influence of genetic factors from that of the shared family environment.

By demonstrating a higher concordance in MZ twins compared with DZ twins, our data show conclusively that RP has a genetic basis, with a doubling of risk of RP in first-degree relatives of affected individuals. This justifies the search for disease susceptibility genes. To date only 2

### Table 2. Concordance and heritability of cold sensitivity, RP, and severe RP

<table>
<thead>
<tr>
<th>Trait</th>
<th>MZ pairs (n = 700)</th>
<th>DZ pairs (n = 726)</th>
<th>Best fitting model</th>
<th>h² (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(C(+/+))</td>
<td>(D(+/-))</td>
<td>Cc %</td>
<td>(C(+/+))</td>
</tr>
<tr>
<td>Cold sensitivity</td>
<td>140</td>
<td>197</td>
<td>59</td>
<td>105</td>
</tr>
<tr>
<td>RP</td>
<td>26</td>
<td>90</td>
<td>37</td>
<td>17</td>
</tr>
<tr>
<td>Severe RP</td>
<td>23</td>
<td>86</td>
<td>35</td>
<td>14</td>
</tr>
</tbody>
</table>

*RP = Raynaud’s phenomenon; MZ = monozygotic (identical) twin pairs; DZ = dizygotic (nonidentical) twin pairs; 95% CI = 95% confidence interval; h² = heritability estimate; C(+/+ = concordant RP positive pairs; D(+/-) = discordant RP pairs; Cc = casewise concordance calculated using the formula Cc = 2C/(2C + D); AE represents best fitting model where A = additive genetic and E = unique environment components of variation.*

### Table 3. Results of thermographic testing and heritability in twin subjects attending for clinical assessment (cold challenge)*

<table>
<thead>
<tr>
<th>Trait</th>
<th>MZ (129 pairs)</th>
<th>DZ (159 pairs)</th>
<th>Model</th>
<th>h² (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline, °C</td>
<td>29.4 ± 3.4 (26.3–35.3)</td>
<td>29.6 ± 3.2 (26.4–35.1)</td>
<td>AE</td>
<td>65 (52–79)</td>
</tr>
<tr>
<td>Drop, °C</td>
<td>6.7 ± 2.0 (0–11)</td>
<td>7.0 ± 1.8 (1.2–12.8)</td>
<td>AE</td>
<td>35 (13–57)</td>
</tr>
<tr>
<td>Rewarm, °C</td>
<td>4.9 ± 3.5 (−1–12)</td>
<td>5.2 ± 3.6 (−1.6–12.5)</td>
<td>AE</td>
<td>24 (5–43)</td>
</tr>
</tbody>
</table>

* Values are the mean ± SD (range) unless otherwise indicated. C(+/+ = concordant trait positive pairs; D(+/-) = discordant pairs; C(−/− = concordant trait negative pairs; see Table 2 for additional definitions.*
such studies have been published, providing limited insight. A genome scan of 6 multicase families showed suggestive linkage at 3 chromosome areas (2). The result of a candidate gene study of 4 vasoactive genes was negative, but the study lacked power (n = 95 cases) (16). Both studies used clinical disease definitions based on subject recall. As yet, no genetic study has included physical responsiveness to cold in characterizing the RP phenotype.

In assessing our findings, a number of methodologic issues need to be considered. Our study was based on women and the findings relate more to primary RP than the secondary form of the condition. The twin study design itself is often criticized because of a potential lack of representativeness of twin samples and potential bias arising from unequal sharing of the common environment in MZ twins compared with DZ twins. These potential biases are likely to have minimal impact in this study because subjects enrolled in the TwinsUK Registry are representative of the UK population with respect to the frequency of common traits and diseases, as well as lifestyle factors (17). None of the twins included in the present study lived together and the differences in environmental sharing have been shown to be minimal (18). We have not found evidence of differences in recall between MZ and DZ twins for past events that might bias results.

One particular difficulty in studying RP is the lack of a gold standard in disease definition. The majority (67%) of those initially classified with severe RP retained the classification after use of the color chart at the interview. While we found good agreement between color charts and clinical criteria at the interview (83% agreement), it was striking that RP symptoms were elicited more frequently at the interview than they were on initial questionnaire. We believe this resulted from the unblinding of twins at the clinical assessment and it highlights the susceptibility of survey data to biased recall. In the present study, responses to the initial questionnaire were used to determine RP prevalence and heritability because they were thought to be a more reliable indicator of disease status.

Unravelling the genetic basis of complex diseases such as RP presents well-recognized challenges. Our observation that the cold challenge response is under partial genetic control provides strong corroboration of the evidence from the subjective data of a heritable basis for RP. Although not designed to provoke an attack of RP, the cold challenge responses can be taken to reflect the physiologic process giving rise to symptoms. We have previously demonstrated that baseline skin temperature, fall in temperature, and rewarming rate are highly correlated and all are associated with the report of RP (7). Demonstrating that these variables are heritable provides further rationale for the use of objective measures to investigate the mechanisms underlying the disease.

RP is associated with other vasospastic conditions such as migraine and hypertension, and we have demonstrated in the same twin population that a shared genetically determined mechanism appears to account for some of this association (19). Understanding the molecular genetic basis of RP and the physiologic processes that determine the wide range of normal peripheral vascular responsiveness may provide useful insight into a range of important pathologic conditions.

AUTHOR CONTRIBUTIONS

Drs. Cherkas and MacGregor had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study design. Cherkas, Carter, Howell, Black, Spector, MacGregor.

Acquisition of data. Cherkas, Carter, MacGregor.

Analysis and interpretation of data. Cherkas, Spector, MacGregor.

Manuscript preparation. Cherkas, Williams, Black, Spector, MacGregor.

Statistical analysis. Cherkas, Williams, MacGregor.

REFERENCES


17. Andrew T, Hart DJ, Sneider H, de Lange M, Spector TD,
Appendix 7

Statement of Support

I confirm the following contributions from Kevin Howell to the publication below:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Contribution</th>
</tr>
</thead>
</table>
- Thermography protocol design (100%)
- Data collection (75%)
- Preparation of manuscript (75%) |

Dame Carol M BLACK
Professor of Rheumatology
Royal Free Hospital
London
24th August 2009
Infrared thermography for the assessment of localised scleroderma in children

KJ Howell, G Martini, KJ Murray, RE Smith, CM Black

Centre for Rheumatology, Royal Free and University College Medical School, London

*Dipartimento di Pediatria, Università di Padova, Italia

1Department of Paediatric Rheumatology, Great Ormond Street Hospital for Sick Children, London

2Department of Medical Physics, Royal Free Hospital, London

Summary

Localised scleroderma is a rare connective tissue disease, often of paediatric onset, characterised by the fibrosis of skin and underlying tissue. Infrared thermography can be used for the detection of skin inflammation associated with active localised scleroderma lesions. In this retrospective study, 48 thermograms of 25 lesions were reviewed independently by two rheumatologists and judged “thermography +ve” or “thermography −ve” on the basis of observed temperature differences between normal and involved skin. There was discordance between observers in only 1/48 lesions. In comparing thermographic assessments with clinical descriptions of lesion activity recorded at the time of imaging, we found thermography had a sensitivity for the detection of active localised scleroderma lesions of ~90%, and a specificity of ~80%. We conclude thermography has utility for the serial assessment of localised scleroderma. Results are reproducible between trained observers.

Key words: localised scleroderma, infrared thermography, sensitivity and specificity, children

Infrarot-Thermographie zur Beurteilung der lokализierten Sklerodermie bei Kindern


Schlüsselwörter: lokalisierte Sklerodermie; Infrarot-Thermographie, Sensitivität und Spezifität, Kinder

Introduction

Localised scleroderma is an autoimmune connective tissue disorder characterised by the fibrosis of skin and underlying tissue. The disease is usually of paediatric onset but is rare, with an incidence of approximately 27 per million per year in the United States (1).

Localised scleroderma is further sub-classified according to the distribution and extent of the skin lesions, although the nomenclature is somewhat arbitrary (1,2,3,4). Morphea describes irregular areas of fibrosis across the torso or limbs. These skin patches may be localised to
small areas, or more generalised. Linear scleroderma is characterised by lesions extending linearly across the skin surface. Such lesions commonly affect limbs and are often strikingly unilateral in their distribution over the body. Where localised scleroderma affects the head and face, extending through the scalp, it is termed en coup de sabre morphea.

Of greatest concern to the paediatrician is the significant potential for deformity as the child grows (1,2). Where the lesions cross a joint, contractures may be observed along with synovitis. Growth failure of the limb may occur, with consequent limb-length discrepancies. Atrophic tissue changes are frequently observed at the lesion sites in the form of subcutaneous lipatrophy or loss of muscle bulk. Finally, and most importantly for the thermographer, inflammation is associated with active, extending localised scleroderma lesions. Newly involved areas may appear pink or feel warm to the touch (5).

The aetiology of localised scleroderma is poorly understood, and there is to date no cure (3). It is uncommon for the lesions to remain active into adulthood (1,2), hence treatment focuses on limiting the tissue damage caused in the active phase. The rarity of the disease means there is a paucity of controlled clinical trials on pharmacological treatments. Anti-inflammatory agents such as methylprednisolone, and methotrexate are the drugs most frequently employed for the management of active localised scleroderma. A reliable indicator of lesion activity is, however, required if pharmacological intervention is to be started and stopped at appropriate times (3).

Allen et al (5) were the first to describe the use of infrared thermography for the assessment of localised scleroderma activity. They observed widespread hyperthermic areas across the skin of a 5 year-old boy with generalised morphea. Furthermore, they were able to demonstrate both clinical improvement and cooling of the lesions after 3 months in response to pulsed intravenous methylprednisolone therapy.

Birdi et al (6) performed thermography once on each of 18 childhood linear scleroderma lesions. They asked a single blinded observer to judge their thermograms “positive” if the lesion was greater than 0.5°C warmer than the surrounding tissue or the contralateral limb. All 3 lesions that were still extending were “thermography positive”, along with 3 of the 12 lesions that were recently clinically unchanged. All 3 resolving lesions were judged “thermography negative.”

Whilst such reports suggest thermography may be of use for the assessment of localised scleroderma, they offer no information on the utility of thermal imaging in groups of patients over long periods of treatment time. Nor do they offer any indication of how reproducible the assessment of thermograms is between trained observers.

Infrared thermography has been in clinical use for the assessment of childhood localised scleroderma for patients of the Royal Free and Great Ormond Street hospitals since 1993 (7). All imaging is performed at the Rheumatology Department of the Royal Free Hospital.

Figure 1 shows the inflammation associated with en coup de sabre type morphea affecting the left side of the face of a 5 year-old girl. Figure 2 demonstrates an active localised morphea plaque located at the right side of the anterior chest wall of a 10 year-old boy. In both of these cases the affected area is clearly hyperthermic in comparison to the contralateral site by more than 0.5°C, and the lesion would be considered “thermography positive” by the criteria of Birdi et al (6).

Method

We aimed to demonstrate the specificity and sensitivity of thermography for detecting clinically progressing lesions. Ours was a retrospective study performed using thermograms recorded from Royal Free Hospital patients between 1993 and 2000. All thermograms were reviewed independently by two rheumatologists experienced in the interpretation of infrared thermograms (GM and KM) to facilitate an assessment of the reproducibility of the technique between trained observers. Only cases with adequate serial descriptions of clinical lesion activity were included in the review.

Thermography was performed using the StarSight pyroelectric thermal imager (8) (Insight Vision Systems Ltd., UK). False-colour image production and processing was performed with Thermosoft software (EIC Inc., USA) running on a 486-66 IBM compatible PC. Prior to thermography, all patients sat or stood as appropriate for 15 minutes at a room temperature of 23±1°C with the limbs or body area to be imaged uncovered.
Figure 1
Thermography of *en coup de sabre* type morphoea affecting the left side of the face. Temperature range 27°C - 36°C

Figure 2
Thermography of localised morphoea at the right side of the anterior chest wall. Temperature range 29°C - 38°C

Figure 3a
A “thermography positive” lesion extending along the lateral aspect of the right thigh prior to intravenous methylprednisolone therapy. Temperature range 25°C - 34°C

Figure 3b
The same lesion as in fig. 3a is now “thermography negative” after intravenous methylprednisolone. Temperature range 25°C - 34°C

Figure 4a
A “thermography positive” lesion with significant underlying lipoatrophy sited at the lower back. Temperature range 27°C - 35°C

Figure 4b
The same lesion as in fig. 4a remains “thermography positive” despite being classified “inactive” after intravenous methylprednisolone and oral methotrexate therapy. Temperature range 27°C - 35°C
Seventeen patients were included in the review (6 male, 11 female). Mean age at onset of the disease was 6.9 years (range 1-15 years). Mean age at first thermographic assessment was 11.5 years (range 2-29 years). Only 3 patients had their initial thermographic assessment at an age of 18 years or more. Table 1 outlines the distribution of localised scleroderma subtypes seen across the 17 patients.

Seven patients exhibited more than one clinically suspicious lesion and so 25 lesions were reviewed in total. Nine patients attended for thermography on more than one occasion. Hence 48 thermograms were included in the review. Table 2 details the distribution of number of visits across the 17 patients included in the study. Patients were on a variety of treatment regimes, and medication continued as considered appropriate throughout the period of serial thermographic assessment.

The clinical description of each lesion, recorded at the time of thermographic assessment, was noted from the patients' medical records. All lesions described as "old", "unchanged", or "pale" were classified as "inactive." All lesions described as "new", "extending", "active", or "pink" were classified as "active." Deformities mentioned in the medical records were also noted, including subcutaneous lipoatrophy, reduced muscle bulk, joint contractures and limb growth failure.

Copies of the thermograms recorded at each clinical assessment were next issued to GM and KM along with a description of the lesion position. The rheumatologists were asked to independently assess the thermographic activity in each thermogram, classifying any lesion more than 0.5°C warmer than adjacent tissue or the contralateral site as "thermography positive."

### Results

Table 3a shows the thermographic activity of all 48 lesions as assessed by GM. This observer used thermography to detect clinically active lesions with a sensitivity of 92% and a specificity of 80%.

Table 3b shows the thermographic activity of the same lesions as assessed by KM. This observer achieved a sensitivity of 87% and a specificity of 80%.

Table 4 compares GM and KM in their assessment of thermographic activity for all lesions.
It can be seen that there is discordance between the two observers in only 1 of 48 lesions. All 5 clinically "inactive" lesions judged "thermography positive" by both observers demonstrated subcutaneous lipoatrophy plus at least one other recorded deformity.

Figure 3a (right frame) shows a typical "thermography positive" lesion extending along the lateral aspect of the right thigh of a 5 year-old female. Figure 3b shows thermography of the thighs of the same patient 18 months later. By this time the lesion had responded well to intravenous methylprednisolone and was clinically inactive. The two thighs appear at similar temperatures and the thermogram was judged "thermography negative" by both observers.

Figure 4a shows a "thermography positive" lesion sited at the lower back of an 11 year-old male. This lesion had recently extended and was considered clinically active. Figure 4b shows the same lesion 2 years later. The lesion had not extended further after treatment with intravenous methylprednisolone and oral methotrexate and was now considered clinically inactive. Both observers, however, judged this lesion still "thermography positive." The presence of severe lipoatrophy associated with the lesion may go some way to explaining this "false-positive" thermography result and others like it.

Discussion and Conclusion

Infrared thermography has a sensitivity for the detection of active localised scleroderma lesions of approximately 90%, and a specificity of approximately 80%. Results are reproducible between trained observers. We conclude thermography has utility for the serial assessment of localised scleroderma, particularly in cases where it is not clear from clinical inspection whether the lesion is still extending despite treatment.

Our work further suggests that thermography may be less useful in the study of lesions that have caused severe tissue damage, particularly subcutaneous lipoatrophy. Whilst it is reasonable to assume that such lesions are inflamed initially (9), this would not be an explanation for the hyperthermia observed in these areas once the lesion has ceased to extend i.e. is clinically inactive. Instead we postulate that significant lipoatrophy influences heat transfer through the subcutaneous fat layer. Skin areas with fat loss underlying are probably being heated more effectively than normal from deeper tissues. Conversely, an intact subcutaneous fat layer forms an effective insulating barrier preventing the conduction of metabolic heat to the skin surface. Hyperthermic areas with no underlying lipoatrophy can be considered warm due to inflammatory processes within the dermis itself.

Thermography has great potential in paediatric serial investigations since it is non-ionising, non-invasive, and well tolerated by young patients. The technique is inexpensive to operate and provides an instant result that can assist the physician in making decisions about disease management while the patient is in clinic. Dermatological applications require the best possible thermal imaging performance to facilitate the resolution of small lesions at temperatures often similar to that of healthy tissue. Fortunately the recent developments in thermal imaging technology (10) mean this performance is now available at a cost which is not too prohibitive. Nonetheless, medical thermography is likely to remain the remit of the specialist referral centre for the foreseeable future.

Much work remains to be done in order to gain a full thermographic understanding of localised scleroderma, and development of the technique is continuing at the Royal Free Hospital.

It must be remembered that a thermogram is a 2-dimensional representation of what is in fact a complex 3-dimensional surface. Consistency of viewing angle is therefore critical, but difficult to achieve in some paediatric subjects. Techniques for rendering a thermogram with emissivity-correction for viewing angle (11) may be of particular utility in paediatric dermatological applications.

Although we have shown that thermograms of localised scleroderma can be assessed consistently by independent trained observers, the notion of a truly quantitative assessment of lesion temperature that is less dependent on the observer is nonetheless appealing. The definition of a "region of interest" (ROI), from which thermographic temperature readings are made, would be necessary to produce a truly quantitative measure of lesion inflammation. Specifying a suitable ROI is however difficult in localised scleroderma since the lesion edge is poorly defined. It is also unclear whether the size of an ROI should be revised for follow-up
measurements to account for the rapid growth of paediatric subjects.

Whilst thickening of the dermis in localised scleroderma has been well documented by ultrasonic methods (12,13), there is less data available on changes to deeper structures. The relationship between dermal thickness, subcutaneous fat loss, and the degree of hyperthermia observed with thermography warrants further investigation.

References
7) Howell KJ. Blood flow in skin as investigated by thermography and capillaroscopy. Rheumatology in Europe 1998;27(3):112-115
10) Ring EFJ, Dicks JM. Spatial resolution of new thermal imaging systems. Thermology International 1999;9(1):7-14

Address of the author:
Kevin Howell
Centre for Rheumatology,
Royal Free and University College Medical School, Rowland Hill Street, London NW3 2PF
E-mail: kjhowell@rfhsm.ac.uk

(Manuscript received on 9.9.2000, accepted on 2.10.2000)
Appendix 8

Statement of Support

I confirm the following contributions from Kevin Howell to the publication below:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Thermography protocol design (100%)</td>
</tr>
<tr>
<td></td>
<td>• Data collection (75%)</td>
</tr>
<tr>
<td></td>
<td>• Preparation of manuscript (25%)</td>
</tr>
</tbody>
</table>

Dr G MARTINI
Department of Paediatrics
University of Padova
24th August 2009
Paediatric Rheumatology Series Editor: P. Woo

Juvenile-onset localized scleroderma activity detection by infrared thermography

G. Martini, K. J. Murray, K. J. Howell, J. Harper, D. Atherton, P. Woo, F. Zulian and C. M. Black

Paediatric Rheumatology Unit, Department of Paediatrics, University of Padova, Italy, Paediatric Rheumatology and Dermatology Units, Great Ormond Street Hospital for Children and Academic Unit of Rheumatology and Connective Tissue Diseases, Royal Free Hospital, London, UK

Abstract

Objective. The aim of this study was to define the clinical utility of infrared thermography in disease activity detection in localized scleroderma (LS).

Methods. We retrospectively reviewed 130 thermal images of 40 children with LS and calculated the sensitivity and specificity of thermography, comparing clinical descriptions of the lesions and contemporary thermographs. The reproducibility of thermography was calculated by using the weighted kappa coefficient to determine the level of agreement between two clinicians who reviewed the thermographs independently.

Results. The sensitivity of thermography was 92% and specificity was 68%. Full concordance between the two clinicians was observed in 91% of lesions, with a kappa score of 0.82, implying very high reproducibility of this technique.

Conclusion. Our results demonstrate that thermography is a promising diagnostic tool when associated with clinical examination in discriminating disease activity, as long as it is applied to lesions without severe atrophy of the skin and subcutaneous fat. Further evaluation is needed to determine whether thermography can predict the future progression of lesions.

Key words: Juvenile localized scleroderma, Thermography, Assessment.

Localized scleroderma (LS) or morphoea is a rare disorder in children. In a study of adults and children the incidence of LS was 2.7/100 000 population per yr over a study period of 33 yr [1]. The severity of the disease varies widely from isolated plaques of morphoea to generalized morphoea, and to extensive linear lesions involving the limbs, the trunk and/or the face.

The course of LS is characterized by an initial phase of inflammation followed by progressive fibrosis, which can affect both the skin and the underlying tissues, and ultimately by atrophy. Involvement of deep subcutaneous, muscular and periosteal tissues is particularly important because it interferes with the growth of affected areas. This process potentially leads to irreversible structural deformities, particularly when the lesions affect the face, as in the en coup de sabre form, or the limbs, as in the linear form, resulting in joint contractures and limb length discrepancy [2, 3]. Therefore the aim of therapy is to arrest the disease early in its course in order to prevent the development of cosmetic and functional complications. This objective presupposes reliable and reproducible methods to detect disease activity and evaluate treatment efficacy.

Various treatments have been suggested for LS, such as oral steroids, UV light, γ-interferon, methotrexate, d-penicillamine, intravenous steroids and vitamin D3, but most of the reports are anecdotal or case collections and only two double-blind controlled studies are available [4–7].

Thermography is a non-invasive technique that detects infrared radiation to provide an image of the temperature distribution across the body surface (Fig. 1). The skin temperature, under carefully controlled environmental conditions, is influenced primarily by the state of the skin vasculature or by the conduction to the skin surface of heat generated in deeper tissues. Some of the clinical applications of thermography in rheumatology reported so far are the assessment of inflamed joints [8, 9], the response to cold challenge of...
the hands in Raynaud’s phenomenon [10–12] and the evaluation of skin surface temperature in Paget’s disease [13] and algodystrophy [14, 15].

The aim of this study was to determine if thermography has a potential role in detecting activity of LS lesions. Furthermore, the sensitivity and specificity of this technique were calculated and its reliability was investigated by analysis of the reproducibility between different physicians.

Methods

We retrospectively reviewed all the thermal images of children with juvenile LS managed at London’s three collaborating centres: the Rheumatology Department at the Royal Free Hospital and the Paediatric Rheumatology and Dermatology Units at Great Ormond Street Hospital for Children.

All patients with onset of the disease before the age of 16 yr and who were examined between 1993 and 2000 were included in the study. The diagnosis of morphoea was based on the typical clinical appearance of the skin and soft tissues and associated deformity, confirmed when necessary by histopathology. The patients were separated into different disease subsets on the basis of the clinical appearance of the lesions. Morphoea was diagnosed when a single or few circumscribed sclerotic plaques with hypo- or hyperpigmentation were present on the skin. When there were multiple patches of morphoea (more than three or four) the condition was diagnosed as generalized morphoea. Linear scleroderma was defined when a sclerotic area had a band-like appearance over the limbs, often in an asymmetrical distribution, with subcutaneous fat and muscle bulk loss and often with impaired bone growth. The en coup de sabre subset was diagnosed when linear morphoea affected the face or the scalp, involving underlying subcutaneous tissues, muscle, periosteum and bone. Combinations of the different subtypes were noted.

All the thermographs were performed at the Royal Free Hospital by the same thermographer (KJH) with the same infrared camera (StarSight pyroelectric infrared imager; Insight Vision Systems, Great Malvern, Worcs, UK). Image processing and production were performed with Thermosoft software (EIC, Jenison, MI, USA) running on a 486–66 IBM-compatible PC.

All the patients were scanned in a temperature-controlled room at 23 ± 1°C, 10–15 min after acclimatization, wearing underwear only. Acclimatization is the time required to achieve stability in skin temperature and is considered to average of 15 min [16]. The technician was unaware of the clinical description of the lesions.

The lesions were considered positive to thermography when a substantial area more than 0.5°C warmer than the matching opposite limb or body area site was visible (or 0.5°C warmer than the surrounding skin if bilateral sites were involved or comparison was impossible for technical reasons).

The sensitivity and specificity of thermography in disease activity detection were examined by comparing the thermal imaging result with the clinical assessment derived from the medical records. To be included in the study, the clinical report had to be contemporary with the thermograms and contain extensive clinical description.

To improve the reliability of clinical descriptions, we elected to include only those written by four of the
Table 1. Clinical characteristics of patients

<table>
<thead>
<tr>
<th>Type</th>
<th>n</th>
<th>F:M</th>
<th>Mean age at onset (yr)</th>
<th>Mean age at diagnosis (yr)</th>
<th>No. ANA-positive</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphoea</td>
<td>5</td>
<td>4:1</td>
<td>10</td>
<td>11.3</td>
<td>1/5</td>
<td>MTX (2), d-pen (1), IVS (3), MMF (1)</td>
</tr>
<tr>
<td>Generalized morphoea</td>
<td>2</td>
<td>2:0</td>
<td>7.5</td>
<td>8</td>
<td>1/2</td>
<td>MTX (2), d-pen (2), IVS (2), OS (1), none (1)</td>
</tr>
<tr>
<td>ECDS</td>
<td>8</td>
<td>3.5</td>
<td>3.8</td>
<td>6.6</td>
<td>6/8</td>
<td>MTX (3), IVS (2), OS (2), none (3)</td>
</tr>
<tr>
<td>Linear</td>
<td>11</td>
<td>8:3</td>
<td>5</td>
<td>5.5</td>
<td>9/10</td>
<td>MTX (4), d-pen (3), IVS (5), OS (2), CyA (1), none (2)</td>
</tr>
<tr>
<td>Linear + morphoea</td>
<td>14</td>
<td>9:5</td>
<td>5.4</td>
<td>5.6</td>
<td>6/7</td>
<td>MTX (9), d-pen (5), IVS (13), none (1)</td>
</tr>
</tbody>
</table>

ECDS, en coup de sabre scleroderma; MTX, methotrexate; d-pen, d-penicillamine; OS, oral steroids; IVS, intravenous steroids; MMF, mycophenolate mofetil; CyA, cyclosporin A.

The clinical characteristics of the patients are shown in Table 1.

The most frequent subgroup was linear scleroderma, isolated or associated with one or more plaques of morphoea (11 and 14 patients respectively). Eight patients had en coup de sabre scleroderma, two of them having patches of morphoea over the rest of the body as well; five patients had localized morphoea and two had generalized morphoea. The characteristics of the different subtypes of scleroderma patients are presented in Table 1.

Twenty-three out of 32 patients (72%) tested for antinuclear antibodies (ANA) were positive.

Fifty per cent of the patients had developed more than two deformities because of the disease: the most frequent were subcutaneous fat loss and tissue atrophy (35 and 30 patients respectively) and muscle bulk reduction (19 patients). Twelve patients presented with severely impaired growth of the affected limb and 11 had developed joint contractures.

Lesions and thermograms

Sixty-eight clinically separable and independently assessed lesions were identified. The sites of the lesions were mainly the limbs (33 lesions on lower limbs, 16 on upper limbs). Ten lesions affected the face and scalp and nine were on the trunk.

Four lesions were present on similar aspects of opposite limbs; in these cases the temperature of the affected area was compared with that of the surrounding skin. For all the other lesions the temperature evaluation was performed by comparing the two matching opposite sites.

Among the 227 thermographs examined, 130 were available from the medical notes. After this selection, 40 patients and 130 thermograms were included in the study. Thirty-five patients were Caucasian, two Afro-Caribbean, two Asian and two Asian/Caucasian. The mean age at onset was 5.7 yr (range 0.5–15 yr) and the female to male ratio was 1.8 (26 females, 14 males). The clinical characteristics of the patients are shown in Table 1.

The most frequent subgroup was linear scleroderma, isolated or associated with one or more plaques of morphoea (11 and 14 patients respectively). Eight patients had en coup de sabre scleroderma, two of them having patches of morphoea over the rest of the body as well; five patients had localized morphoea and two had generalized morphoea. The characteristics of the different subtypes of scleroderma patients are presented in Table 1.

Twenty-three out of 32 patients (72%) tested for antinuclear antibodies (ANA) were positive.

Fifty per cent of the patients had developed more than two deformities because of the disease: the most frequent were subcutaneous fat loss and tissue atrophy (35 and 30 patients respectively) and muscle bulk reduction (19 patients). Twelve patients presented with severely impaired growth of the affected limb and 11 had developed joint contractures.

Lesions and thermograms

Sixty-eight clinically separable and independently assessed lesions were identified. The sites of the lesions were mainly the limbs (33 lesions on lower limbs, 16 on upper limbs). Ten lesions affected the face and scalp and nine were on the trunk.

Four lesions were present on similar aspects of opposite limbs; in these cases the temperature of the affected area was compared with that of the surrounding skin. For all the other lesions the temperature evaluation was performed by comparing the two matching opposite sites.

Among the 227 thermographs examined, 130 were included in the study, each of them having a contemporary detailed clinical description of the corresponding lesion derived from the medical notes. Fifty per cent of the lesions had been examined by thermography and recorded more than once, 25% being examined three or more times over the disease course.

Clinical–thermography agreement

The thermographic activity of all lesions assessed by GM and KJM is shown in Table 2. The sensitivity and specificity reached by GM were 92 and 68% respectively.
(positive predictive value 0.65), with full agreement between the clinical report and thermography in 47/51 active lesions and 54/79 inactive lesions.

KJM achieved a sensitivity of 86% and a specificity of 68% (positive predictive value 0.55). This observer found agreement between the clinical description and the thermography result in 44/51 active lesions and 54/79 quiescent lesions.

All the 11 lesions described as new according to the clinical descriptions were also positive on thermography. In 25 out of 79 examinations, lesions described as inactive were positive on thermography for both observers. We noticed that this phenomenon appeared to depend on lesion duration. The average duration of so-called false positive lesions was 58.6 months compared with 45 months in all the others. We observed that lesions that were warm on thermography but described as quiescent were in general older lesions with a severe degree of atrophy and subcutaneous fat and muscle bulk loss.

In order to confirm this observation, we calculated the sensitivity and specificity of thermography in lesions with duration 2 yr. We found that they were 81 and 87.5% respectively (positive predictive value 0.87), with full agreement between the clinical description and the thermography result in 42/50 lesions, 21/24 clinically active and 21/26 inactive.

Furthermore, the site of the lesion influenced the disagreement between the clinical description and the thermography result. Thirty-nine per cent of thermograms performed on lesions affecting the face and the scalp were falsely positive in comparison with 17 and 12% of those performed on limb and trunk lesions respectively.

**Inter-observer agreement**

GM and KJM performed an independent review of 112/130 thermograms: both were unaware of the thermal image report and the clinical description of the corresponding lesion. The result of the comparison is shown in Table 3. There was full concordance between the two observers in 102/112 (91%) lesions examined and the $\kappa$ score was 0.82 (95% confidence interval 0.714–0.926), which implies almost perfect agreement between the two physicians in thermography interpretation.

**Discussion**

Disease activity detection in LS is a fundamental problem, both in the evaluation of the need for treatment and the assessment of its efficacy over time. Unfortunately, there are no objective methods of assessing disease activity and laboratory tests are not helpful for this purpose.

Several skin scoring systems have been proposed and validated for systemic sclerosis based on skin thickening extension and its changes over time [20], but these scores are not applicable to isolated lesions. The serial measurement of the lesions is often unreliable because of the difficulty in defining the exact border of the lesion itself. Similarly, the sensation of warmth to touch, suggesting activity of the lesion, is a subjective finding.

The use of thermography for activity detection was described for the first time by Allen et al. in 1987 [21]. The authors reported a case of a patient with rapidly progressive generalized morphoea and they monitored the response to treatment with pulse intravenous methylprednisolone by clinical assessment, laboratory tests and thermography examination. They showed that clinical improvement after therapy was confirmed by cooling of the lesions on thermography.

Birdi et al. [22] examined once by thermography 18 lesions in 11 children with linear scleroderma. In that study, all three extending lesions were positive and all three resolving lesions were negative on thermography, while three out of 12 clinically unchanged lesions were thermography-positive.

Such reports, suggesting a possible role for thermography in the assessment of morphoea, do not give any information on its use over time, nor do they indicate either its reproducibility between different observers or the factors possibly affecting its reliability.

In our study we compared thermographic assessments with clinical descriptions of lesion activity recorded at the time of imaging, and we found that thermography had a sensitivity of 92% and a specificity of 68%. Specificity was even higher in new and recent lesions (87.5% in lesions examined within 2 yr of onset), with a high positive predictive value (0.87). This observation suggests that thermography is helpful in investigating the blood flow changes occurring in the initial stages of the disease.

---

**Table 3. Clinical observer (GM–KJM) agreement**

<table>
<thead>
<tr>
<th></th>
<th>KJM positive</th>
<th>KJM negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM positive</td>
<td>56</td>
<td>7</td>
<td>63</td>
</tr>
<tr>
<td>GM negative</td>
<td>3</td>
<td>46</td>
<td>49</td>
</tr>
<tr>
<td>Total</td>
<td>59</td>
<td>53</td>
<td>112</td>
</tr>
</tbody>
</table>
The rate of false positivity of thermal images increases in old lesions characterized by skin atrophy, loss of subcutaneous fat and reduction of muscle bulk. Similarly, the thermograms performed on lesions of the face or scalp, where the skin and subcutaneous tissues are thinner and thus more rapidly altered by the atrophic process occurring in the disease, are more frequently positive despite the clinical features of inactivity. An intact subcutaneous fat layer constitutes an insulating barrier against metabolic heat conduction towards the skin surface. Therefore we could suggest the hypothesis that these observations may be explained by an alteration in heat transfer through the atrophic subcutaneous fat layer, resulting in skin hyperthermia detected by the infrared camera. On the other hand, hyperthermia observed over an area without lipoatrophy is likely to be caused by the microcirculatory changes occurring during the inflammatory process taking place within the dermis [23].

In the present study we evaluated the level of agreement between different observers in interpreting thermal images in order to calculate the reproducibility of this technique as a measure of its reliability. Our results show very high reproducibility between observers ($\kappa = 0.82$), strongly suggesting the value and reliability of the technique in assessing LS lesion activity. As with all diagnostic tools, thermographic examination of the area of interest has to be applied in the appropriate clinical context, with the purpose of either confirming or excluding a suspected clinical diagnosis or narrowing the diagnostic possibilities.

Our results demonstrate that thermography is a promising diagnostic modality when associated with clinical examination in discriminating disease activity, as long as it is applied to lesions without a severe degree of skin and subcutaneous fat atrophy. Further investigations are necessary to study the relationship between dermal thickness, subcutaneous fat loss and the degree of hyperthermia observed with thermography.

Thermography is likely to have application in monitoring the response to treatment, as suggested by Allen et al. [21]. The retrospective nature of our study does not allow us to speculate on this issue or to evaluate the possible role of thermography in predicting the disease course. For this purpose, prospective studies are required with serial thermographic examinations in order to evaluate disease progression or arrest, to develop methods to quantify the degree of hyperthermia and to define the lesion extension clearly.

References

Appendix 9

Statement of Support

I confirm the following contributions from Kevin Howell to the publication below:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibel L, Sampaio MC, Visentin MT, Howell KJ, Woo P, Harper</td>
<td>• Study design (20%)</td>
</tr>
<tr>
<td>JI. Evaluation of methotrexate and corticosteroids for the treatment of localized scleroderma (morphoea) in children. Br J Dermatol 2006;55:1013-1020</td>
<td>• Thermography protocol design (100%)</td>
</tr>
<tr>
<td></td>
<td>• Data collection (20%)</td>
</tr>
<tr>
<td></td>
<td>• Preparation of manuscript (25%)</td>
</tr>
</tbody>
</table>

Dr L WEIBEL

University Hospital of Zurich
Switzerland
24th August 2009
Evaluation of methotrexate and corticosteroids for the treatment of localized scleroderma (morphoea) in children

L. Weibel, M.C. Sampaio,* M.T. Visentin,† K.J. Howell,‡ P. Woo§ and J.I. Harper

Department of Paediatric Dermatology and §Paediatric Rheumatology, Great Ormond Street Hospital for Children, London WC1N 3JH, U.K.
*Serviço de Dermatologia, Hospital Célio Pierro, JID Ipaussurama, S/N-Campinas-SP, Brazil
†Department of Paediatric Rheumatology, University of Padova, Padova, Italy
‡Department of Rheumatology, Royal Free Hospital, London, U.K.

Correspondence
Lisa Weibel.
E-mail: weibel@gosh.nhs.uk; lisaweibel@yahoo.de

Accepted for publication
3 May 2006

Key words
Localized scleroderma, methotrexate, methylprednisolone, outcome, prednisolone, treatment

Conflicts of interest
None declared.

This study was presented in part at the British Association of Dermatologists’ 85th Annual Meeting in Glasgow in July 2005.

Summary

Background Localized scleroderma (LS) or morphoea is often considered to be a benign self-limiting condition confined to the skin and subcutaneous tissue. However, the course of the disease is unpredictable and severe functional and cosmetic disability may result. Drug treatment with systemic corticosteroids in combination with methotrexate has been reported to be beneficial in LS, but data in children is limited.

Objectives To evaluate the efficacy and tolerability of systemic corticosteroids in combination with methotrexate in children with LS.

Methods Treatment and outcome of 34 patients with LS were retrospectively analysed. Pulsed intravenous methylprednisolone was given, followed by oral prednisolone on a reducing regimen and maintenance treatment with methotrexate. We assessed treatment outcome clinically and by thermography and monitored adverse events.

Results From the onset of treatment, the disease stopped progressing in 94% of the patients. All patients demonstrated significant clinical improvement within a mean time of 5.7 ± 3.9 months. Mean duration of follow-up over the treatment period and beyond was 2.9 ± 2.0 years. In 16 (47%) patients therapy was discontinued when the disease was considered to be inactive clinically; however, seven (44%) of the 16 developed a relapse, necessitating repeat treatment. At last follow-up (range 0.2–7.0 years), 24 (71%) of the 34 patients had completely inactive disease. Observed adverse events were moderate and transient and no patient had to stop therapy.

Conclusions These data suggest that systemic corticosteroids and methotrexate in combination are beneficial and well tolerated in the treatment of children with LS. Because of the risk of relapse after discontinuing therapy, long-term monitoring is mandatory.

Localized scleroderma (LS) or morphoea is a recognized connective tissue disorder characterized by hardening and thickening of the skin due to an increased density of collagen. The course of LS includes an early inflammatory stage with hyperaemia of the skin, followed by fibrosis, sclerosis and, finally, atrophy.1 LS shows a great variety in its clinical presentation and has been classified into plaque or circumscribed, linear including scleroderma en coup de sabre, generalized, morphoea profunda (deep), pansclerotic and combined forms.1,2 LS is usually considered to be a condition confined to the skin and subcutaneous tissue and of a benign, self-limiting nature.

However, it often affects underlying muscle and bone and, importantly, extracutaneous manifestations of LS can be found in almost one-quarter of affected children.3 At the more severe end of the spectrum, the disease can progress over years and cause significant atrophy, growth retardation, irreversible structural deformities, joint contractures and severe functional, cosmetic and psychological disabilities.

The aim of therapy is to arrest the disease early in its course in order to prevent the development of cosmetic and functional complications. The management and treatment of severe LS is challenging. There is no specific therapy...
available. Drugs are usually directed towards suppressing inflammation and collagen alteration. Numerous treatments, such as penicillamine, antimalarial drugs, retinoids, calcitriol, calcipotriol, imiquimod, ciclosporin, interferon gamma and ultraviolet (UV) A irradiation, have been used for the treatment of LS, with varying degrees of success and often limited effects on linear and deep forms of LS.\textsuperscript{4–8} Corticosteroids and low-dose methotrexate have repeatedly been reported to be beneficial in the treatment of LS and in children methotrexate has been the most frequently used drug within the last 5 years.\textsuperscript{9–12} However, to date the evidence for the efficacy and safety of a systemic treatment with corticosteroids and methotrexate in children with LS is limited. We evaluated a treatment protocol using intravenous methylprednisolone (IVMP) in the acute phase and/or oral prednisolone in combination with long-term methotrexate in a cohort of paediatric patients with LS. The aim of the study was to determine whether this treatment protocol was effective and safe. We furthermore wanted to identify factors influencing treatment response and important monitoring measures for adverse events and to evaluate the role of thermography for monitoring disease progress.

Materials and methods

Patients

We retrospectively reviewed the case notes of children treated for LS at Great Ormond Street Hospital for Children between 1998 and 2005. All patients who received therapy in the form of a combination of systemic corticosteroids and methotrexate were included in the study. The diagnosis of LS was made clinically by a paediatric dermatologist (J.I.H.) and paediatric rheumatologist (P.W.). The subtype of LS, as well as the site and extent of the lesions, were noted with the help of clinical photographs where necessary. We evaluated baseline characteristics regarding demographic, clinical and laboratory features. Extracutaneous manifestations of the disease, including complications and autoimmune conditions, were noted. Laboratory findings at the start of treatment included erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), antinuclear antibodies and smooth-muscle antibodies.

Treatment protocol

Figure 1 provides an outline of the standard treatment protocol that was used in most patients. Induction therapy included two courses of high-dose IVMP, each containing three pulses, given on two consecutive weeks. Oral prednisolone was started after the first course of IVMP, stopped during the second course of IVMP and then continued on a reducing regimen. Maintenance treatment with weekly methotrexate was started 1 week after the second course of IVMP. Within the treatment regimen we noted dose, duration and route of administration of corticosteroids and methotrexate.

Clinical outcome measures

We collected detailed information on treatment response and adverse events over the therapy period and beyond. To date there are no universally recognized criteria for clinical activity. Therefore we defined the disease as clinically ‘active’ if new lesions appeared; previous lesions increased in size; the lesions were erythematous or warm to touch; there was oedema with thickening of the skin and/or pain related to joints or muscles. Lesions described as having a pale or brownish colour, without thermal changes, of static size, with softening and/or atrophy of the skin and no related pain, were defined as clinically ‘inactive’. ‘Arrest of disease progression’ was defined as no extension of the lesions and no further functional or cosmetic impairment. ‘Clinical improvement’ was defined as softening of the skin, fading signs of inflammation and improvement of previous joint impairment. During maintenance treatment the need for repeat administration of IVMP due to re-activation of disease was described as ‘flare of disease’ whereas the re-activation after discontinuing treatment was defined as ‘relapse’.

Thermography

We and others have previously described the method of thermography to detect disease activity in LS.\textsuperscript{13,14} All thermographs included in this study were performed at the Royal Free Hospital by the same thermographer (K.J.H.), using the same infrared camera from 1998 to 2000 (StarSight pyroelectric infrared imager; Insight Vision Systems, Great Malvern, Worcs, U.K.) and from 2001 to 2005 (FLIR SC 500 ‘Thermacam’; Flir Systems, West Malling, U.K.). Patients were assessed prior to treatment and again during follow-up. Lesions were considered ‘active’ on thermography when the affected area was more than 0.5 °C warmer than the matching opposite body area. For this study, a total of 130 thermal images were assessed. To evaluate the inter-observer reproducibility, two observers experienced in thermography independently reviewed the images blinded to both the clinical description.
of activity and to the thermography report. Thermograms that were ‘active’ prior to treatment were reviewed over a follow-up period of at least 2 years.

Adverse events

Adverse events were noted during the entire treatment period. For the administration of IVMP, the children were admitted and monitored closely. During treatment with oral prednisolone, weekly blood pressure and urine analysis tests were performed. Laboratory monitoring during maintenance treatment included full blood count, electrolytes, urea, creatinine and liver function tests (which included serum alanine aminotransferase, ALT) every 4–6 weeks.

Statistical analysis

All data were collected from the patient charts and entered into a computerized spreadsheet. Mean, standard deviation (SD) or percentage were calculated for the overall sample and subgroups. Comparisons were made with the use of Student’s t-test, Fisher’s exact test or the χ² test, as appropriate. Linear correlations were described by the Pearson correlation coefficient (r). The null hypothesis was rejected with a two-sided P-value of < 0.05. All analyses were performed with the use of SPSS 11.0 for Macintosh (Statistical Package for the Social Sciences, SPSS Inc., Chicago, IL, U.S.A.).

Results

Patients

Thirty-four patients with LS were included in the study. The relevant epidemiological, clinical and laboratory features of the patients are shown in Tables 1 and 2. All but three patients (91%) had linear morphea or a combination with the linear subtype. Twenty (59%) children presented with a variety of deformities: tissue atrophy and/or muscle bulk reduction was found in 17 (50%) patients, joint restriction and growth impairment of the affected limb were present in six (18%) and five (15%) children, respectively.

Extracutaneous involvement affected 24 (71%) patients and the most frequent manifestations were articular, muscular and/or bone related, as found in 17 (50%) cases. Three (9%) children presented with vitiligo and two (6%) with ocular involvement. In 19 patients (56%) the disease affected more than 5% of the total body surface area.

Treatment

Twelve (35%) children had previously been treated with different agents, but without clinical improvement: topical corticosteroids, systemic therapy (ciclosporin, penicillamine, prednisolone and methotrexate) and UVA irradiation were used in 10 (29%), three (9%) and one (3%) patient, respectively. Twenty-eight (82%) children with acute progressive disease received the standard protocol as shown in Figure 1. In eight of these children with little sign of inflammation only one course of IVMP was administered. A group of six (18%) patients (nos 1, 6, 7, 10, 12 and 22) with milder disease (mainly limited to the skin and with minimal signs of acute inflammation) was treated with oral corticosteroids and methotrexate only.

In Table 1 the patients are listed in order of their time of follow-up from 0:2 to 7:0 years. The mean duration of follow-up was 2±9 ± 2:0 years. Nineteen (56%) patients (nos 16–34) had a follow-up time of over 2 years.

The mean initial dose of oral prednisolone was 0:6 ± 0:34 mg kg⁻¹ daily. Of the total 34 patients, 28 (82%) had stopped prednisolone after a mean time of 13 ± 5 months.

Maintenance treatment with methotrexate was started at a mean initial dose of 10:0 ± 3:5 mg m⁻² per week. This was increased when there were still signs of ongoing low-grade disease activity in 26 (76%) patients over the treatment period to a mean maximum dose of 12:4 ± 4:3 mg m⁻² per week. Maintenance treatment with weekly methotrexate was started orally in all but two (94%) patients (nos 9 and 25), who received subcutaneous injections due to an extensive and aggressive disease for which increased drug bioavailability was required. Over the total treatment period 14 (41%) patients switched from oral to subcutaneous administration of methotrexate. This was because of gastric intolerance and the intention of increased bioavailability in nine and five patients, respectively.

Clinical outcome

The most relevant clinical outcome measures are listed in Table 3 and shown in Figure 2. No patients dropped out.

From the onset of treatment, the disease stopped progressing in 32 (94%) of the 34 patients. The mean time to achieve a definite clinical improvement was 5±7 ± 3:9 months. There was no correlation found between the time of responding with clinical improvement and the time of disease duration (r = 0:061, P-value 0:7). At the 2-year follow-up, which included 19 patients, 14 (74%) had completely inactive disease, four (21%) some ongoing low-grade activity and one (5%) clinically active disease. At last follow-up (between 0:2 and 7:0 years) for all 34 patients, 24 (71%) had completely inactive disease.

For the four patients with a flare of disease during maintenance treatment all responded well to a repeat administration of IVMP.

In the 16 children in whom treatment with methotrexate was stopped, the disease remained inactive for a mean time of 20 ± 12 months prior to stopping treatment. Seven (44%) of the 16 patients had a relapse that occurred between 5 and 32 months after discontinuing maintenance treatment (mean time to relapse 16 ± 12 months). In comparing those patients who relapsed with those who did not, no significant difference was found in the duration of maintenance treatment, nor
in the duration of inactive disease. The patients who relapsed had a significantly longer follow-up time than those who did not relapse (5.5 ± 1.4 and 2.2 ± 1.5 years, respectively, P-value < 0.001).

The management of relapse included the following measures: repeat administration of IVMP, restart of oral prednisolone and methotrexate and restart of methotrexate only in three, two and two cases, respectively. They all responded

Table 1  Clinical baseline features of 34 patients with morphoea

<table>
<thead>
<tr>
<th>Patient</th>
<th>Type of morphoea</th>
<th>Site of lesions</th>
<th>Age at start of treatment (years)</th>
<th>Autoantibodies</th>
<th>Extracutaneous involvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Linear: trunk, limbs</td>
<td>Right leg</td>
<td>13.9</td>
<td>Negative</td>
<td>Glucose impairment</td>
</tr>
<tr>
<td>2</td>
<td>Linear: trunk, limbs</td>
<td>Right abdomen, leg</td>
<td>8.9</td>
<td>Sm-ab</td>
<td>Muscle bulk reduction</td>
</tr>
<tr>
<td>3</td>
<td>Generalized</td>
<td>Trunk, legs, arms</td>
<td>6.2</td>
<td>Negative</td>
<td>Joint restriction and deformity</td>
</tr>
<tr>
<td>4</td>
<td>ECDS</td>
<td>Right forehead, cheek</td>
<td>15.2</td>
<td>Negative</td>
<td>Vitiligo, facial hemi-atrophy</td>
</tr>
<tr>
<td>5</td>
<td>Linear: trunk, limbs</td>
<td>Right lower leg</td>
<td>10.8</td>
<td>ANA 1 : 320 (s), sm-ab</td>
<td>None</td>
</tr>
<tr>
<td>6</td>
<td>ECDS</td>
<td>Right forehead, eye, scalp</td>
<td>9.1</td>
<td>Sm-ab</td>
<td>None</td>
</tr>
<tr>
<td>7</td>
<td>ECDS</td>
<td>Left forehead</td>
<td>7.6</td>
<td>Negative</td>
<td>None</td>
</tr>
<tr>
<td>8</td>
<td>Linear: trunk, limbs</td>
<td>Left leg</td>
<td>11.8</td>
<td>ANA 1 : 1280 (h)</td>
<td>None</td>
</tr>
<tr>
<td>9</td>
<td>Linear + plaque</td>
<td>Trunk, legs, face, right arm</td>
<td>5.6</td>
<td>Negative</td>
<td>Joint restriction and deformity</td>
</tr>
<tr>
<td>10</td>
<td>Linear: trunk, limbs</td>
<td>Left buttock, leg</td>
<td>14.8</td>
<td>NM</td>
<td>Calcification cuts with discharge</td>
</tr>
<tr>
<td>11</td>
<td>Linear: trunk, limbs</td>
<td>Right trunk, buttock, leg</td>
<td>5.3</td>
<td>Negative</td>
<td>Muscle bulk reduction</td>
</tr>
<tr>
<td>12</td>
<td>ECDS</td>
<td>Right forehead, nose</td>
<td>13.9</td>
<td>Sm-ab</td>
<td>None</td>
</tr>
<tr>
<td>13</td>
<td>Linear: trunk, limbs</td>
<td>Right arm</td>
<td>5.5</td>
<td>Sm-ab</td>
<td>Joint restriction and deformity</td>
</tr>
<tr>
<td>14</td>
<td>Linear: trunk, limbs</td>
<td>Right trunk, arm</td>
<td>13.9</td>
<td>ANA 1 : 80 (h), sm-ab</td>
<td>Joint restriction and deformity, limb length discrepancy, muscle bulk reduction, Bell’s palsy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient</th>
<th>Type of morphoea</th>
<th>Site of lesions</th>
<th>Age at start of treatment (years)</th>
<th>Autoantibodies</th>
<th>Extracutaneous involvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>ECDS</td>
<td>Left forehead, mouth, chin</td>
<td>3.7</td>
<td>Negative</td>
<td>None</td>
</tr>
<tr>
<td>16</td>
<td>Linear: trunk, limbs</td>
<td>Right trunk</td>
<td>12.8</td>
<td>Negative</td>
<td>None</td>
</tr>
<tr>
<td>17</td>
<td>ECDS</td>
<td>Left forehead, scalp</td>
<td>7.0</td>
<td>ANA 1 : 80 (s), sm-ab</td>
<td>Conjunctivitis, astigmatism</td>
</tr>
<tr>
<td>18</td>
<td>Linear: trunk, limbs</td>
<td>Left groin</td>
<td>7.7</td>
<td>ANA 1 : 40</td>
<td>Pyostomatitis vegetans</td>
</tr>
<tr>
<td>19</td>
<td>Linear: trunk, limbs</td>
<td>Left leg, buttock</td>
<td>8.3</td>
<td>NM</td>
<td>Limb length discrepancy, muscle bulk reduction</td>
</tr>
<tr>
<td>20</td>
<td>ECDS</td>
<td>Right forehead, eye, scalp</td>
<td>7.7</td>
<td>ANA 1 : 80 (s)</td>
<td>None</td>
</tr>
<tr>
<td>21</td>
<td>ECDS</td>
<td>Right nose, mouth, tongue, neck</td>
<td>4.1</td>
<td>Negative</td>
<td>Jaw and nose deformity</td>
</tr>
<tr>
<td>22</td>
<td>Profunda</td>
<td>Right lower arm, hand</td>
<td>6.9</td>
<td>Sm-ab</td>
<td>Limb length discrepancy, muscle bulk reduction</td>
</tr>
<tr>
<td>23</td>
<td>Linear: trunk, limbs</td>
<td>Right groin, leg</td>
<td>6.0</td>
<td>Negative</td>
<td>Muscle bulk reduction</td>
</tr>
<tr>
<td>24</td>
<td>Linear: trunk, limbs + ECDS</td>
<td>Right trunk, leg, arm, left forehead</td>
<td>8.4</td>
<td>Negative</td>
<td>Vitiligo, strabism, muscle spasms, muscle bulk reduction</td>
</tr>
<tr>
<td>25</td>
<td>Generalized</td>
<td>Right trunk, face, both legs, arms</td>
<td>7.4</td>
<td>ANA 1 : 5120 (h), anticardiolipin-ab</td>
<td>Weight loss, fatigue, muscle bulk reduction, joint restriction and deformity</td>
</tr>
<tr>
<td>26</td>
<td>ECDS</td>
<td>Left forehead, scalp</td>
<td>10.1</td>
<td>NM</td>
<td>None</td>
</tr>
<tr>
<td>27</td>
<td>Linear: trunk, limbs</td>
<td>Left trunk, back</td>
<td>7.0</td>
<td>ANA 1 : 80 (s), sm-ab</td>
<td>None</td>
</tr>
<tr>
<td>28</td>
<td>Linear: trunk, limbs</td>
<td>Left trunk, buttock, leg</td>
<td>5.4</td>
<td>NM</td>
<td>Muscle bulk reduction</td>
</tr>
<tr>
<td>29</td>
<td>Linear: trunk, limbs</td>
<td>Right leg, groin</td>
<td>2.3</td>
<td>Negative</td>
<td>Limb length discrepancy, muscle bulk reduction</td>
</tr>
<tr>
<td>30</td>
<td>ECDS</td>
<td>Right forehead, nose, chin, scalp</td>
<td>7.4</td>
<td>Sm-ab</td>
<td>Jaw deformity, dental problems</td>
</tr>
<tr>
<td>31</td>
<td>Linear: trunk, limbs</td>
<td>Right trunk, arm</td>
<td>4.6</td>
<td>NM</td>
<td>Lack of breast growth, deformity</td>
</tr>
<tr>
<td>32</td>
<td>ECDS</td>
<td>Right eye, cheek</td>
<td>3.7</td>
<td>Negative</td>
<td>None</td>
</tr>
<tr>
<td>33</td>
<td>Linear: trunk, limbs</td>
<td>Left trunk, leg</td>
<td>5.5</td>
<td>ANA 1 : 640 (s)</td>
<td>Vitiligo, muscle bulk reduction</td>
</tr>
<tr>
<td>34</td>
<td>Linear: trunk, limbs</td>
<td>Left trunk, arm, leg</td>
<td>9.8</td>
<td>ANA 1 : 80 (s)</td>
<td>Joint restriction, limb length discrepancy, muscle bulk reduction</td>
</tr>
</tbody>
</table>

ECDS, scleroderma en coup de sabre; sm, smooth muscle; ab, antibodies; ANA, antinuclear antibodies; h, homogeneous; s, speckled; NM, not measured.

© 2006 The Authors
Journal Compilation © 2006 British Association of Dermatologists • British Journal of Dermatology 2006
well to re-treatment apart from one patient, who is awaiting follow-up.

We reviewed the patients who had a flare of disease during maintenance treatment and/or a relapse after stopping treatment (n = 9), comparing them with those who did not have a flare or relapse (n = 25). Those with a flare and/or relapse were younger at disease onset (3.3 ± 1.5 years vs. 5.3 ± 3.7 years, P-value 0.029) and had a longer follow-up (5.0 ± 1.8 years vs. 2.1 ± 1.5 years, P-value 0.001). No other differences were found in this comparison.

We also compared the group of patients who were treated with oral corticosteroids and methotrexate only (n = 6) with those who additionally received IVMP at the initiation of treat-

ment (n = 28). There was no difference found in time of clinical improvement (5.0 ± 3.9 and 5.7 ± 3.9 months, respectively, P-value 0.710) or prevalence of relapse (zero and seven patients, respectively, P-value 0.306).

**Thermography**

All but one patient had thermography performed prior to the start of treatment (Table 4, Fig. 3). In nine (35%) patients thermography became inactive during follow-up (Fig. 4). Of the 13 patients in whom follow-up thermography remained ‘active’, 10 (77%) had clinically inactive lesions at the same time of follow-up. For this reason, thermography was considered to be ‘false positive’ in these cases. All these 10 patients had linear scleroderma (morphoea) and five of them had en coup de sabre. Seven of the 10 patients had stopped treatment but three of these seven (43%) subsequently developed a relapse of their condition. The risk of relapse was therefore not greater in the ‘false positive’ group compared with all the other treated patients.

**Table 2** Baseline characteristics of children with morphoea

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>n = 34</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female-to-male ratio</td>
<td>2:1:1</td>
</tr>
<tr>
<td>Age at disease onset (years), mean (SD)</td>
<td>4.8 (3.4)</td>
</tr>
<tr>
<td>Disease duration at diagnosis (years), mean (SD)</td>
<td>2.3 (1.8)</td>
</tr>
<tr>
<td>Disease duration at start of treatment (years), mean (SD)</td>
<td>3.4 (2.4)</td>
</tr>
<tr>
<td>Age at start of treatment (years), mean (SD)</td>
<td>8.2 (3.5)</td>
</tr>
<tr>
<td>Morphoea subtype, n (%)</td>
<td>18 (53)</td>
</tr>
<tr>
<td>Linear: trunk, limbs</td>
<td>11 (32)</td>
</tr>
<tr>
<td>ECDS</td>
<td>2 (6)</td>
</tr>
<tr>
<td>Generalized</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Deep</td>
<td>2 (6)</td>
</tr>
<tr>
<td>Combined</td>
<td>24 (71)</td>
</tr>
<tr>
<td>Exacutaneous involvement</td>
<td>1/33 (3)</td>
</tr>
<tr>
<td>C-reactive protein &gt; 7 mg dl⁻¹</td>
<td>5/31 (16)</td>
</tr>
<tr>
<td>Erythrocyte sedimentation rate ≥ 20 mm h⁻¹</td>
<td>10/29 (34)</td>
</tr>
<tr>
<td>Autoantibodies</td>
<td>10/29 (34)</td>
</tr>
<tr>
<td>Antinuclear antibodies</td>
<td>10/29 (34)</td>
</tr>
<tr>
<td>Smooth-muscle antibodies</td>
<td>10/29 (34)</td>
</tr>
<tr>
<td>ECDS, scleroderma en coup de sabre.</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3** Clinical outcome of paediatric patients with morphoea

<table>
<thead>
<tr>
<th>Outcome measures</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrest of disease progression from the onset of treatment</td>
<td>32/34 (94)</td>
</tr>
<tr>
<td>Two-year follow-up</td>
<td></td>
</tr>
<tr>
<td>No detectable clinical activity</td>
<td>14/19 (74)</td>
</tr>
<tr>
<td>Number of patients off treatment</td>
<td>1/19 (5)</td>
</tr>
<tr>
<td>Last follow-up (range 0-2-7/0 years)</td>
<td></td>
</tr>
<tr>
<td>No detectable clinical activity</td>
<td>24/34 (71)</td>
</tr>
<tr>
<td>Number of patients off treatment</td>
<td>10/34 (29)</td>
</tr>
<tr>
<td>Flare of disease during MT</td>
<td>4/34 (12)</td>
</tr>
<tr>
<td>Stop of MT at inactive disease</td>
<td>16/34 (47)</td>
</tr>
<tr>
<td>Duration of MT (months), mean (SD)</td>
<td>32 (12)</td>
</tr>
<tr>
<td>Relapse after stop of MT</td>
<td>7/16 (44)</td>
</tr>
</tbody>
</table>

MT, maintenance treatment with methotrexate. Except where noted. *With the need for repeat intravenous methylprednisolone.

Fig 2 This curve demonstrates the clinical improvement (defined as softening of the skin, fading signs of inflammation and improvement of previous joint impairment and/or pain) of the 34 patients over time after starting treatment with systemic corticosteroids and methotrexate.

**Table 4** Thermography

<table>
<thead>
<tr>
<th>Details</th>
<th>n = 34 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermography prior to treatment*</td>
<td></td>
</tr>
<tr>
<td>Active</td>
<td>26/33 (79)</td>
</tr>
<tr>
<td>Inactive</td>
<td>6/33 (18)</td>
</tr>
<tr>
<td>No agreement between observers</td>
<td>1/33 (3)</td>
</tr>
<tr>
<td>Follow-up of active lesions on initial thermogram*</td>
<td></td>
</tr>
<tr>
<td>Became inactive</td>
<td>9/16 (55)</td>
</tr>
<tr>
<td>No change</td>
<td>13/16 (50)</td>
</tr>
<tr>
<td>No agreement between observers</td>
<td>4/26 (15)</td>
</tr>
</tbody>
</table>

*All thermal images of at least 2 years of follow-up were reviewed independently by two observers experienced in thermography and blinded to both the clinical description of activity and to the thermography report.
Adverse events

Adverse events over the treatment period are shown in Table 5. They were all mild and transient and no patients dropped out of treatment due to adverse events. Because of gastric intolerance nine (26%) patients switched from oral to subcutaneous administration of methotrexate with good improvement. In one case (no. 16) the infusion of IVMP was stopped because of raised blood pressure. Methotrexate was stopped for 3 weeks and transiently reduced in one and two patients, respectively, because of a raised ALT. Methotrexate was stopped for a week in one case, because of lymphopenia.

**Table 5** Adverse events over treatment period

<table>
<thead>
<tr>
<th>Adverse events over treatment period</th>
<th>n = 34 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>26 (76)</td>
</tr>
<tr>
<td>Nausea during MT</td>
<td>14 (41)</td>
</tr>
<tr>
<td>Elevated liver enzymes during MT</td>
<td>6 (18)</td>
</tr>
<tr>
<td>Glucosuria during IVMP and/or oral prednisolone</td>
<td>5 (15)</td>
</tr>
<tr>
<td>Cushingoid habitus after IVMP and during oral prednisolone</td>
<td>5 (15)</td>
</tr>
<tr>
<td>Lymphopenia &lt; 1500/μL during MT</td>
<td>4 (12)</td>
</tr>
<tr>
<td>Abdominal discomfort during oral prednisolone and MT</td>
<td>4 (12)</td>
</tr>
<tr>
<td>Mild systemic hypertension during IVMP</td>
<td>3 (9)</td>
</tr>
<tr>
<td>Hyperglycaemia during IVMP</td>
<td>3 (9)</td>
</tr>
<tr>
<td>Headache during MT</td>
<td>3 (9)</td>
</tr>
<tr>
<td>Mouth ulcers during MT</td>
<td>3 (9)</td>
</tr>
<tr>
<td>Neutropenia &lt; 1500/μL during MT</td>
<td>2 (6)</td>
</tr>
</tbody>
</table>

MT, maintenance treatment with methotrexate; IVMP, intravenous methylprednisolone. Serum alanine aminotransferase above age-adjusted cut-off levels.

Discussion

This retrospective study is the largest published series of children with LS treated with corticosteroids and methotrexate. Clinical assessment revealed that all patients improved markedly with this treatment and that no patient had to discontinue therapy due to adverse events.

LS may be limited to the skin and subcutaneous tissue only and show a self-limiting course without significant sequelae. However, the course and prognosis of LS is unpredictable and depends on the variant of the disease. The linear subtype of LS is the most common form in children, and together with the deep variant, it can be responsible for severe morbidity as described in a recent multinational study including 750 children with LS. The clinical activity of LS generally persists for 3–6 years, but re-activation may occur. For the treatment of localized plaque lesions topical steroids, calcipotriol ointment, imiquimod or UVA irradiation may be appropriate. All other forms of LS must be considered as potentially severe.

In adult patients beneficial effects of oral corticosteroids were reported in a follow-up study that included 17 patients with severe LS. However, after a mean treatment duration of 18 months, six (35%) experienced a relapse after discontinuing therapy. Within the last decade, methotrexate has gained...
attention as a new approach for sclerotic skin diseases. In systemic sclerosis, methotrexate was shown to be effective in a double-blind, placebo-controlled study of 29 adult patients.\textsuperscript{18} Beneficial effects of low-dose methotrexate (15 mg weekly) were demonstrated in nine adult patients with widespread morphoea during a 24-week trial.\textsuperscript{11} The mechanism through which low-dose methotrexate improves skin fibrosis remains poorly understood. It may act directly on the skin fibroblasts or the skin improvement might be due to its anti-inflammatory effect.\textsuperscript{19} Again, in adults, Kreuter et al.\textsuperscript{12} showed pulsed high-dose IVMP (1000 mg daily on three consecutive days monthly for at least 6 months) combined with low-dose methotrexate (15 mg weekly) to be beneficial and safe in 15 patients with LS.

In children, there are only limited data to validate treatment with systemic corticosteroids and methotrexate for LS. Uziel et al.\textsuperscript{10} described beneficial effects of combined methotrexate (0.3–0.6 mg kg\textsuperscript{-1} weekly) and pulsed IVMP (30 mg kg\textsuperscript{-1}, daily on three consecutive days monthly for 3 months) in 10 children with LS. To our knowledge, no other reports have previously evaluated this treatment regimen in children.

The baseline characteristics of the patients included in our study were similar overall to those reported by Uziel et al.\textsuperscript{10} The high prevalence of extracutaneous manifestations in our group is likely to be related to the selection of patients having acute progressive disease and the high number with the linear subtype of LS.\textsuperscript{3}

Unlike in the previous studies, we used a slightly different treatment protocol by giving a course of pulsed IVMP on three consecutive days, repeated after 1 week, as an induction therapy, followed by oral corticosteroids on a reducing regimen.\textsuperscript{10,12} However, the treatment response with an immediate lack of disease progression and definite clinical improvement corresponds well with the outcomes previously reported.\textsuperscript{10,12} These findings support the role of corticosteroids as effective ‘inducing agents’ for rapidly reducing the early inflammatory phase of the disease. Like Kreuter et al.,\textsuperscript{12} we did not find a correlation between duration of disease and time of response to treatment.

The group of patients who were treated only with oral corticosteroids and methotrexate did not show a difference in time of response or prevalence of relapse. This result is limited by selection bias, as these patients were considered to have a milder disease.

In the patients who discontinued therapy, a relatively high relapse rate of 44% was observed. It was felt that in all these patients maintenance treatment was discontinued only after an adequate time of therapy and duration of inactive disease. However, re-activation occurred in almost half within 4–28 months after stopping treatment. The only risk factor we were able to identify for flare of disease during maintenance treatment and/or relapse, was young age at disease onset. From our data we would advocate that maintenance treatment should be continued for at least 2 years and that after stopping treatment these patients are regularly monitored for at least 5 years.

In the present study the overall tolerability of the treatment protocol was high and no patient had to discontinue therapy due to adverse events. The mild and reversible adverse effects observed in our study correspond to the frequently occurring adverse reactions of low-dose methotrexate and/or high-dose corticosteroids that have previously been reported.\textsuperscript{10,12} However, prior to start of treatment patients need to be carefully assessed for associated medical problems of a renal, cardiac or endocrine nature. We suggest that monitoring during the treatment period should include blood tests every 4–6 weeks (full blood count, electrolytes, urea, creatinine and liver function tests, in particular ALT), weight and height at clinic visits, and while on corticosteroids, blood pressure and urinalysis.

Evaluation of activity of skin lesions in patients with LS has proved challenging to clinicians and investigators. Thermography has been validated as an assessment tool in patients with LS to detect disease activity with a sensitivity of 92% and specificity of 68%.\textsuperscript{13,14} In our study, thermography assessment prior to treatment revealed a sensitivity of 78.8% with full agreement between the two observers. It proved to be a helpful tool in demonstrating and measuring disease activity prior to treatment and in guiding the clinician in the decision about whom, when and how to treat. However, looking at thermography during the follow-up of treatment, thermography was considered to be ‘false positive’ in 10 patients, i.e. clinically inactive but thermographically positive. As described earlier by Martini et al.,\textsuperscript{14} this is probably due to skin atrophy, loss of subcutaneous fat and reduction in muscle bulk, characteristic of old LS lesions. The fact that the two observers experienced in thermography did not reach an agreement in 15% of the evaluations of follow-up thermograms underlines the difficulties in interpreting thermography during follow-up. Despite standardized conditions the patients’ general temperature levels often varied between follow-up visits causing difficulties in comparing thermograms over time. However, with the lack of other validated methods to detect disease activity in LS, we feel that thermography remains a useful, noninvasive tool that we would recommend for baseline assessment. We have further work in progress to evaluate the application of thermography in monitoring response to treatment and to investigate other techniques to detect disease activity in LS in a prospective setting.

This study is subject to a number of important limitations: study design, which is retrospective and not double-blinded and placebo-controlled; the lack of correlative laboratory disease activity markers; and the predominant use of clinical judgement as the marker of response. Skin scoring systems have been validated for the assessment of widespread skin thickening in patients with systemic sclerosis.\textsuperscript{11,20} However, we believe their use for monitoring isolated lesions of LS in children is limited. Measuring the size of skin lesions over time is considered to be inaccurate because it is often difficult to define the exact borders of the lesions and in a growing child an increase in lesion size can mean normal growth.

In conclusion, we believe that children with LS should be identified early, evaluated appropriately, treated promptly and
monitored carefully. Despite the limitations, our data suggest that a combination of systemic corticosteroids and methotrexate is beneficial and well tolerated as treatment for LS in children.

Acknowledgments

Dr Lisa Weibel was financially supported by a nonrestricted grant from the Novartis Foundation and the Theodor und Ida Herzog-Egli-Stiftung, Switzerland. We would like to thank Jane Stevens, Clinical Nurse Specialist at Great Ormond Street Hospital for her help in coordinating the management of our patients.

References

Appendix 10

I confirm the following contributions from Kevin Howell to the publication below:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Contribution</th>
</tr>
</thead>
</table>
• Data collection (50%)  
• Analysis of results (70%)  
• Preparation of manuscript (90%) |

**Patricia Woo CBE**

Professor of Paediatric Rheumatology, UCL,
Consultant in Paediatric Rheumatology
Great Ormond Street Hospital
London
Validation of a protocol for the assessment of skin temperature and blood flow in childhood localised scleroderma

Kevin J. Howell1, Antonia Lavorato2, María Teresa Visentin3, Roy E. Smith2, Gerald Schaefer4, Carl D. Jones5, Lisa Weibel6, Christopher P. Denton1, John I. Harper6 and Patricia Woo7

1Department of Rheumatology, Royal Free Hospital, London NW3 2QG, UK, 2Department of Medical Electronics, Royal Free Hospital, London NW3 2QG, UK, 3Department of Paediatrics, Vicenza City Hospital, Vicenza, Italy, 4School of Engineering and Applied Science, Aston University, Birmingham B4 7ET, UK, 5Medical Imaging Group, Department of Mathematical and Computing Sciences, Faculty of Advanced Technology, University of Glamorgan, Pontypridd CF37 1DL, UK, 6Department of Paediatric Dermatology, Great Ormond Street Hospital for Children, London WC1N 3JH, UK and 7Department of Paediatric Rheumatology, Great Ormond Street Hospital for Children, London WC1N 3JH, UK

Background/purpose: Localised scleroderma (LS) is the most common form of scleroderma seen in children, and usually presents unilaterally. Infrared thermography (IRT) and laser Doppler (LD) have both been reported to be useful in assessing the active, inflammatory stage of LS. We developed and validated a protocol using these techniques for the assessment of unilateral LS activity in children.

Method: We investigated the spatial variability and repeatability of LD measurements from adult control forearm skin, and the inter- and intra-operator reproducibility of both LD blood flow trace analysis and IRT skin temperature analysis. Software was developed to produce overlay images of thermograms onto digital photographs of skin sites. In a group of seven adult control subjects, we established the normal range for skin temperature and LD blood flow at six standardised sites (forehead, cheek, abdomen, back, arm and leg), and measured contralateral differences in readings. In a group of 34 children with LS, we investigated the skin temperature and LD blood flow in unaffected skin at the same six sites.

Results: In adults, physiological variability in LD blood flow and skin temperature between the two sides of the body was found to be greater than the uncertainty introduced into the measurements by (inter alia) limited intra- or inter-operator reproducibility. The cheek displayed the highest mean asymmetry in both skin temperature (0.5 °C) and LD blood flow (40%).

Conclusion: Our protocol combines IRT, LD and photography for LS assessment in children, and establishes a normal range of readings in line with other authors.

Key words: localised scleroderma – infrared thermography – laser Doppler flowmetry – image registration – paediatrics

© 2009 John Wiley & Sons A/S
Accepted for publication 15 May 2009

Localised scleroderma (LS), or morphea, is a rare connective tissue disorder, but it is the most common form of scleroderma seen in children (1). LS begins with an active inflammatory stage (Fig. 1a), during which time the lesion extends. This is followed by fibrosis and thinning of the dermis, and often atrophy of subcutaneous fat and muscle (Fig. 1b) (2). Although the condition can occur bilaterally, the majority of LS cases present with fibrotic lesions on one side of the body.

The current treatment for severe cases of active, extending LS is a combination of systemic corticosteroids and methotrexate (3, 4). The appropriate administration of these therapies presupposes reliable and reproducible methods both for the detection of initial disease activity and also the assessment of therapeutic efficacy.

Infrared thermography (IRT) has previously been reported to be of value for the assessment of the inflammatory changes associated with LS (Fig. 1c) (5, 6). Our group has shown IRT to be a sensitive technique for the detection of active LS lesions (7–9). Specificity is limited, however, by ‘false-positive’ thermograms recorded from older LS lesions, which often appear ‘hot’ despite their clinical inactivity (4,8). This is thought to be due to changes in heat conduction through the skin from deep tissues in lesions associated with extensive subcutaneous atrophy.
Because IRT measures skin temperature, it is only a surrogate measure of blood flow (10). The laser Doppler (LD) technique provides a direct measure of blood flow, although fibre-optic LD methods can only assess blood flow within very small volumes of tissue under the probe tip. Serup and Kristensen (11) used LD to assess blood flow in LS lesions, and LD has also been used for the serial assessment of an LS lesion in response to therapy (12).

In order to improve the reliability of our assessments of LS activity, we proposed a new protocol that combined IRT, LD and digital photographic techniques to provide a record of the temperature, blood flow and appearance of each LS lesion. To assist in relating warm areas in thermograms to the physical extent of each lesion, software algorithms were developed to enable registration of infrared thermograms and clinical photographs.

Herein we describe our proposed protocol, along with experiments to validate the technique before use for unilateral LS activity assessment in children. The validation studies comprised of evaluating:

- The spatial variability of LD measurements from control forearm skin.
- The repeatability of LD measurements recorded consecutively from the same area of control forearm skin.
• The inter- and intra-operator reproducibility of LD blood flow trace analysis.
• The inter- and intra-operator reproducibility of IRT skin temperature analysis.
• The normal range for temperature and LD blood flow readings from six standardised skin sites in seven healthy adult control subjects, plus the extent of asymmetry inherent in contralateral IRT and LD readings in this group.
• The range of temperature and LD blood flow readings from unaffected skin contralateral to the lesion in a group of children with unilateral LS.

Method

Study protocol
The Research Ethics Committee of Great Ormond Street Hospital approved our protocol for the assessment of skin in adult controls and paediatric patients.

Equipment
For the measurements we used:

• An SC500 Thermacam uncooled focal plane array thermal imager (FLIR Systems, West Malling, UK), acquired by the Royal Free Hospital in 2001. Thermogram capture and temperature analysis was performed using Thermacam Researcher 2002 software (FLIR Systems).

• A 3 Mpixel Canon Powershot A70 digital camera (Canon (UK) Ltd, Reigate, UK) with ×3 optical zoom (shown mounted onto the SC500 Thermacam in Fig. 2)

• An MBF3D laser blood flow monitor, laser wavelength 810nm (Moor Instruments, Axminster, UK, Fig. 3).

Measurements
All subjects underwent an imaging protocol as follows:

Assessment took place at ‘measurement points’ marked by a clinician (M. V.) with an adhesive aluminium foil triangle of 2 cm shortest side length. The measurement point could then be identified in the thermograms and photographs as the area to which the aluminium arrow pointed. For temperature calculations, the thermographic region of interest was defined as a square of 2 cm side length at the measurement point.

Each subject partially disrobed as appropriate. After an acclimatisation period of 15 min the measurements were performed in the following order:

• IRT appropriate standard views.
• Digital photography corresponding to the thermographic standard views.
• LD flowmetry performed consecutively at each measurement point three times. The monitor was operated in ‘laser-trip’ mode so...
that the laser output ceased if the probe was detached from the skin.

Figure 1d shows a digital photo of the face of a control subject (K. H.) with measurement points marked by aluminium foil on either side of the forehead. Figure 1f shows the corresponding thermogram, with the thermographic regions of interest (ROIs) for temperature calculation included (also exactly corresponding to the areas of LD flow measurement).

LD data was analysed using Moorsoft for Windows software (Moor Instruments). More than 10 s of stable data was selected from the trace from each measurement using the ‘mark’ facility within Moorsoft (Fig. 4), and the average flow in arbitrary units (AU) calculated over a 10 s period. The three time-averaged flux figures calculated for each measurement point were then tabulated in the Microsoft Excel spreadsheet programme. The mean flux over all readings for each measurement point was then calculated, rejecting the most outlying reading of the three if flow varied substantially on repeated measurement at the same site.

Image registration
An image registration software to produce an ‘overlay’ composite image of each thermogram and photo was developed. As the two images are not only of different modalities, but are also typically taken from different perspectives and at slightly different times, this represents a non-trivial problem. Image registration is, however, a well-studied area and we were able to adapt standard approaches to our specific needs.

Image registration algorithms can in general be divided into two different types: landmark-based algorithms where users have to set control points in both images, and intensity-based algorithms that rely only on the image information (13). Landmark-based registration consists of four main stages:

- During the feature detection stage, distinguishing characteristics such as corners, edges, centres of gravity, etc. are identified, either manually or automatically. This identification of landmarks is performed on both the reference (fixed) and sensed (moving) images.
- The optimisation stage controls estimation of transform parameters that geometrically map landmarks between the fixed and moving images. We found an affine transform to be sufficient for these purposes.
- Upon selection of appropriate transform parameters, pixel values which are mapped into non-integer co-ordinates are interpolated in order to establish their value. This represents the image re-sampling stage.
- The feature matching stage is achieved through the use of a similarity metric in which a degree of closeness or accuracy of alignment between corresponding landmarks is calculated.
In intensity-based image registration methods, the feature detection stage is omitted. As a consequence, the transform optimisation and feature matching stages are performed using pixel intensities (or functions thereof) instead of landmarks. Intensity-based image registration algorithms comprise the following components:

- The spatial mapping of intensities throughout the alignment process is achieved with a transform component.
- An interpolation component is used to evaluate intensities at non-discrete locations.
- The metric component calculates a measure of alignment accuracy.
- Optimisation of the similarity measure, using a search space defined by transform parameters, is achieved with an optimisation component.

In order to minimise the user interaction to arrive at a good overlay a two-step approach was implemented. In the first instance, intensity-based registration is applied. This step is carried out completely automatically by the software. Of crucial importance here is the choice of an appropriate similarity metric. As images of two different modalities are involved direct comparison of pixel values will not lead to accurate registration. We therefore used the mutual information metric (14) that is based on the joint probability distribution of the two modalities. The resulting composite image is then shown to the user who can accept it or not. If the overlay is not accepted (because of inaccurate alignment) in a second step landmark-based registration is performed and the user prompted to mark corresponding control points in both the thermogram and the photo.

Once the two images are geometrically aligned, a composite image is generated. This is simply performed by computing a weighted sum of the respective pixel values of the photo and the thermogram. The user has the possibility to interactively change these weights and hence put more emphasis on either of the two modalities. Figure 1g shows a generated overlay composite image of the digital photo in Fig. 1d and the thermogram in Fig. 1f.

**Image database**

Mindful of the large number of medical images of various formats that the protocol would create, we identified appropriate database software for the archiving of the thermograms, photos and overlay composites. We used the ImagedB shareware application (Focus Software Solutions, Inverness, UK), which allows the archiving and display of all the image formats used by the protocol (.jpg, .bmp). All images can be tagged with a user-modifiable descriptor file, enabling each image to be identified as, for example, a thermogram of the head recorded from normal skin from subject KN130698 during protocol visit 1. Images meeting any combination of tag criteria can be displayed via a hierarchical menu structure (Fig. 5). ImagedB has proved invaluable for...
the quick identification of images for research purposes, and also when reviewing patients in the Great Ormond Street LS clinic. ImagedB is backed-up regularly to a secure area of the Royal Free Medical School server, and can be exported onto CD-ROM if required.

**Spatial variability of LD measurements from control forearm skin**

The spatial variability of LD readings across skin was investigated by means of a square $3 \times 3$ template placed on the forearm of a healthy adult control subject (Fig. 6). The nine circular holes in the template allowed the LD probe to be applied to the skin in a grid pattern. The distance between the cell centres was 16 mm, giving a surface area across which the experiment was performed of approximately $10 \text{ cm}^2$. After 15 min of acclimatisation to a room temperature of $15^\circ\text{C}$, a flux reading, $x_i$, was measured for a 10-s period from each cell.

The spatial variability was defined as:

$$\text{Spatial variability} = \text{SD} \left( \frac{x_i - \bar{x}}{\bar{x}} \right)$$

expressed as a percentage figure.

**Repeatability of LD measurements recorded consecutively from the same area of control forearm skin**

The repeatability of LD readings from exactly the same skin site was investigated by marking the forearm of a healthy adult subject with one of the aluminium foil triangles described in ‘Study protocol’. The flux was measured for 10 s on 10 occasions at the triangle tip; each time the probe was detached from the skin and replaced after 1 min. The repeatability was calculated from Eq. (1) with $x_i$ representing an individual 10 s measurement at the triangle tip.

The experiment was repeated by a second operator, and the mean repeatability achieved by the two operators was quoted.

**Intra- and inter-operator reproducibility of LD blood flow trace analysis**

The LD flowmeter values were output to a PC via a serial link, and these data were converted in real-time into graphical format (flux traces) using MoorSoft for Windows. To calculate perfusion values over a specific duration, the operator selected the appropriate part of the trace, and the software calculated the perfusion statistics. The protocol specified the analysis of 10-s bursts of LD data, but for assurance of measurement stability rather longer signals were acquired on each occasion. Consequently, different periods of 10-s duration could be selected within the same trace, potentially leading to different results.

To assess the reproducibility of the LD trace analysis, 840 skin site measurements recorded during the adult control and paediatric patient studies described in ‘Normal range of skin temperature and LD flow readings in control adults’ and ‘Normal range of skin temperature and LD flow readings in control skin of children with LS’ below were analysed twice by the same operator (intra-operator reproducibility), and once each by two different operators (inter-operator reproducibility).

The corresponding coefficients of reproducibility were given by:

**Intra-operator reproducibility:** the SD of the difference between pairs of means obtained by the same operator on repeating the analysis, divided by the average of each pair of means.

**Inter-operator reproducibility:** the SD of the difference between pairs of means obtained by the two operators, divided by the average of each pair of means.

**Intra- and inter-operator reproducibility of IRT skin temperature analysis**

In our protocol, skin temperature was extracted from the thermograms at the same point where LD measurements were made. Using the Thermacam Researcher software (FLIR Systems), ROIs were placed on each thermal image as described in ‘Study protocol’ and Fig. 1f, and
the mean pixel value in degrees Celsius within each ROI was output to a spreadsheet.

As with the LD trace analysis, we assessed the inter- and intra-operator reproducibility of the thermal image analysis. Using thermograms recorded in the studies detailed in ‘Normal range of skin temperature and LD flow readings in control adults’ and ‘Normal range of skin temperature and LD flow readings in control skin of children with LS’ below, a single operator placed a total of 262 ROIs twice at the same sample of measurement points. Two operators then performed the analysis at the same 262 measurement points. Intra- and inter-operator reproducibility was defined as the SD of the difference between pairs of means, expressed in degrees Celsius.

**Normal range of skin temperature and LD flow readings in control adults**

We applied the protocol described in ‘Study protocol’ to assess control skin in seven healthy adult subjects [four females, three males, age 34.5 ± 15.7 years (mean ± SD)] drawn from volunteer staff at the Royal Free and Great Ormond Street Hospitals. The purpose of this phase of the study was to confirm if the measurements were practicable before the involvement of paediatric patients, and to establish reference readings of temperature and LD blood flow from human skin at six body sites typically affected by LS.

Figure 1e shows the six body areas chosen for the assessments, and the left column of Table 1 describes the measurement points. Each measurement point was randomly allocated by MV within each of the six areas on one side of the body. A further six measurement points were placed at precisely contralateral locations, to allow an evaluation of the degree of asymmetry occurring in readings from opposite sides of the body.

**Normal range of skin temperature and LD flow readings in control skin of children with LS**

We next turned our attention to measurements from unaffected skin (on one side of the body) in a group of 34 paediatric patients with LS who underwent the protocol outlined in ‘Study protocol’ for assessment of their disease [23 females, 11 males, age 11.3 ± 3.9 years (mean ± SD)]. In total, 86 healthy skin sites were assessed in these children. The skin sites were divided into the same six groups used for classifying the measurements in adults, shown in Fig. 1e. The right column of Table 1 details the distribution of measurement points, dictated in this instance by the anatomical position of each LS lesion. Because we had recorded no control data from adults specifically from the groin or buttock regions, data collected from these areas during the patient arm of the study was grouped with the ‘abdomen’ and ‘back’ adult data for comparative analysis.

**Results**

**Spatial variability and repeatability of LD measurements from control forearm skin**

The spatial variability of LD flux across the 10 cm² surface area of forearm skin was 13%.

**Repeatability of LD flux (mean of two operators), based on 10 readings from a single point** was 9%.

**Intra- and inter-operator reproducibility of LD blood flow trace analysis**

For both the inter-operator and intra-operator data, the spread of the differences between pairs of readings divided by the mean of readings followed a normal distribution. Hence the reproducibility could be calculated by fitting the data to a Gaussian curve and calculating σ for the distribution.

The intra-operator reproducibility, based on the analysis of 840 LD traces twice by the same operator, was 1%.

<table>
<thead>
<tr>
<th>TABLE 1. Distribution of measurement points across six body areas in the studies of healthy adult control skin, and healthy skin in children with LS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>7 healthy adult controls</strong></td>
</tr>
<tr>
<td>Six reference points per subject, plus six contralateral readings</td>
</tr>
<tr>
<td><strong>Body site [number of readings from healthy skin]</strong></td>
</tr>
<tr>
<td>Cheek [14]</td>
</tr>
<tr>
<td>Leg [14]</td>
</tr>
<tr>
<td>Total 84 points</td>
</tr>
</tbody>
</table>

LS, localised scleroderma.
The inter-operator reproducibility, based on the analysis of the 840 traces by two different operators, was 5%.

Intra- and inter-operator reproducibility of IRT skin temperature analysis

The differences in pairs of temperature readings also followed a normal distribution for the intra-operator and inter-operator reproducibility data, and so $\sigma$ was once again calculated from the best-fit Gaussian curve. Reproducibility was calculated from temperature data expressed in degree Celsius.

The intra-operator reproducibility, based on the analysis of 262 ROIs twice by the same operator, was 0.2°C.

The inter-operator reproducibility, based on the analysis of 262 ROIs by two different operators, was 0.3°C.

Normal range of skin temperature and LD flow readings in control adults

Figure 7 shows the mean temperature recorded from the six body areas by means of IRT, along with the distribution of the readings about the mean value. The forehead proved to be the warmest site, and also demonstrated the least variation about the mean $T$ value. The coolest body sites were at the periphery (arms and legs). Cheek skin showed the greatest variation in temperature about the mean $T$ value.

Figure 8 shows the equivalent LD flux data. Here, when compared with the temperature data, a rather different distribution of the readings across the body sites was seen. The highest LD flux was still noted at the head sites, however, and the lowest flux at the periphery. Measurements from the cheek were once again the data distributed most widely about the mean reading.

Figure 9 plots the differences in temperature recorded between contralateral sites at the six body areas in the seven adult controls. The highest mean asymmetry in temperature between the two sides of the body was recorded at the cheek (0.5°C) but individual measurements of asymmetry as high as 1.2°C were encountered (at the leg).

Figure 10 plots the differences in LD flux between contralateral sites, with the differences expressed as a percentage of the mean value calculated from the two sides of the body. Here again, the cheek showed the greatest mean asymmetry (40%). All individual readings of asymme-
try above 60% were recorded at either the cheek or the leg.

**Normal range of skin temperature and LD flow readings in control skin of children with LS**

Figures 11 and 12 show the mean skin temperature and LD flux recorded in the healthy skin of paediatric LS patients. In these figures we use the ‘Abdomen’ and ‘Back’ category labels to save space on the chart: data were in fact collected from the broader ‘anterior trunk/groin’ and ‘posterior trunk’ body areas detailed in Table 1.

Temperature and flux showed very similar trends across the body sites as for adult controls.

There were significant differences in mean skin temperature between adults and children at the forehead ($T_{\text{adult}} = 34.2 \, ^\circ\text{C}$ vs. $T_{\text{child}} = 33.7 \, ^\circ\text{C}$, $P = 0.03$, $t$-test), posterior trunk/back ($T_{\text{adult}} = 32.9 \, ^\circ\text{C}$ vs. $T_{\text{child}} = 31.5 \, ^\circ\text{C}$, $P = 0.05$, $t$-test) and arm ($T_{\text{adult}} = 32.4 \, ^\circ\text{C}$ vs. $T_{\text{child}} = 31.6 \, ^\circ\text{C}$, $P = 0.04$, $t$-test). However, the only significant differences in LD blood flow between adults and children were recorded at the leg ($\text{LDF}_{\text{adult}} = 9.3 \, \text{AU}$ vs. $\text{LDF}_{\text{child}} = 11.7 \, \text{AU}$, $P = 0.02$, $t$-test).

**Discussion**

Ours is the first published protocol to combine thermographic skin temperature data, LD blood flow measurements, and digital photography for the assessment of disease activity in LS.

For IRT and LD flowmetry to be sensitive measures of LS disease activity, the techniques need to show minimal spatial and temporal variability in healthy skin. Additionally, if the protocol is to be of value for serial assessment of patients, good intra-operator repeatability is essential. Repeatable measurements between different operators (inter-operator) are also necessary if the protocol is to be adopted across multiple centres.

Our pilot study has also gathered information about the normal range of skin temperature and LD flux: without this it is not possible to reliably conclude which are elevated readings suggestive of active LS in patients.

Our study found a spatial variability for LD flux across a 10 cm$^2$ area of forearm skin of 13%, and a repeatability of 9% for measurements repeated from a single point. In comparison, the variability introduced to the measurements by operator analysis of the LD traces was low (1% intra-operator, 5% inter-operator). Hence spatial and temporal heterogeneity due to physiological factors is the predominant contributor to variability in LD readings. At the human forearm, this variability may be considered to be of the order of 10% over a few minutes and a few square centimetres. Provided, therefore, that measurements are made from localised areas over short periods of time, contralateral LD readings that differ by little $>$10% ought to be readily detectable. Further work is required to determine the spatial variability in LD flux across small surface areas at sites other than the forearm.

Figure 8 shows that the mean contralateral difference in LD flux in adults is in fact of the order of 20% at most body sites, and around 40% at the cheek. In addition, high individual differences in some subjects at the cheek, arm and leg might limit the sensitivity of LD flowmetry for detecting inflammation in LS plaques at these body sites.

The intra-operator dependence on skin temperature readings was just 0.2 $^\circ\text{C}$, and analysing the measurement site temperatures with two different operators increased the uncertainty to 0.3 $^\circ\text{C}$.

In their study of childhood LS, Birdi et al. (6) assumed a contralateral temperature difference of 0.5 $^\circ\text{C}$ or greater (taking into account the entire lesion by simple visual inspection of the thermogram) would be suggestive of active, inflamed
LS. We subsequently used this temperature cut-off in a larger study of childhood LS at the Royal Free and Great Ormond Street Hospitals. Our data from healthy adult controls shows that the mean contralateral temperature difference was indeed <0.5 °C at all body sites. We found some individual contralateral differences to be >0.8 °C, however, at the cheek, back and leg. Our data suggest that, when considering localised ROIs instead of the thermogram ‘as a whole’, contralateral temperature differences of the order of 1 °C can be expected at some body sites in healthy individuals.

We did not specifically investigate the variability introduced into the LD and temperature measurements from the intra- and inter-operator dependence on positioning of the triangular marker on the skin. This may be an important factor in serial patient assessments, where the markers need to be placed in the same skin positions several months apart, using the photographic record as a visual guide. Because the error that will arise in positioning the triangular marker is analogous to small variations in positioning the LD probe or thermographic ROI, it would be reasonable to assume that the uncertainty induced would be of a similar magnitude to that reported herein.

Whilst mean skin temperature was similar overall between adults and children, we found the skin to be significantly cooler in children at the forehead, back and arm. This was not attributable to disparity in skin blood flow because there were no statistically significant differences in LD flux between the subject groups at any of these three body sites. Children may have cooler skin than adults at certain body sites due to variations in skin thickness, limb size and bulk or metabolic rate.

The only difference we found in skin blood flux between adults and children was a higher flux at the legs in children, and this was a very small absolute difference in flow AU.

In general our data showed a nearly identical distribution across body sites for both LD flux and skin temperature, when comparing healthy skin of children with LS and adult controls.

Both Stücker et al. (15) and Park et al. (16) gained similar LD results to ours when investigating the heterogeneity of cutaneous microvascular blood flow across different body sites and age groups. It is noteworthy that all three studies have drawn comparable conclusions, despite differences in the specification of the LD apparatus used, which can affect both the volume and depth of the skin flow measurements made.

Our study only recruited a paediatric patient group with LS, so we were not able to measure contralateral differences in skin temperature and LD flux in healthy children. It might be reasonable to consider that healthy adults and children share similar symmetry between the two sides of the body for skin temperature and blood flow. Using our protocol, initial data from inactive morphea lesions in children supports this hypothesis (17).

In conclusion, we have developed a new protocol for the assessment of blood flow and skin temperature in unilateral forms of LS. The largest source of uncertainty in these measurements arises from the inherent difference between the two sides of the body in blood flow and skin temperature. At most body sites, however, these differences are acceptably small. Work in follow-up to these findings has identified a cut-off value of LD flux elevation for the detection of active LS lesions (17).

References


Address:
Kevin J Howell
Department of Rheumatology
Royal Free Hospital
Pond Street
London NW3 2QG
UK
Tel: +44 20 7472 6550
Fax: +44 870 133 1058
e-mail: k.howell@medsch.ucl.ac.uk
Appendix 11

Statement of Support

I confirm the following contributions from Kevin Howell to the publication below:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Contribution</th>
</tr>
</thead>
</table>
• Thermography and laser Doppler protocol design (100%)  
• Data collection (50%)  
• Data analysis (20%)  
• Preparation of manuscript (20%) |

[Signature]

Dr L WEIBEL

University Hospital of Zurich  
Switzerland  
24th August 2009
Laser Doppler Flowmetry for Assessing Localized Scleroderma in Children

Lisa Weibel,1 Kevin J. Howell,2 Maria Teresa Visentin,3 Alain Rudiger,4 Christopher P. Denton,2 Francesco Zulian,3 Patricia Woo,1 and John I. Harper1

Objective. Assessment of disease activity is a major challenge in the management of children with localized scleroderma. The aim of this study was to evaluate the role of laser Doppler flowmetry (LDF) in comparison with infrared thermography in the detection of scleroderma disease activity.

Methods. In 41 children with localized scleroderma, 111 lesions were assessed on 2 separate occasions, by clinical examination, LDF, and thermography. Measurements from contralateral areas of unaffected skin served as intrapatient controls, and differences in blood flow and temperature were calculated between the corresponding sites. The sensitivity and specificity to detect clinically active lesions were compared between LDF and thermography.

Results. Seventy-five active lesions (34%) and 147 inactive lesions (66%) were identified clinically. The median relative increase in blood flow measured by LDF was +89% (range −69% to +449%) for clinically active lesions and +11% (range −46% to +302%) for clinically inactive lesions (P < 0.001). Thermography showed a median difference in temperature of +0.5°C (range −0.1°C to +4.1°C) and +0.3°C (range −1.9°C to +2.7°C) for clinically active lesions and clinically inactive lesions, respectively (P = 0.024). Using a cutoff level of 39% to indicate increase in blood flow, a sensitivity of 80% and specificity of 77% to detect clinically active lesions were observed; for thermography, no useful cutoff level was identified. The correlation between differences in blood flow and differences in temperature was small, but significant (r2 = 0.120, P < 0.001).

Conclusion. LDF is a helpful, noninvasive diagnostic technique that can be used to discriminate disease activity in children with localized scleroderma, and is more accurate than thermography for this purpose.

Localized scleroderma, or morphea, is a rare connective tissue disorder characterized by hardening and thickening of the skin due to an increased density of collagen (1,2). The course of localized scleroderma includes an early inflammation stage with hyperemia of the skin, followed by fibrosis, sclerosis, and finally, atrophy. The condition usually begins in childhood and is much more common in the pediatric population than is systemic sclerosis. The disease shows a great variety in its clinical presentation and has been classified into the clinical subtypes of plaque or circumscribed morphea, linear scleroderma, including scleroderma en coup de sabre, as well as generalized morphea profunda (deep), pansclerotic, and combined forms (1,3).

In general, localized scleroderma is considered to be a condition confined to the skin and subcutaneous tissue and characterized by a benign, self-limiting nature. However, it often involves underlying muscle and bone. Of note, extracutaneous manifestations of the disease can be found in almost one-quarter of affected children (4). At the more severe end of the spectrum, the disease can progress over several years and may cause substantial atrophy, growth retardation, irrevers-
ible structural deformities, joint contractures, and severe functional, cosmetic, and psychological disabilities.

In children, the current treatment for severe localized scleroderma is a combination of systemic corticosteroids and low-dose methotrexate (5–8). The management of severe localized scleroderma is challenging, and the detection of disease activity remains a fundamental problem, both in the evaluation of the need for treatment and the assessment of therapeutic efficacy over time. Clinical examination is subjective and sometimes remains unsatisfactory, and laboratory tests are not helpful for this purpose. Reliable and reproducible methods are needed to detect and monitor disease activity.

Thermography has previously been reported to be a helpful tool to assess disease activity in children with localized scleroderma (5,9,10). However, a limitation of this technique became evident when thermography was found to yield false-positive results in the assessment of patients with older lesions of localized scleroderma, characterized by marked atrophy of the skin, subcutaneous fat, or muscles (5,9).

Laser Doppler flowmetry (LDF) is a noninvasive method for the measurement of cutaneous microcirculation and has a broad range of applications in dermatology and microvascular surgery (11–13). It remains unknown whether blood flow levels measured by LDF correlate with clinical findings related to disease activity in juvenile localized scleroderma. In this study we assessed disease activity in a cohort of pediatric patients with localized scleroderma, by clinical examination and, simultaneously, by LDF and infrared thermography. Our aim was to evaluate the role of LDF in detecting disease activity in juvenile localized scleroderma, and to define the sensitivity and specificity of this technique in comparison with infrared thermography.

**PATIENTS AND METHODS**

**Patients.** The study was performed between April 2005 and November 2006 at the Pediatric Dermatology and Rheumatology Units at Great Ormond Street Hospital for Children, in collaboration with the Rheumatology Department at the Royal Free Hospital (London, UK). Children with a minimum age of 2 years were recruited. The diagnosis of localized scleroderma was made clinically by an experienced pediatric dermatologist (JIH) and pediatric rheumatologist (PW). Children with generalized localized scleroderma in whom a right-to-left-side comparison between affected and normal skin was impossible were excluded. Written informed consent was obtained from the children’s parents. The study protocol was approved by the local ethics committee.

**Study design.** All patients were assessed by the same investigators (LW and KJH) according to a standardized protocol, which included clinical examination, LDF, and thermography. During the study period, every patient underwent evaluations at 2 different time points, performed a minimum of 2 months apart.

All patients were assessed clinically by the same investigator (LW). The subtype of localized scleroderma, the site and extent of the lesions, and current treatment were noted. In every child, 3 measurement points were selected to represent the areas of suspected disease activity, as follows: 1) the border of the affected skin; 2) the areas in which lesions were spreading or had most recently spread; and 3) in the case of a single lesion, 3 different sites across the affected area, or in the case of multiple lesions, 3 different lesions in separate areas.

These selected sites were marked with small arrow-shaped aluminum foil stickers. For single, small lesions in localized scleroderma, as was identified in some children with scleroderma en coup de sabre, the number of measurement points was reduced, as appropriate. Corresponding measurement points were marked at the contralateral, unaffected side of the body, to serve as intrapatient control sites (Figure 1). All marked sites were assessed clinically, followed by thermography and LDF evaluations. At the second study visit, the same measurement points were assessed, with reference to digital photographs of these same sites.

**Clinical assessment.** The marked areas of localized scleroderma were described clinically as either active or inactive. Lesions that were spreading and/or showing erythema were defined as active, whereas all others were described as inactive. We noted which of the marked sites represented the most active lesion in a patient. This was defined as the site with the maximum extent of inflammation and/or spreading. In patients with no clinically active lesions, the most recently active area was noted.

**Thermography.** The technique of thermography used in this study was carried out as has been previously described (9). Using this technique, we recorded the temperature over an area of 1 × 1 cm at all measurement points and control sites. The thermograms were obtained using the same infrared camera (FLIR SC 500 Thermacam; Flir Systems, West Malling, UK) to assess the skin of each patient (Figure 1).

**Laser Doppler flowmetry.** LDF is a noninvasive method for the measurement of cutaneous microcirculation (11–14). The tissue volume assessed has a surface area of ~1 mm² and a depth of 1–2 mm, which means that, predominantly, the vasculature of the papillary dermal bed is assessed. We used an MBF3D laser Doppler monitor with a laser wavelength of 810 nm (Moor Instruments, Axminster, Devon, UK). Measurements were performed using an optical-fiber probe attached to the skin with a self-adhesive disc. Blood flow was monitored for 10 seconds at each skin site sequentially. Measurements were performed 3 times, and the mean flow, in arbitrary flux units, was calculated at each site (Figure 1). A maximum of 1 reading per site was discounted to remove “noisy” data attributable to movement artifacts.

**Statistical analysis.** The absolute difference in temperature (expressed in °C) between the affected side and the corresponding, contralateral control site was calculated for each measurement point. Blood flow levels tend to show a broad variation among different body areas (15). We therefore...
calculated the relative difference in blood flow between a lesion and the corresponding control site as follows: \[
\frac{(\text{blood flow of affected site}) - (\text{blood flow of unaffected site})}{\text{blood flow of unaffected site}} \times 100.
\]
Results are expressed as the percentage increase (+) or decrease (−) in blood flow.

We analyzed the data from all patients, but also in separate analyses of the children with scleroderma en coup de sabre and those with lesions on the trunk and/or limbs, since patients with en coup de sabre lesions may represent a distinct clinical subgroup in whom clinical signs of inflammation are often lacking despite the presence of active disease. We considered all assessed lesions in each patient (usually 3 sites per patient) as being representative of the overall clinical picture in the patient. Since multiple measurements per patient were not independent from each other, and since the number of assessed sites per patient was sometimes smaller than 3, it could be argued that this introduced bias into the analysis. We therefore undertook a subanalysis of the clinically most active site or the most recently active site in each child, thus analyzing only 1 site per patient.

Results are expressed as the mean ± SD, median (range), or percentage, as appropriate. Comparisons were made with the use of the t-test, Fisher’s exact test, or chi-square test.

Table 1. Main characteristics of the 41 patients with localized scleroderma

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female-to-male ratio</td>
<td>2.2:1</td>
</tr>
<tr>
<td>Disease subtype, no. (%)</td>
<td></td>
</tr>
<tr>
<td>En coup de sabre</td>
<td>18 (44)</td>
</tr>
<tr>
<td>Linear of trunk/limbs</td>
<td>19 (46)</td>
</tr>
<tr>
<td>En coup de sabre combined with linear of trunk/limbs</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Deep</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Age at disease onset, years</td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>5.1 ± 2.9</td>
</tr>
<tr>
<td>Median (range)</td>
<td>4.5 (1.5–11.6)</td>
</tr>
<tr>
<td>Age at study entry, years</td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>10.2 ± 3.0</td>
</tr>
<tr>
<td>Median (range)</td>
<td>10.0 (4.7–15.7)</td>
</tr>
<tr>
<td>Treatment at study entry, no. (%)</td>
<td></td>
</tr>
<tr>
<td>MTX</td>
<td>16 (39)</td>
</tr>
<tr>
<td>Systemic corticosteroids and MTX</td>
<td>6 (15)</td>
</tr>
<tr>
<td>No treatment</td>
<td>19 (46)</td>
</tr>
<tr>
<td>Pretreatment*</td>
<td>16 (39)</td>
</tr>
<tr>
<td>Off treatment†</td>
<td>3 (7)</td>
</tr>
</tbody>
</table>

* Patients were started on treatment with systemic corticosteroids and methotrexate (MTX) after the first study assessment.
† Having stopped treatment with systemic corticosteroids and MTX.
square test, as appropriate. Linear correlations were described by the Pearson’s product-moment correlation coefficient (r and \( r^2 \)). The quality of the diagnostic tests was described by the area under the receiver operating characteristic (ROC) curve and its 95% confidence interval (95% CI). In addition, cutoff values were identified in order to obtain the highest sensitivity and specificity values in combined analyses. The null hypothesis was rejected by setting the significance level as a 2-sided \( P \) value less than 0.05. All analyses were performed with the use of SPSS for Macintosh OS X (version 11.0; SPSS, Chicago, IL).

**RESULTS**

**Characteristics of the patients.** Forty-five patients with localized scleroderma were enrolled in the study. Four of the children were unable to attend a second study visit, and were therefore excluded from the analysis. A total of 41 patients was included, among whom 111 lesions were identified for analysis. The epidemiologic and clinical features of the patients are summarized in Table 1. Thirty-nine patients had either linear localized scleroderma or a combination with the linear subtype, while 2 patients (5%) had the deep subtype. In 18 patients (44%), >5% of the total body surface area was affected. The mean ± SD disease duration at the time of study entry was 5.1 ± 3.5 years. The mean ± SD time interval between the 2 study visits was 7.9 ± 5 months. Thus, a total of 222 affected skin lesions was assessed over the study period, and each was compared with its respective control site.

**Findings of clinical assessment, LDF, and thermography.** Clinical examination of the skin sites revealed 75 active lesions (34%) and 147 inactive lesions (66%). The median relative increase in blood flow measured by LDF was +89% (–69% to +449%) for clinically active lesions and +11% (–46% to +302%) for clinically inactive lesions (\( P < 0.001 \)) (Figure 2A). Thermography showed a median difference in temperature of +0.5°C (–0.1°C to +4.1°C) for clinically active lesions and +0.3°C (–1.9°C to +2.7°C) for clinically inactive lesions (\( P = 0.024 \)) (Figure 2B). Table 2 demonstrates the differences in blood flow and temperature for the clinical subgroups of patients with en coup de sabre lesions and those with other types of localized scleroderma lesions.

The area under the ROC curve for the relative difference in blood flow, as a measure of the ability of LDF to detect clinically active lesions, was 0.80 (95% CI 0.73–0.87; \( P < 0.001 \)), as shown in Figure 3A. A cutoff level of 39% as an indicator of increase in blood flow had a sensitivity of 80% and specificity of 77% to detect clinically active lesions. In contrast, the area under the ROC curve for absolute differences in temperature, as a measure of the ability of thermography to distinguish between clinically active and clinically inactive lesions, was only 0.59 (95% CI 0.52–0.67; \( P = 0.025 \)). A cutoff level of 0.5°C as an indicator of increase in temperature

<table>
<thead>
<tr>
<th>Table 2. Differences in blood flow and temperature between clinically active and clinically inactive lesions in patients with and those without scleroderma en coup de sabre lesions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative difference in blood flow, median (range) %</td>
</tr>
<tr>
<td>En coup de sabre lesions (n = 94)</td>
</tr>
<tr>
<td>Clinically active (n = 28)</td>
</tr>
<tr>
<td>Clinically inactive (n = 66)</td>
</tr>
<tr>
<td>( P )</td>
</tr>
<tr>
<td>Lesions other than en coup de sabre (n = 128)</td>
</tr>
<tr>
<td>Clinically active (n = 47)</td>
</tr>
<tr>
<td>Clinically inactive (n = 81)</td>
</tr>
<tr>
<td>( P )</td>
</tr>
</tbody>
</table>
had a sensitivity of 52% and specificity of 58% to detect clinically active lesions. Table 3 shows the details of the ROC curves for the differences in blood flow and differences in temperature in patients with en coup de sabre lesions and those with other types of localized scleroderma lesions.

**Correlation between differences in blood flow and differences in temperature.** When comparing the differences in blood flow and differences in temperature, we found a small, but significant correlation between the 2 measures, with a Pearson’s correlation coefficient of 0.447 (n = 222; r² = 0.120, P < 0.001). The correlation between the LDF findings and the thermography findings did not markedly improve when calculated for the following subgroups: clinically active lesions (n = 75; r = 0.354, r² = 0.125, P = 0.002), clinically inactive lesions (n = 147; r = 0.268, r² = 0.072, P = 0.001), scleroderma en coup de sabre lesions, including combined forms (n = 94; r = 0.254, r² = 0.065, P = 0.018), and localized scleroderma lesions other than the en coup de sabre subtype (n = 128; r = 0.405, r² = 0.164, P < 0.001).

**Subanalysis of a single lesion per patient.** In the subanalysis involving the clinically most active site or most recently active site in each patient, we measured a total of 82 lesions in 41 patients, in comparison with the respective control site. Blood flow assessed by LDF remained significantly higher in active lesions (n = 43) than in inactive lesions (n = 39) (median relative difference in blood flow +100%, range +2% to +449% in active lesions versus +13%, range −34% to +288% in inactive lesions; P < 0.001). However, the subanalysis assessment by thermography did not show any significant difference between clinically active lesions and clinically inactive lesions.

In this subanalysis, the area under the ROC curve for the LDF findings was slightly larger (0.88, 95% CI 0.80–0.96) than that described for all 222 lesions (P < 0.001) (Figure 3B). In contrast, the area under the ROC curve for the thermography findings was even smaller (0.48, 95% CI 0.35–0.61) than that described for all 222 lesions (P = 0.066). A cutoff level of 45% for the relative difference in blood flow had a sensitivity and specificity of 86% and 85%, respectively. In this subanalysis, there was no correlation between differences in blood flow and differences in temperature (P = 0.076). Moreover, the results of this subanalysis did not differ when patients with and those without scleroderma en coup de sabre lesions were analyzed separately (results not shown).

### DISCUSSION

This is the first study to evaluate LDF as a tool to detect disease activity in a large cohort of children with localized scleroderma. Blood flow measured by LDF was significantly increased in clinically active localized scleroderma lesions. We found that LDF had a sensitivity of 80% and specificity of 77% to detect active lesions,

<table>
<thead>
<tr>
<th>Relative difference in blood flow</th>
<th>Absolute difference in temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC (95% CI)</td>
</tr>
<tr>
<td>En coup de sabre lesions (n = 94)</td>
<td>0.88 (0.81–0.95)</td>
</tr>
<tr>
<td>Lesions other than en coup de sabre (n = 128)</td>
<td>0.75 (0.65–0.85)</td>
</tr>
</tbody>
</table>

* Values are the area under the curve (AUC) (95% confidence interval [95% CI]). P values are in comparison with the AUC values of the other group for each measure.
if the blood flow was found to be increased by at least 39%. The subanalysis of only 1 lesion per patient showed an even greater sensitivity and specificity of LDF (86% and 85%, respectively). Thermography, in contrast, was not helpful in the detection of disease activity in the children in this study.

The clinical presentation of localized scleroderma is varied. Some lesions, particularly the scleroderma en coup de sabre type, often do not show any obvious clinical signs of inflammation, although the disease remains active and continues to progress. In this situation, as well as during followup of treatment, it is often difficult to determine whether a lesion is active, and this cannot be determined by clinical examination only.

LDF is a noninvasive method for the measurement of cutaneous microcirculation. It has been used to investigate a range of medical conditions, such as inflammatory skin disorders, vasospastic vascular disorders, neuropathies, the response of microcirculation to neurotransmitters, tumors, ulcers, and burns, and has been used to monitor the microcirculation in skin flaps or grafts following intestinal, orthopedic, or plastic surgery procedures (11–13). The use of LDF for the assessment of localized scleroderma in adults has been described in 2 small studies and 1 case report (16–18). Serup and Kristensen (16) performed LDF in 15 adults with localized scleroderma, and Kalis et al (17) applied this method in 16 adults with localized scleroderma. As in our study, those authors found blood flow to be increased in localized scleroderma lesions. However, they did not describe levels of blood flow in relation to clinical disease activity.

Serup and Kristensen (16) performed LDF measurements in the sclerotic center of plaques, as well as in perilesional areas, and found blood flow to be elevated in both sites. In our study, we limited the assessment to the borders of lesions, in order to standardize the measurements. Kalis et al (17) assessed blood flow in the center of localized scleroderma plaques, and distinguished between clinically progressive, static, and resolving lesions. Those authors found increased blood flow in equal proportions in all groups (mean ±40% [±SD 10%]), whereas in our study, blood flow was significantly higher in the clinically active lesions, by a median of +89%. However, compared with normal skin, we also recorded a slight increase in blood flow in clinically inactive lesions. It remains unclear whether this is clinically relevant.

LDF values do not show any relevant intrainsidividual differences between corresponding body sites in healthy subjects (15). It is possible that the slightly higher blood flow in clinically inactive lesions reflects persistent changes in the microcirculation caused by disease progression. However, it is more likely that some inflammation is ongoing in a proportion of these lesions, and not detectable by clinical examination alone. In fact, sudden reactivation of clinically stable localized scleroderma, and thus reiterates the need for reliable and reproducible methods. Histopathologic investigations represent the gold standard, but repeated skin biopsies are not an appropriate option to monitor localized scleroderma in children.

The results of this study suggest that LDF is an important adjunct to clinical examination for the purpose of detection of disease activity. Prospective studies are needed to further investigate the potential prognostic value of LDF in localized scleroderma. LDF can be used to investigate only a small skin area, and measurements are susceptible to motion artifacts. The more recently developed technique of laser Doppler imaging (LDI) allows evaluation of blood flow over a specific area of skin while avoiding contact with the skin surface (11). Nevertheless, current LDI devices are of limited use in children, because of their slow scanning times. With further development, LDI may become a valuable technique.

In the present study thermographic imaging did not prove to be helpful in distinguishing between active and inactive localized scleroderma lesions. This is in contrast to the findings in previous reports from our institution and others (5,9,10). Martini et al (9) reported a sensitivity and specificity of thermography of 92% and 68%, respectively, to detect active lesions. In that retrospective study, thermograms were visually assessed, and a lesion was considered thermography positive if it appeared at least 0.5°C warmer than the matching opposite body site. This is different from the method used in the present study, in which we limited the area of thermography to a small site within a localized scleroderma lesion and measured the actual temperature in °C instead of making a global visual judgment.

We suspect that the main reason for the thermography results obtained in the present study is that considerable numbers of long-standing localized scleroderma lesions were included. Previous studies showed that thermography produced false-positive results in patients with older localized scleroderma lesions, in whom marked atrophy of skin, subcutaneous fat, and
muscle was present (5,9). In these situations, positive findings on thermography probably represent increased heat conduction from deeper tissues, rather than active inflammation in the affected skin. This mechanism is likely to have contributed to the poor correlation between the results of LDF and those of thermography as seen in the present study. It was previously shown, in healthy volunteers, that skin temperature does not correlate well with cutaneous blood flow, which supports the notion that thermography is not a good measure of skin perfusion (14).

Although the precise childhood incidence is unknown, juvenile localized scleroderma is a rare condition, and our single-center analysis involved a large cohort of 41 children. A limitation of this study is the lack of investigator blinding. However, this is unlikely to have influenced our results, since the techniques of LDF and thermography are essentially operator independent (15). In this study all patients were assessed consistently by the same investigator, and we did not evaluate any interobserver variability. More studies are needed to identify a gold standard for the detection of disease activity in children with localized scleroderma, particularly a method that would be suitable to utilize in the clinical routine.

Our results thus show that LDF is a valuable, noninvasive diagnostic technique that can be used to discriminate disease activity in localized scleroderma, and is more accurate for this purpose than thermography. A prospective study in which we are evaluating the application of LDF in monitoring disease progression, response to treatment, and the relationship between changes in blood flow and skin structure in children with localized scleroderma is ongoing. By integrating spatial and flow data, the newer method of LDI represents an interesting technique for future study.

AUTHOR CONTRIBUTIONS
Dr. Weibel had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.


Acquisition of data. Weibel, Howell, Visentin.


Statistical analysis. Weibel, Visentin, Rudiger.

REFERENCES
Appendix 12

Statement of Support

I confirm the following contributions from Kevin Howell to the publication below:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-field-of-view thermal image calibration system for medical thermography applications. Int J Thermophys 2008 ;29 :1123-1130</td>
<td>• Field trial data collection (33%)</td>
</tr>
</tbody>
</table>

Dr R SIMPSON
National Physical Laboratory
Teddington
24th August 2009
In-Field-of-View Thermal Image Calibration System for Medical Thermography Applications

R. C. Simpson · H. C. McEvoy · G. Machin · K. Howell · M. Naeem · P. Plassmann · F. Ring · P. Campbell · C. Song · J. Tavener · I. Ridley

Published online: 2 May 2008
© Crown copyright. Reproduced by permission of the Controller of HMSO and the Queen’s printer for Scotland 2008

Abstract  Medical thermography has become ever more accessible to hospitals, medical research, and clinical centers with the new generation of thermal cameras, which are easier to use and lower in cost. Some diagnostic techniques using thermal cameras are now regarded as standardized, such as the cold challenge test for Raynaud’s phenomenon. The future for medical thermography appears to be improved accuracy, standardization, and establishment as a mainstream medical imaging methodology. Medical thermography standardization, quantitative measurements, image comparison, and multi-center research trials all require thermal cameras to provide a demonstrably traceable, accurate, and reliable temperature output. To this end, the National Physical Laboratory (NPL) has developed a multi-fixed-point source that serves as an in-image calibration system, thereby providing a reliable means for radiometric image validation. An in-field-of-view fixed-point validation system for thermal imaging has successfully been developed, tested, and validated at NPL and has undergone field trials at three clinical centers in the UK. The sources use the phase change
plateaux of gallium–zinc eutectic, gallium, and ethylene carbonate. The fixed-point sources have an estimated cavity emissivity of greater than 0.998, a plateau longevity of nominally 3 h at ambient conditions, a stability of 0.1°C, or better, over that period, a repeatability of 0.1°C or better, and an estimated temperature uncertainty of ±0.4°C ($k = 2$). In this article, the source specifications and design as well as testing, validation, and field trial results are described in detail.

**Keywords**  Ethylene carbonate · Fixed point · Gallium · Gallium–zinc · Thermal image

1 Introduction

Medical thermographers use thermal imagers to diagnose and monitor patient health. In most applications, thermal images are used for anomalous variations in skin temperature. These applications typically require a thermal imager to have good thermal and spatial resolution. The temperature information is also used for periodic monitoring of patient progress and image comparison in a clinical center and between clinical centers, as well as procedural development and standardization. These applications require a thermal imager to have, in addition to the previous requirements, good medium- to long-term stability and a regular accredited calibration traceable to the International Temperature Scale of 1990 (ITS-90). Without such, any temperature measurements made will carry significant and often unquantified measurement uncertainty.

It is clear that many medical thermography applications do not require absolute temperature measurements (e.g., when used in a purely imaging mode, there is little desire to carry out regular traceable calibrations—other than to ensure that the imager is operating to its full spatial and thermal capability). However, previous work by National Physical Laboratory (NPL) assessing the performance of thermal imagers used in medical thermography identified poor or non-existent traceability to ITS-90 with a temperature dispersion between the imagers in the region of ±2°C [1]. The disparity was much larger than expected by the medical thermography community and would be very significant for any work involving absolute rather than relative temperature measurement, such as standardization, charting patient progress (either of the disease or of the response to treatment), and image comparison in-center and cross-center.

In response to this, NPL has been working with UK medical thermographers to put in place a robust traceability regime [1]. The developments described here are the next step in putting clinical thermography on a firm traceable basis. NPL has developed an in-image fixed-point calibration system, covering the normal body-temperature range (Fig. 1). This system will provide a means of calibrating the image temperature for high-level thermography measurement and thereby introduce increased measurement rigor and confidence while being a tool to aid thermography standardization and national/international comparisons.
2 Measurements

In order to provide an in-image calibration system, it was concluded that three fixed-point blackbody sources would be required to cover the normal medical thermography temperature measurement range, nominally 25–35°C. The normal target-to-imager focal distance in medical thermography is nominally 1 m and a reasonable estimate of array size for imagers would be a minimum 256 × 256. A source aperture size of 26 mm would provide an image size to cover a minimum of 9 pixels (a 3 × 3 block) with which to obtain an average of the temperature from the blackbody aperture. With medical thermography used for a large variety of applications, it was decided that the fixed-point blackbodies should be able to operate without additional temperature control once initiated, allowing them to be orientated around the patient under investigation. A nominal uncertainty requirement of ±0.2°C for the fixed-point temperature was agreed in order to provide an effective calibration system.

A number of materials were considered for the fixed points, both high-purity metals and high-purity organics. The selection was largely governed by previous knowledge of their use as temperature fixed-point materials and their proximity to the required temperature range. The materials finally selected for use were gallium–zinc (∼25°C), gallium (29.7646°C), and ethylene carbonate (36.3°C). The gallium–zinc eutectic melting point has never, to our knowledge, been used as a fixed point for thermal metrology. While some work has been carried out in identifying its
characteristics [2], there is limited information concerning its suitability as a fixed point for temperature-measurement applications. The gallium melting point is a very well-known defining fixed point of the ITS-90, and its performance is well-documented. The ethylene carbonate freezing point has also been used as a thermal metrology fixed-point, although not as widely as gallium. It was also used recently in a project to produce an infrared ear thermometry fixed-point validator [3].

Construction material selection was based on fixed-point material compatibility and surface emissivity. While the cavities could be painted with high emissivity paint, it was concluded that this would decrease their overall suitability for use in a clinical environment. The selected cavity construction material was black Teflon (PTFE). This was known to be compatible with the proposed fixed-point materials, providing a robust fixed point with a high surface emissivity [3].

The blackbody cavity design comprised a cylindro-cone with dimensions of 150 mm in length, 26-mm aperture diameter, and a 120° cone for the back wall. The cavity emissivity, for isothermal conditions, was calculated to be 0.9983 using the previously determined Teflon (PTFE) surface emissivity value, over the 8–13 µm range, of 0.79 [3]. The fixed-point cells were constructed by Isothermal Technology Ltd.

The validation testing consisted of repeatability, stability, precise temperature definition, and uncertainty estimation. Dry block Peltier coolers/heaters were used to heat/cool the fixed-point cells. Several preliminary measurement runs were performed to establish the controller settings and initiation procedure for each fixed-point cell. A Land Cyclops C300 IR thermometer, which has an operating wavelength of 8–13 µm, was used as the transfer radiation thermometer and the NPL’s ammonia heat-pipe blackbody was used as the comparison blackbody [4].

For the gallium-based fixed points, the procedure was to initially cool the fixed point to 10°C for 2–4 h using the Peltier heater/coolers to ensure that the fixed-point material was completely frozen. Each fixed-point cell was then ramped to a temperature just above its melting-point temperature. At this stage, the cells were then either left to pass straight through their melting plateaux, taking approximately 4–5 h, while being monitored with a Land Instruments Cyclops 300 (C300) IR thermometer mounted vertically above the fixed point, or alternatively, the cells were removed from the Peltier unit after having completed approximately 2–3 h of their plateaux, and then monitored using the C300 thermometer mounted horizontally.

The initial measurements were used to determine controller settings for the Peltier units and to measure plateau stability. The subsequent measurements were used to determine the fixed-point temperature, repeatability, and longevity of plateaux. In this second case, the C300 was used to compare the fixed-point source to the NPL ammonia heat-pipe blackbody.

The procedure for initiating the ethylene carbonate fixed point was as follows. While solid, the cell was placed into a Dewar containing water at approximately 75°C, enough to cover almost the entire fixed point, leaving only about 1 cm at the top of the fixed point uncovered. After 1 h, the pre-boiled water was replaced with fresh hot water. At this point, the fixed-point cell was inverted a number of times to mix the fixed-point material. After a further hour, the fixed point was removed and again inverted to ensure good mixing. The water in the Dewar was replaced with tap water, nominally at 20°C, and the fixed point was again put in the Dewar with the water at
the same level. After a further hour, the water in the dewar was removed and replaced with fresh tap water. After a further hour, the fixed point was removed and initiated using a physical shock (a sharp tap on the bench) and monitored using the C300 IR thermometer.

Several melt and freeze cycles of each fixed-point source were measured. From this data, the fixed points’ stability, repeatability, radiance temperature, and overall uncertainty were determined. Following the successful completion of the validation, the fixed points were transported to three clinical centers for field trials. Each clinical center had the fixed-point system for a month, and the fixed points were used in the centers’ normal applications.

3 Results

The temperature determination of the fixed-point source was made using the C300 IR thermometer as the transfer radiation thermometer and the NPL ammonia black-body source as the comparison blackbody source. The average temperatures of the sources were: gallium–zinc = 25.3°C; gallium = 29.8°C; and ethylene carbonate = 35.9°C. An uncertainty of ±0.4°C \((k = 2)\) was attributed to each cell (Table 1).

The temperature stability of each of the cells was determined from the standard deviation of a number of fixed-point plateaux (Figs. 2–4). The mean of the standard deviation for all of the melts and freezes was as follows: gallium–zinc eutectic = 0.08°C (14 melts); gallium = 0.08°C (12 melts); and ethylene carbonate = 0.07°C (8 freezes).

The length of the fixed-point melts/freezes that could be expected in normal in situ use was determined over a number of repeat cycles. These were performed with no

<table>
<thead>
<tr>
<th>Uncertainty component</th>
<th>Standard uncertainty uncertainty (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blackbody reference source</td>
<td>0.03</td>
</tr>
<tr>
<td>Alignment—reference source</td>
<td>0.06</td>
</tr>
<tr>
<td>Alignment—fixed-point source</td>
<td>0.11</td>
</tr>
<tr>
<td>Stability—reference source</td>
<td>0.00</td>
</tr>
<tr>
<td>Stability—fixed-point source</td>
<td>0.06</td>
</tr>
<tr>
<td>DVM accuracy (measuring C300 output)</td>
<td>0.00</td>
</tr>
<tr>
<td>SSE reproducibility (C300)</td>
<td>0.02</td>
</tr>
<tr>
<td>Residual SSE (C300)</td>
<td>0.13</td>
</tr>
<tr>
<td>Resolution (C300)</td>
<td>0.01</td>
</tr>
<tr>
<td>Repeatability</td>
<td>0.04</td>
</tr>
<tr>
<td>Emissivity of fixed point</td>
<td>0.08</td>
</tr>
<tr>
<td>Combined uncertainty</td>
<td>0.22</td>
</tr>
<tr>
<td>Expanded uncertainty ((k = 2))</td>
<td>0.44</td>
</tr>
</tbody>
</table>
insulation and in typical environmental conditions of ambient humidity and temperature. These results could, therefore, be thought of as worst case. The typical melt plateaux were: gallium–zinc eutectic = 3–4 h; gallium = 3–4 h; and ethylene carbonate = 1–2 h.

The repeatability of the fixed-point cells was determined from the standard deviation of the mean for a number of precise temperatures determined from plateau melts/freezes: gallium–zinc eutectic = 0.11°C (7 melts); gallium = 0.09°C (5 melts); and ethylene carbonate = 0.06°C (5 freezes).
The clinical centers carrying out the field trial measurements reported that the fixed-point cell temperature performance was found to be satisfactory within the performance of the measurement instruments. The longevity of the plateaux for each fixed point was measured both with and without insulation, and the time periods were found to be nominally 2 and 4 h, respectively. It was commented that, in a clinical thermography environment, a 2-h plateau would probably not be sufficient, especially in consideration of the considerable time spent on fixed-point cell preparation—an automated system should help to remove this issue.

The fixed-point sources were used as validation sources in cold challenge hand tests with some success. It was decided that a specific mount would have to be constructed to enable the hands and the cells to be aligned on the same focal plane. The main points of criticism were the un-automated fixed-point cell preparation, the lack of an alignment mount for the fixed points, and the aperture size could potentially require enlargement.

4 Conclusions

Three fixed-point cells for the in-field-of-view calibration of thermal imagers, for use in medical applications, have been successfully constructed, tested, and validated. They successfully provide a means of calibrating thermal images by being operated in the field of view during measurement of a subject. The final uncertainty determination of the fixed-point cells was higher than expected, but the largest uncertainty components come from the transfer radiation thermometer used. Further work will be carried out with a transfer radiation thermometer having better target definition, temperature resolution, and stability, and it is envisaged that the uncertainties will be significantly reduced, maybe by as much as a factor of two.
Field trials of the sources have been successfully carried out. The feedback from the clinicians was largely positive concerning the capability of the fixed-point sources. There were no technical problems with the performance capability of the fixed-point sources; the plateau length, temperatures, and uncertainties were all fit-for-purpose. Suggestions were made concerning the fixed-point initiation and control systems. The fixed-point system was a proof-of-concept; therefore, the initiation and control systems were simple and were not optimized as a commercial system. The field trial results will be reported in more detail in a future paper to be written by the authors.


References

Appendix 13

Statement of Support

I confirm the following contributions from Kevin Howell to the publication below:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Howell KJ, Smith RE. Guidelines for specifying and testing a thermal camera for medical applications. Thermology International 2009;19:5-12</td>
<td>• Preparation of manuscript (70%)</td>
</tr>
</tbody>
</table>

Dr RE SMITH

Head of Medical Electronics

Royal Free Hospital

London

24th August 2009
Guidelines for specifying and testing a thermal camera for medical applications

Kevin J Howell, Roy E Smith*
Departments of Rheumatology and *Medical Electronics, Royal Free Hospital, London. NW3 2QG. UK

SUMMARY
An infrared thermal camera used for medical thermography operates as a Medical Device, but few thermal imagers on the market meet the requirements of the EU Medical Devices Directive. The onus is therefore on the purchaser to ensure that the risks arising from the use of thermography in a medical context are identified and minimised.

Herein we present guidelines for the procurement and operation of an infrared thermal imager for medical thermography. Camera specifications, whole-lifetime costs, risk assessment, ongoing quality assurance and service/maintenance are among the issues addressed.

We encourage the reader to adopt an analytical approach to the procurement of medical thermographic equipment, in order to ensure safe practices and reliable measurements.

KEY WORDS: infrared thermal camera, medical thermography, risk assessment, quality assurance

ANLEITUNG ZUR SPEZIFIZIERUNG UND AUSWAHL EINER WÄRMEKAMERA ZUM EINSATZ IN DER MEDIZIN
Eine Infrarot-Kamera, die zur medizinischen Thermographie verwendet wird, ist als Medizin-Gerät zu betrachten, auch wenn nur wenige der am Markt zur Verfügung stehenden Wärmekameras die Voraussetzungen der EU-Direktive für Medizin-Geräte erfüllen. Daher liegt der Nachweis und die Verringerung des Risikos, das beim medizinischen Einsatz der Thermographie entsteht, beim Käufer und Anwender dieser Technologie.

Hier wird eine Anleitung für die Beschaffung und den Betrieb einer Infrarotkamera für die medizinische Thermographie gegeben. Die Spezifikationen, die Kosten für die Lebensdauer der Kamera, Risikoeinschätzung, kontinuierliche Qualitätssicherung sowie Service und Instandhaltung werden besprochen.

Die Leser werden ermutigt, bei der Beschaffung von Gerätschaften zur medizinischen Thermographie analytisch vorzugehen, um einerseits sowohl eine sichere Anwendung der Technologie als auch zuverlässige Messungen zu gewährleisten.

SCHLÜSSELWÖRTER: Infrarotkamera, Medizinische Thermographie, Risikoeinschätzung, Qualitätssicherung

Thermology international 2009: 19: 5-14

Background
Medical infrared thermography (MIRT) has a history going back some 40 years [1]. At that time there was limited regulation of medical technology. Today, in the ionizing imaging modalities, image optimisation in the face of dose reduction has driven the need for Quality Assurance (QA) checks. The Medical Device Directive, issued by the European Union, has introduced the concept of the "medical device" [2]. The UK Medicines and Healthcare Products Regulatory Agency (MHRA) defines a "medical device" as an instrument used for the purpose of, inter alia:

- diagnosis or monitoring of disease, injury or impairment
- investigation of the anatomy or of a physiological process.

Therefore, for medical thermographic applications, in both clinical and research settings, an infrared thermal imager unquestionably operates as a "medical device."

Healthcare institutions must work to clearly defined policies for the procurement and management of medical devices. Such policies are designed to ensure that

- correctly specified equipment is operated appropriately by competent staff, and
- an ongoing programme of funded equipment maintenance and QA is in place to ensure the equipment continues to perform within specification.

A properly implemented medical equipment management policy is a tool to ensure that risks to the patient and the health care institution are minimised. Risks can arise from the consequences of equipment failure (e.g. non-availability of the device in clinic), from misdiagnosis (because the device is not operating properly) or from some other adverse event (e.g. trips, falls or electrical fault hazards) [3]. Cogent procurement and management will ensure that, over the life of the equipment, value-for-money is obtained.

MIRT is still in transition. After a period of decline - due to exaggerated claims and disappointments - the introduction of reliable and sensitive focal plane array (FPA) imagers has led to resurgence in the numbers of cameras being procured for medical applications. Despite this increase in in-
terest, no guidance is yet published relating specifically to the procurement and equipment management of imagers for medical thermography. The potential therefore exists for exposure to risk from inappropriately procured thermal imaging equipment, operated without any QA plan.

Herein we describe the basic principles of procurement and equipment management pertaining to MIRT. Whilst local issues will inevitably necessitate amendments to our guidelines, our intention is to provide broad advice which provokes the initiation or refinement of medical equipment management at thermographic centres.

Procuring a thermal imager

In England, the MHRA document DB2006(05) [4] outlines the process by which medical devices should be procured and managed within the English National Health Service (NHS). Central to this document are principles of good financial governance, the concept of best value for money for NHS institutions (and how to win it), and an emphasis on costing equipment over its whole lifetime. A device with an inexpensive purchase price may prove extremely costly over its entire life if, for example, service and repair costs are not competitive.

Fig. 1 charts the life-cycle of a typical thermal imager in medical use. Below we will consider the key elements to this life-cycle in detail.

Figure 1
Typical life-cycle of a medical device

Identifying a clinical need

The provision of any new clinical service or research application starts with a number of decisions prior to the actual procurement of medical equipment:
- “What are the diagnostic needs of our clinical service?”
- “What equipment do we need to procure in order to fill that diagnostic need?”
- “In what environment should the equipment be operated, and by whom?”

In the case of thermography, meticulous research is therefore required at the outset to ensure infrared imaging is an appropriate modality for assessment of the clinical question in hand. It may be necessary to perform thermography in an accurately temperature-controlled room. Trained staff who can dedicate appropriate time to this exacting discipline will need to be identified. A number of educational courses in MIRT are available, which will add specific competence in thermographic techniques to the core knowledge of physicians or clinical scientists.

Specifying the thermal imager and talking to suppliers FPA technology has revolutionised the thermal imaging market, bringing to the customer an extensive choice of high-performance, competitively-priced infrared cameras. The equipment available can now meet the most demand-
ing medical thermographic needs. Identifying the imager most appropriate for those needs, however, can be difficult.

The vast majority of thermal imagers sold worldwide are destined for use in industrial, security or defence applications. Medical thermography remains a niche market. A key problem for healthcare customers is that biomedical know-
lledge amongst thermal imager suppliers is, at best, extremely limited. The specification of thermal imagers, and the presentation of their sales material, is normally also geared to non-biomedical markets. A good sales team will work closely with the customer to identify the product most appropriate to their need. The ultimate responsibility for ensuring the equipment is properly specified for medical use, however, will lie with the purchasers. They should have a basic understanding of some of the key thermal imager specifications, as discussed below

Detector type

Modern FPA detectors fall into two categories: "cooled" and "uncooled" [5]. Choice of detector is the first decision facing the customer.

Cooled detectors are typically arrays of photon sensors, whereby incident IR radiation causes electrons to move across the energy gap of a doped semiconductor material. Since this process will occur at any temperature above absolute zero, it is necessary to cool the detector to reduce inherent noise in the image. Compact and robust mechanical coolers for thermal detectors are now readily available, and the use of liquid nitrogen to cool detectors is now not necessary. Nonetheless, mechanical cooling introduces a moving part into the imager, which is quite likely to be the earliest source of equipment failure. Mean time between failures (MTBF) of mechanical coolers is usually quoted at several thousand hours [6]. For "high-end" medical applications (e.g. measuring fast dynamic processes [7,8] or small temperature variations across one image), cooled detectors remain the technology of choice due to their high sensitivity and rapid response times.

Uncooled detectors are typically arrays of resistance micro-
 bolometers. There is a change in electrical resistance of the detector material in response to a temperature change elicited by incident IR radiation. Imager life is likely to be dictated by the life of the detector array itself. Uncooled detectors tend to have slower response times than their cooled counterparts, and a somewhat lower sensitivity. For many biomedical applications, which involve slowly changing temperatures of the magnitude of a few tenths of a Kelvin or more, these limitations of uncooled detectors may not be important. More significant can be the drift in output of uncooled detectors in response to changes in ambient temperature, and the time taken for the detector to reach stability after switch-on (which can be minutes or even hours) [9]. The performance of uncooled detectors has improved in the last five years, however, and many imagers perform very adequately within minutes of switch-on in a stable temperature controlled room. Such data may not be available from the manufacturer, however, and if the user conducts their own investigations there is no guarantee that another "identical" machine will behave in the same way.

Thermal accuracy and sensitivity

The thermal sensitivity of a thermal imager reflects its ability to detect temperature differences between two parts of the same image, and is typically a fraction of one tenth of a Kelvin for modern FPA imagers (best for cooled systems). The thermal accuracy poses a bigger challenge in medical thermography. This is the absolute accuracy of the temperature reading, which must be measured against some accepted standard (ideally traceable to ITS-90, the international temperature scale of 1990) [10]. The quoted thermal accuracy of high-end imagers is often little better than budget models: typically ±2 K at ambient temperatures. This implies an inherent inaccuracy of perhaps 20% of the range of human skin temperature, and is a potential severe limitation to the application of thermography across multiple healthcare institutions.

Fortunately, this disappointing thermal accuracy arises from the limitations of factory calibration of thermal imagers. As we will see later, thermal accuracy can be greatly improved by the use of traceable, in-situ, calibration sources. A quoted thermal accuracy of ±2 K is therefore not necessarily a problem, although provision must be made for in-situ calibration of the thermal imager in use.

Optics, array size and resolution

The infrared lens, which forms the thermal image on the detector, is one of the most expensive components of a thermal imager. Many cameras accept interchangeable lenses for normal, "close-up," or "wide-angle" views, so it will be necessary to discuss which lenses might be required with the supplier.

In medical applications we would often like the spatial resolu-
tion of the imager to be as high as possible, implying that finer detail in the infrared image can be detected. Since each detector element in the focal plane array is responsible for a pixel in the output image, resolution and array size are related. Standard array size is typically 320 x 240 pixels for medical use, but smaller arrays may be useable for some applications. A new generation of larger arrays are now coming onto the market, with some imagers now incorporating up to 640 x 512 pixel arrays.

Despite the relationship between array size and resolution, differences in the quality of the array, optics and imager processing firmware mean that resolution will vary between imagers with the same array size. Hence the only way to reliably assess imager resolution is to observe the minimum resolvable detail in the image. Fig. 2 shows a desktop test array developed for this purpose by Prof. Francis Ring and co-workers at the Royal National Hospital for Rheumatic Diseases in Bath, UK [11]. The array consists of a pattern of vertical and horizontal heated bars spaced apart at a variety of distances, to allow comparison of the minimum resolvable detail of different imagers in both the x and y planes. When viewing this test array, any image distortion due to the lens optics is also clearly visible.

Image capture and analysis software

Medical thermal imagers only rarely are used as standalone devices. More commonly, the image output is captured to a
computer where appropriate image processing, reporting and storage take place. Thermal imaging software is therefore a key element of the medical device package.

The suitability of imaging software can only really be assessed by using it, so it is important to ask the supplier for a demonstration. If the customer’s own computer is to be used to run the software, the demonstration should take place on that machine to ensure that there will be no software or hardware compatibility issues. It is often helpful to make a list of requirements that the software will need to fulfil specific to the clinical application. Some more generic questions to ask might be:

- Can the software capture timed sequences of images, or just single frames?
- What analysis tools are available? Can regions of interest relevant to medicine be drawn for the extraction of temperature data?
- Are the values extracted by the software accurate? Has the program been written under a quality system? The quality of images, and accuracy of temperatures extracted, is dependent on the integrity of the software code. Regular calibration and QA of the thermal imager will be completely undermined if the software compromises the data output from the imager.
- Are the tools for reporting of images adequate? Can intuitive colour palettes be applied to images, and are they exportable to word-processing, graphics, or spreadsheet packages?
- How will the images and reports be made available? Is PACS an option? Are images compatible with the DICOM standard [12]?
- How are images stored, tracked, encrypted and backed up [13]?

Whole lifetime costs of a thermal imager

The purchase process will depend on the institution purchasing the thermal imager. For many research applications, the purchaser will be free to choose the imager of choice from a number of quotes, subject to the constraints of budgets and the procurement rules of grant awarding bodies etc. For clinical applications within organisations like the NHS, the preferred procurement method is by competitive tender, where bids to supply equipment that meets a defined specification are submitted by suppliers, and the winning bid is that which represents best overall “value for money,” as judged by pre-defined criteria. The competitive tender process has been shown to provide best “value for money” for NHS institutions, but it is not always well-suited to procuring unusual or “one-off” items like thermal imagers. Many of the criteria that constitute “value for money” for the device will not be previously defined, and will need to be identified carefully before the tender process can proceed. Misjudgements in this process can lead to clearly inferior equipment achieving the winning bid.

Assessing the “whole lifetime” cost of a medical device is essential for the good financial governance of the clinical service the device assists in providing. Institutions need to budget not only for the purchase cost of a thermal imager (including sales tax, where applicable), but also the ongoing costs of operating the device throughout its projected life. For thermography this is likely to include annual service and calibration of the imager. A decision also needs to be made regarding the financial arrangements for coping with unexpected breakdown of the equipment. Most suppliers are willing to place thermal imagers under service contracts. Otherwise, the user will need to cope with the cost of repairs as and when unexpected equipment failure occurs. Returning the imager to the supplier for planned maintenance or remedial repairs after a breakdown may incur less tangible costs that should also be included in the budget. This includes the loss of revenue from thermographic procedures that cannot be performed while the equipment is unavailable.

An accurate assessment of whole lifetime cost of a thermal imager is dependent on good estimates for the likely life of the device. For cooled detectors, the cooler life is likely to be the limiting factor, and this is normally given in manufacturers’ specifications. For uncooled detectors, where detector life is probably the limiting factor, an accurate prediction of lifetime is more problematic. Detector arrays are notoriously variable in their longevity, and stories abound in the medical thermography profession of users with identical models of FPA imager who have achieved greatly differing imager lifetimes. As with much integrated micro-circuitry, this is probably a shortcoming in the consistency of the detector manufacturing process. Uncooled detector MTBF is rarely quoted by manufacturers, but 30,000 hours is not untypical [6] (this is a summary statistic and provides
Risk assessment
Prior to the purchase of the thermal imager, a thorough initial assessment of the risks associated with the use of the imager should be carried out. Thermal imagers are not normally sold as “medical devices” and are therefore not CE marked as such to MDD 93/42 EEC. In the eyes of European law, therefore, thermal imagers are not proven fit and safe for medical use, and the onus is on the thermographer to procure and operate the device in a manner which minimises risk to the patient.

Risk can be defined as:
(consequence of an event) x (probability that the event will occur).

Table 2 shows one approach [3] to risk assessment, with the consequences of an event scaled from 1 to 5 (1 = negligible consequence, 5 = multiple fatalities), and the probability of an event also scaled from 1 to 5 (1 = impossible, 5 = certain). Risk scores of 15 or above can be considered “high risk,” while conversely scores of 3 or less pose “low risk” or “no risk.” This is often applied when analysing adverse incidents, however, it is equally applicable in prospective risk assessment.

At the procurement stage the major risk to the organisation relates to the non-compliance to the MDD. As there are very few thermal cameras meeting the MDD on the market, the actions arising from the initial risk assessment relate to management processes the user can implement to best overcome this deficiency.

Acceptance of the new medical thermal imager
All the preliminary work of costing, specification and risk assessment is of course performed with the aim of delivering to the institution an appropriate thermal imager which will pass an acceptance test, and move smoothly and expeditiously into clinical use.

Acceptance test
Immediately after delivery, the imager and any attached electronic components should undergo an appropriate acceptance test as outlined in DB2006(05) [4]. This typically consists of:
- Paperwork and asset logging: has all the correct documentation, such as instruction manuals, been supplied with the imager? Add the equipment to the institution’s database of medical equipment, and label the device with its asset number.
- Visual inspection: any apparent damage to the casing, lenses or connectors?
- Functional check: does the device function in line with the supplier’s information?
- Basic electrical safety: e.g. continuity of earth connection, leakage current (inter alia), consistent with IEC601 – the standard for medical electrical equipment [14]. Computer power supplies can have earth leakage currents that far exceed the limits of IEC601.
- An assessment of the instrument (and its stand) under PUWER (Provision and use of work equipment regulations in the UK – from EU directive 89/655/EEC) (15)

Operational risk assessment
Now is the moment to perform a repeat risk assessment, this time with the device operating in situ, and bearing in

---

**Table 1**
Example costs for procuring and operating an uncooled thermal camera in the first year.

<table>
<thead>
<tr>
<th>Description</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermal camera</td>
<td>€30,000</td>
</tr>
<tr>
<td>Test equipment, camera stand and isolation transformer</td>
<td>€8,000</td>
</tr>
<tr>
<td>Annual service</td>
<td>€1,500</td>
</tr>
<tr>
<td>Annual downtime (QA and service)</td>
<td>€10,000</td>
</tr>
<tr>
<td><strong>Total for first year</strong></td>
<td><strong>€49,500</strong></td>
</tr>
</tbody>
</table>

---

**Table 2**
Qualitative risk assessment matrix, adapted from [3]

**QUALITATIVE RISK ASSESSMENT MATRIX – LEVEL OF RISK**

<table>
<thead>
<tr>
<th>CONSEQUENCES</th>
<th>Impossible</th>
<th>Rare</th>
<th>Unlikely</th>
<th>Moderate</th>
<th>Likely</th>
<th>Certain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negligible – 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Minor – 1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Serious – 2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Major – 3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>Fatality – 4</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>Multiple Fatalities – 5</td>
<td>0</td>
<td>4</td>
<td>8</td>
<td>12</td>
<td>16</td>
<td>20</td>
</tr>
</tbody>
</table>

KEY: [ ] No risk □ Low risk □ Moderate risk □ Significant risk □ High risk
mind the findings of the acceptance test. Infrared thermography, as a non-contact, non-ionising imaging technique poses only very minimal physical risk to the patient per se. However, other risks to the patient and staff (operator) may arise. These can include:

- Electrical safety hazards (e.g. risk to cardiac patients who come into physical contact with the camera or stand, if the equipment is poorly electrically isolated). An isolation transformer may be appropriate depending on the result of an electrical safety test at acceptance.
- Trip and fall hazards. (Patients and staff can trip over trailing wires or other lab equipment, camera stands can be knocked over, imagers can fall from their stands). Consider a unipod medical imaging stand for the thermal imager, rather than the unstable and inflexible tripods often offered with thermal imagers. Cold water challenges may allow water and electricity to mix with explosive consequences.
- Inappropriate arrangement of equipment in the lab. The layout with the imager present may be ergonomically unacceptable for patients and staff.
- Risk of wrong diagnosis. (Erroneous and misleading temperature readings arising from poor imager calibration or QA). Without a careful programme of imager quality assurance, this is undoubtedly the greatest risk to the patient posed by thermography.
- Long hours in front of a poorly-configured visual display or badly designed workstation introduce further hazards for the operator (Display screen regulations - Directive 90/270/EEC) [16].

Thus risk assessment plays a vital role in fulfilling the obligations of healthcare institutions to comply with health and safety regulations that protect patients and workers. A formal risk assessment is shown in appendix 1.

Calibration and quality assurance

The thermal imager supplier undertakes to supply equipment that performs within specification. As discussed above, however, that specification is unlikely to be rigorous enough to satisfy the requirements of medical thermography. Quoted temperature accuracy is typically as poor as ±2 K, and there is drift in readings after initial switch-on of uncooled systems. The typical accuracy specified by many suppliers arises because most imagers are used in industrial settings where the dynamic range is broad, ambient temperature may vary greatly, or the imager may be expected to be used immediately after switch-on. Fortunately most of these limitations can be overcome by operating the imager within careful constraints (e.g. stable ambient temperature, not imaging until the detector reaches stability), but it is vital the performance of a newly-procured imager is fully investigated at the outset to enable an appropriate operating protocol to be adopted. Furthermore, these investigations should be regularly repeated as part of an ongoing process of imager quality assurance: camera performance can and does change with use. Where the imager no longer performs within the medical specification required, remedial action will be required.

Temperature calibration

Some form of in situ device which can provide a known traceable standard temperature is essential to validate the absolute accuracy of the thermal imager. This normally takes the form of a black-body source i.e. a target set of targets which can be held at an appropriate temperature [17]. The radiant emissivity of the target(s), ε must be known very precisely, and is typically as near to unity as design will allow. The temperature output of the thermal imager can then be compared against a range of target temperatures to produce a calibration curve for the thermographic device. Ideally, a source should also be included within the field of view during the imaging of the patient, in order to validate the temperature readings within each specific image.

Many of the black body sources commercially available are not appropriate for medical use because they only provide calibration across an inappropriate temperature range. Producing a reliable, cost-effective black body source which will operate at near-ambient temperatures - i.e. in the medical range of approximately 293 – 313 K (20°C – 40°C) - with an accuracy of better than ± 0.1 K is technically quite challenging. However, a small number of devices which meet this requirement are available. An ice-water mixture provides an approximation of a black body source, but requires the camera to be mounted pointing downwards and is outside the medical range. It is nonetheless the basis for some QA (see below).

In the UK, the National Physical Laboratory has developed a Thermal Imager Validation System (TIVS) [18] which is ideally suited to the calibration of medical thermal imagers. TIVS consists of portable black body sources utilising either the solid-liquid phase transition or the eutectic temperature of eutectics. They retain highly stable temperatures at their respective phase transitions for prolonged periods. The use of more than one source facilitates the validation of the imager output at more than one temperature point within the medical temperature range. The sources are small and robust enough to be included within the imaging field of most medical images, thus providing a true in situ validation of each thermal image during thermographic procedures.

The accuracy one is looking to achieve from a thermal imager will depend on the particular medical application, and the imager purchased. In a steady ambient temperature, even an uncooled modern FPA imager which has stabilised after switch-on should be achieving an accuracy of a very few tenths of a degree Celsius, and this is sufficient for most medical applications.

Other quality assurance tests

Plassmann et al [9] describe a number of further quality assurance tests that should be performed on thermal imagers immediately after procurement, and also regularly on an ongoing basis while the imager is in medical service. These tests are also outlined on the World Wide Web at www.medimaging.org The tests can be performed using ice and water to give a basic degree of confidence in imager performance, but the use of an appropriate black body source as the temperature reference is preferable for some of the tests.
- Offset drift after switching on. Uncooled imagers can exhibit a drift in temperature reading of several degrees Celsius over hours, prior to the detector reaching acceptable stability. This is vital to investigate: it dictates how long the user must wait before the imager is performing optimally, and the findings should be built into the operating protocol for thermographic studies.

- Long term offset drift. Provided ambient temperature remains constant, and the detector has reached stability after switch-on, the thermal imager should give a reproduceable measurement of black body target temperature over different days. Variations of more than a few tenths of a degree might invalidate longitudinal thermographic studies.

- Offset variation over range. This is a check to ensure the imager is accurate at all temperatures across the medical range, and is particularly easy to perform using the TIVS system.

- Thermal flooding. This investigates the effect on measured target temperature of introducing a second target at a different temperature into the field of view. The effect should be minimal, but the performance of FPA imagers varies in this respect.

Service and maintenance by the supplier

The final aspect of thermal imager equipment management is the maintenance service offered by the supplier. This falls into three categories: ad hoc repair; maintenance contract; and planned preventative maintenance and calibration. Before entering into any contract it is important to understand what is being offered. Once again a risk based approach is sensible. How often does a camera fail? What is the consequence for the clinical service (loss of revenue, rescheduling appointments)? What is the claimed response time of the company? And what do other users say is the reality? Where is the service carried out? Is the workshop accredited? Where are spares held? Is it better to put the cost of a maintenance contract towards a second camera? How is the imager calibrated against a black body source during annual service? Is the calibration across a sensible range of temperatures for medical applications? If not, it may be an expensive exercise which gives less validity to medical measurements than inexpensive in situ calibrating equipment tailored to the medical temperature range. Is the calibration traceable to ITS-90?

Sending an imager to the supplier for routine maintenance potentially creates a gap in the clinical thermography service. Few medical thermography centres have the luxury of owning more than one thermal imager, but this is the preferred situation since the second imager, with known performance characteristics, can take on some of the clinical load in the absence of the first. It also offers your clinical service resilience - in a financial model the cost of a camera may be small compared with other fixed overheads.

Some suppliers can offer a loan instrument during the service period. However, since the performance characteristics described in section 3 above will need to be known about the loan device to validate any clinical work, this is not always necessarily an acceptable solution. It cannot be assumed that even identical models of thermal imager have identical performance characteristics.

Unlike the validation of the thermal imager in situ against a black body source, the black body calibration which occurs at the service centre will result in adjustments being made to the imager firmware which permanently alter its output. Hence, the calibration must be certified by the supplier i.e. documentation should be supplied which details the changes that were made to the instrument, and accredited - i.e. those changes are traceable and standardised. In this way, the user should be able to account for any changes in the imager performance that occur on its return from manufacturer's service. Time must be set aside for re-assessing the imager using the in-house protocol before it returns to clinical service.

Discussion

Inexpensive, high-performance focal plane array thermal imagers have presented the medical thermal imaging community with an excellent opportunity over the last decade. For the first time, high quality thermographic measurements are attainable for medical applications by almost any practitioner. Ongoing improvements in imager technology mean this is an exciting time for medical infrared thermographers [19].

With this technologically-driven opportunity, however, have come threats to the discipline of infrared thermography. As infrared equipment becomes easier to acquire at relatively low cost, the likelihood it will be misused by the uninitiated increases. Thermal imagers procured for medical use should not be considered straightforward pieces of consumer electronic equipment, but rather specialised physiological measurement tools which must be subject to rigorous quality assurance protocols. Provided the highest standards of application of the technology prevail, the benefits will be seen to have greatly outweighed the risks.

In this paper, we have limited our discussion to technical performance of the imager. It should not be forgotten by the reader that other opportunities to misapply thermography in medicine exist. In particular, the importance of applying rigorous, reproducible patient preparation and image capture protocols has been clearly demonstrated [20, 21]. Analysis of thermal images with consistent, appropriate statistical tools and "regions of interest" is also vital.

Arguably the biggest risk to both patient and user from thermography, however, remains the misuse of the thermal imager through ignorance of its technical performance. This risk will only be minimised by education of thermographers (and potential thermographers) about the technology "inside the camera case." In this, suppliers could potentially play an important role. The medical community must in turn communicate its requirements from infrared technology to suppliers. In particular, there needs to be rapid adoption of in situ methods of thermal imager validation which are appropriate to the medical temperature range. Without this, meaningful multi-centre trials utilising thermographic measurement are impossible - a limitation which has invalidated major studies in the past [22].
In conclusion, we would encourage the reader to adopt a thorough and analytical approach to the procurement of a medical thermal imager. Consider the whole-life cost of your purchase to ensure value for money, and ascertain the exact specification of infrared imager you will require. A careful risk assessment should be performed, and the performance of the imager should be investigated both at the commissioning stage, and on an ongoing basis throughout the imager lifetime. The supplier can also contribute to imager quality assurance, but talk with your service centre about how they can offer maintenance that is really applicable to medical measurements.

These efforts, which are analogous to little more than the precautions we might take in keeping a motor vehicle in roadworthy condition, will ensure years of reliable infrared measurements for your thermographic service.

References

1. Ring EFJ. The historical development of thermometry and thermal imaging in medicine. Thermology International 2003; 13:33-57


11. Ring EFJ, Dicks JM. Spatial resolution of new thermal imaging systems. Thermology International 1999;9:7-14


14. IEC 60601-1-1 Ed. 2.0 b:2000, Medical electrical equipment - Part 1-1: General requirements for safety - Collateral standard: Safety requirements for medical electrical systems. IEC TC/SC 62A


20. Ammer K, Ring EF. Repeatability of the standard view both dorsal hands. Results from a training course on medical infrared imaging. Thermology International 2004;14:99-102


Address for Correspondence

Kevin J Howell, Clinical Scientist
Royal Free Hospital, Department of Rheumatology
Pond Street London. NW3 2QG. UK
Tel. +44(0)2074726550, email. k.howell@medsch.ucl.ac.uk

(Received on 15.11.2008, accepted on 10.1.2009)
Appendix 1. Example risk assessment for a thermal imager in medical use.

<table>
<thead>
<tr>
<th>Department/Area:</th>
<th>Date of Assessment:</th>
<th>Assessment No:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microvascular Laboratory</td>
<td>5/3/2008</td>
<td>12345</td>
</tr>
<tr>
<td>Rheumatology Department</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**DESCRIPTION OF ACTIVITY**

**USE OF AN FPA THERMAL CAMERA TO PRODUCE THERMAL IMAGES OF PATIENTS**

Patients – both adults and children - will be imaged. They will be asked to disrobe as required, and to equilibrate to ambient in the room with the camera. They may be imaged standing or sitting. The images taken will then be used to support clinical diagnosis. They might undergo a cold challenge of the hands with water at 15 °C. All equipment should be used according to manufacturer’s instructions.

**HAZARD IDENTIFICATION**

1. Inappropriate or incorrect use
2. Failure to set up the equipment safely for the working environment
3. Imaging equipment is not classified as a medical device
4. Equipment does not comply with requirements of IEC601 for electrical safety of medical devices.
5. Failure to maintain the equipment
6. Failure to perform initial image QA
7. Failure to perform ongoing QA
8. Failure to plan for contingencies – including equipment failure and planned replacement.

**PERSONS AT RISK** (who might be harmed and how? are there any groups especially at risk?)

Patients requiring clinical service, staff, and visitors

1. Images obtained and their interpretation may be incorrect (Patient)
2. Trips, slips, and accidental contact with equipment (All)
3. Equipment might not be fit for medical purpose (Patient)
4. Exposure to electricity (All)
5. Equipment fails and no image is obtained (Patient)
6. No understanding of imager performance (Patient)
7. Change in image quality resulting in misdiagnosis (Patient)
8. Suspension of service (Staff, Patients)

**EXISTING PRECAUTIONS** (what controls are currently in place?)

1. Professional capacity of thermographer
2. Ergonomic design of laboratory – equipment away from patient flow. Patient and visitors not left unsupervised.
3. Recognition by thermographer and clinicians of limitations of equipment.
4. Protection against fluid ingress, isolation transformer if leakage currents excessive.
5. Annual assessment of equipment.
6. Acceptance QA
7. On-going QA
8. Contingency plan not in place

<table>
<thead>
<tr>
<th>other risk assessments</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ANALYSIS OF EXISTING PRECAUTIONS (are the precautions effective; if not, why not?)

1. Effective
2. Partially effective – environment is cramped
3. Effective
4. Isolation transformer; electrical items mounted above for level away from sink and cold challenge area
5. Effective.
6. Adequate.
7. Insufficient resource for proper QA programme.
8. Not sufficient

ACTION REQUIRED (what further action needs to be taken to control/reduce the risk?)

<table>
<thead>
<tr>
<th>Action</th>
<th>By whom</th>
<th>Target date</th>
<th>Date completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Draw-up business contingency plan</td>
<td>KH</td>
<td>01/06/2008</td>
<td></td>
</tr>
<tr>
<td>Write business case for QA support</td>
<td>KH</td>
<td>01/07/2008</td>
<td></td>
</tr>
<tr>
<td>Feed into area refurbishment</td>
<td>KH</td>
<td>31/12/2008</td>
<td></td>
</tr>
</tbody>
</table>

REVIEW DATE
General review – 31/12/2008

PERSON(S) CARRYING OUT THE ASSESSMENT
Kevin Howell -
Roy Smith – Head of Medical Electronics

Print: ___________________ Signature: ___________________ Date: ___________