Quality assurance for a multi-centre thermography study

EJ Marjanovic[1,2] , J Britton [3], KJ Howell [4], AK Murray[1,2,5]

[1] Centre for Musculoskeletal and Dermatological Research, University of Manchester, M13 9PL, UK
[2] Salford Royal Foundation, NHS Trust Manchester Academic Health Science Centre, Salford, M6 8HD, UK
; [3]Medical Physics Department, Old Medical School, Leeds General Infirmary, Leeds, LS1 3EX, UK
[4]Microvascular Diagnostics, UCL Institute of Immunity and Transplantation, Royal Free Hospital, London, NW3 2QG; UK
[5] Photon Science Institute, University of Manchester, M13 9PL, UK

SUMMARY

In this study we describe the quality assurance programme developed to evaluate the performance of six thermal imaging devices in the VALIDS study, a project set up by a United Kingdom (UK) group to investigate the reliability of thermography for the assessment of Raynaud's phenomenon secondary to systemic sclerosis. At each of the six centres the start-up drift, accuracy and repeatability of the thermal cameras was assessed by imaging an Isotech 988 blackbody source. All of the thermal cameras, except one (FLIR A35SC) had stable measurements within a 30 minute warm-up period. All of the thermal cameras, except the FLIR A35SC had accurate measurements within ±2°C of the blackbody temperature. Repeatability, as demonstrated by the calculation of the within-subject standard deviation, was less than 0.5°C for 5 of the 6 cameras; the exception being the FLIR Agema Thermovision 570 thermal camera. In conclusion the results confirmed that the thermal imaging device with less exacting accuracy specifications (FLIR A35SC) did not match the performance of the higher-specification devices, and it is questionable if it is fit for purpose when being used clinically. All the other devices demonstrated acceptable accuracy and repeatability for clinical use. These quality assurance methods, when employed along with rigorous patient preparation, image capture and analysis protocols, provide essential confidence in valid and reliable temperature measurements across multiple clinical centres.

KEY WORDS: Thermal camera, quality assurance, medical thermography

QUALITÄTSSICHERUNG EINER MULTIZENTRISCHEN THERMOGRAPHIE-STUDIE

In dieser Studie wird ein Qualitätssicherungsprogramm beschrieben, das für die Leistungsbewertung von sechs Wärmebildkameras in der VALIDS Studie entwickelt worden war. Dies ist ein Projekt, das von einer Gruppe im Vereinigten Königreichs (UK) gegründet wurde, um die Zuverlässigkeit der Thermographie für die Beurteilung des sekundären Raynaud-Phänomen bei systemischer Sklerose zu untersuchen. In jedem der sechs Zentren wurde mittels des Schwarzkörperstrahlers Isotech 988 die Stabilität der Messung nach dem Einschalten, sowie Genauigkeit und Wiederholbarkeit der Messung der Wärmebildkamera beurteilt. Alle Wärmekameras mit Ausnahme Einer (FLIR A35SC) boten innerhalb einer 30-minütigen Aufwärmphase stabile Messungen. Alle Wärmebildkameras mit Ausnahme der FLIR A35SC zeigten akkurate Messungen innerhalb von ± 2 °C der Temperatur des Schwarzkörpers. Die Standardabweichung von wiederholten Messungen wurde als Maß der Wiederholbarkeit verwendet und ihr Wert lag bei 5 der 6 Kameras unterhalb von 0,5°C; ausgenommen hiervon ist die Wärmekamera FLIR Agema Thermovision 570. Abschließend bestätigten die Ergebnisse, dass das Wärmebildgerät mit weniger exakten Spezifikationen (FLIR A35SC) nicht die Leistung der Geräte mit höherer Spezifikation erzielte, und es fraglich ist, ob es für eine Verwendung in der klinischen Temperaturmessung geeignet ist. Alle anderen Geräte zeigten eine für den klinischen Einsatz akzeptable Genauigkeit und Wiederholbarkeit der Temperaturmessung. Gemeinsam mit dem strengen Befolgen von Anleitungen für die Vorbereitung der Patienten sowie der Erfassung- und Analyse von Wärmebildern, fördern diese qualitätssichernden Methoden wesentlich das Vertrauen, dass Temperaturmessungen aus mehreren klinischen Zentren valide und zuverlässig sind.

SCHLÜSSELWÖRTER: Wärmebildkamera, Qualitätssicherung, Medizische Thermographie

Thermology international 2018, 28(1) 7-13

Introduction

Infrared thermography is a well-established imaging method for skin temperature measurement in biomedical research (1), but a requirement for better quality assurance of the technique has been identified (2). All temperature measurements should be traceable to the International Temperature Scale of 1990 (ITS-90) (3), and to ensure reliable outcomes the performance of thermal imagers should be regularly checked against an accredited reference standard. It is therefore important that all thermal imaging cameras used in clinical studies are subject to a rigorous and standardised quality assurance programme to ensure consistency of camera performance.

Howell and Smith (4) have described a protocol for the procurement and quality assurance of thermal imagers in medical use, and detailed standards for the application of thermography in fever screening have been published by the International Standards Organization (5).

Simpson et al (6) investigated the performance of several thermal imagers in routine medical use, finding considerable variation in camera stability and agreement with blackbody reference temperature. Uncontrolled, these differences in performance represent a potentially significant source of error in multi-centre studies where many thermal

imagers are used to collect temperature data across several sites.

The VALIDS study (7) was a project set up by a United Kingdom (UK) group to investigate the reliability of thermography with repeated hand cold challenge for the assessment of Raynaud's phenomenon secondary to systemic sclerosis across six UK tertiary referral centres. Herein we describe the quality assurance programme developed to ensure traceable temperature measurements at all sites, and present the performance results of the six different thermal imaging devices at each of the centres employed for the VALIDS data collection.

The aim of the study presented was to compare the performance of six thermal imaging cameras at different sites using a series of simple quality assurance measurements. Some of the measurements were also undertaken at the end of the VALIDS study to assess any changes in the operating parameters of the thermal cameras.

Method

Thermal cameras

Each of the thermal cameras put forward for the VALIDS study (7) was routinely used as part of the clinical service in each of the centres except for Centre 6, where the camera was loaned from FLIR (West Malling, UK). All the other thermal cameras were manufactured by FLIR and utilised uncooled focal plane array (FPA) sensors. The models of the cameras used were as follows; Centre 1- FLIR Agema Thermovision 570; Centre 2- FLIR SC305; Centre 3- FLIR A320G; Centre 4- FLIR A40; Centre 5- FLIR A305SC and Centre 6- FLIR A35SC. The spatial resolution of each of the cameras was 320*240 pixels, except for the FLIR A35SC which had a resolution of 320*256. The noise equivalent temperature difference (NETD) for each of the cameras was approximately 50mK. The accuracy range given by the manufacturer for the cameras in Centres 1-5 was ±2°C (or ±2%), whereas for Centre 6 (FLIR A35SC) it was stated as ±5°C (or ±5%). The receiver bandwidth of the detector for all cameras was 7 - 13 microns. Each thermal camera was fitted with a standard 25° x 18.8° lens, where the focus can be altered from 40-100cm from the front of the camera. ThermaCamResearcher software version 2.11, supplied by FLIR, was used to capture the images over a cross-over ethernet connection between a laptop and the thermal cameras at each of the centres.

Figure 1:
Picture of the set up used when imaging the blackbody source

Set-up

To examine thermal camera performance an Isotech 988 blackbody source (Isothermal Technology Limited, Pine Grove, Southport, Merseyside, England), which had been calibrated against a primary standard at the National Physical Laboratory (Hampton Rd, Teddington), was imaged at each of the six centres at the beginning and end of the study period. The measurements were performed with the blackbody positioned on a table approximately one metre from the thermal camera, with the room temperature set to approximately 23°C at each centre. The thermal camera emissivity was set to 0.98. A picture of the blackbody source being imaged by the thermal camera is shown in Figure 1.

Warm-up stability

Measurement of start up drift is important as it provides information on the amount of time required before a thermal imaging camera can be used optimally, and before clinical images can be acquired. To examine the stability of each imager during its warm-up period, the blackbody was set to 30°C and left to stabilise for 15 minutes. The thermal camera was then switched on, and images were acquired every five minutes for half an hour. An example of an image acquired by one of the cameras is shown in Figure 2. The warm-up data was only acquired at the beginning of the VALIDS study.

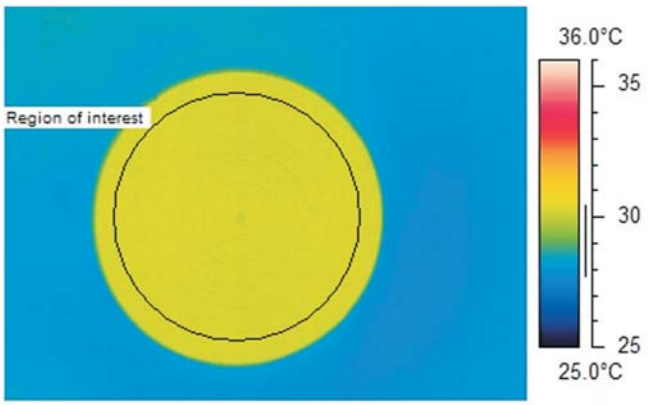


Figure 2:
Image of black body source acquired using a FLIR SC305 Thermal Imaging camera acquired at the start of the warm up stability test. The black body source was set to 30°C and the camera was measuring an average over the region of interest of 30.3°C. The region of interest made up 75% of the total area of the image of the black body source.

Accuracy and repeatability

To examine thermal camera accuracy and repeatability, the thermal camera and blackbody were switched on for a minimum of 30 minutes prior to image acquisition. The blackbody was then set to a temperature of 18°C and an image was acquired; the temperature of the blackbody was then increased in 2°C steps and further images acquired up to 40°C. These measurements were obtained before (baseline) and after or near the completion of the VALIDS study (follow-up) approximately 3 months later. In this experiment the blackbody source was taken as the reference temperature or "gold standard," and the performance accuracy of the cameras in measuring temperature was assessed against this device.

Analysis

Analysis of the blackbody images was performed using FLIR ThermaCam Researcher software version 2.11. For each image, a circular region of interest that encompassed approximately 75% of the area of the blackbody cavity was positioned in the centre of the image, and the mean pixel value was calculated.

The data was analysed using IBM SPSS Statistics version 22 (IBM Corporation, Armonk, N.Y. USA).

The repeatability of the thermal imaging measures was examined by calculating the within-subject-standard deviation of the data acquired from the blackbody source at all temperature points before (baseline) and after the completion of the VALIDS study (follow-up) approximately 3 months later. The within-subject standard deviation was calculated by dividing the sum of squares by its degrees of freedom to get the estimate of variance. The square root of this is the estimate of the within-subject standard deviation. This method is described in more detail by Bland & Altman in their paper on measurement error (8).

The agreement of the camera measurements against blackbody reading was examined by using Bland and Altman analysis (9). Bland and Altman analysis plots the mean of the thermal imaging camera and blackbody readings against the difference between the readings. The systematic "bias" is calculated as the mean difference of the measurements and the "limits of agreement" are given by ±1.96 standard deviations either side of the bias.

Results

Thermal camera warm-up

Figure 3 shows the temperature recorded by each thermal camera in 5 minute intervals from switch on. All of the thermal camera temperature measurements were within 2°C of the actual blackbody temperature (30°C), and all camera measurements tended towards 30°C by 30 minutes, with the exception of the measurements obtained by the camera in Centre 6. It was also observed that the temperature measurements for cameras in centres 3, 4 and 5 were not quite stable after 30 minutes, but the variations were less than 1°C. This is probably caused by small temporal and spatial temperature variations within the room, which are very difficult to control. However, collective opinion between the participating centres is that this is unlikely to have a major impact on the clinical results acquired.

Figure 3:
Temperature reading variation as the thermal cameras warm up from switch-on (blackbody source temperature 30°C)

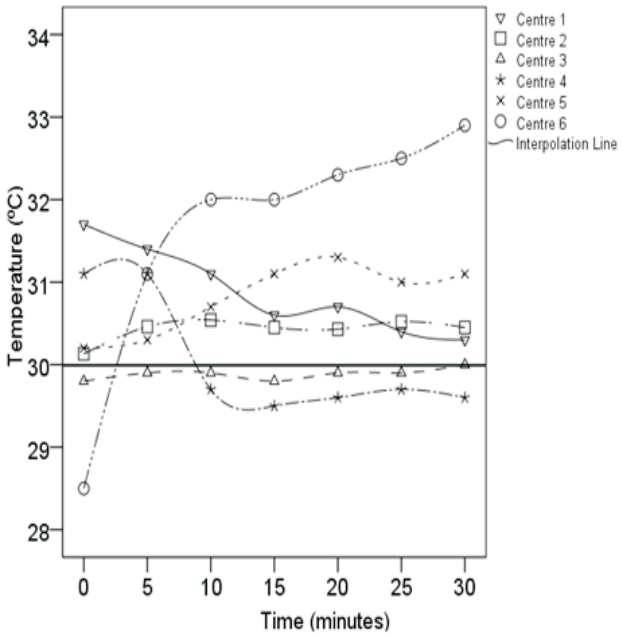


Table 1
Within-subject standard deviation for the baseline and follow-up thermal camera measurements acquired before and after the study period (across all blackbody temperature points).

Centre	Within-subject-standard deviation (°C)
1	0.71
2	0.22
3	0.06
4	0.34
5	0.12
6	0.22

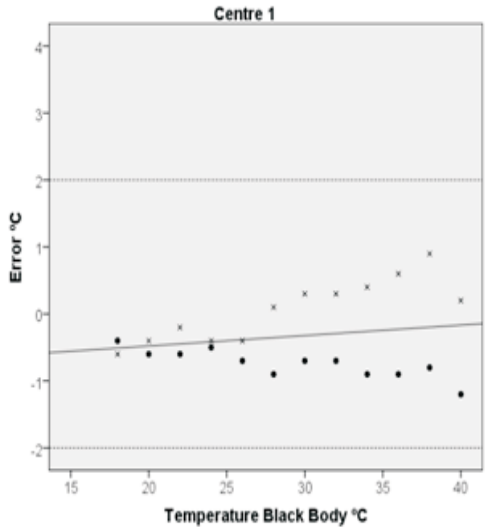
Accuracy and repeatability

Repeatability

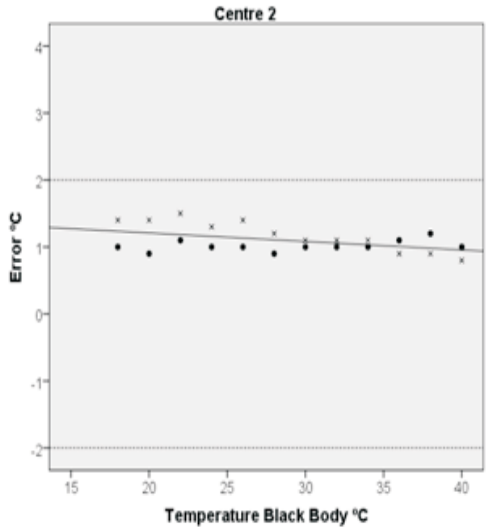
Figures 4a-f show the repeatability of each of the cameras in measuring the temperature of the blackbody source between 18°C and 40°C at baseline and at follow-up. Five of the six cameras behaved similarly at baseline and follow-up. At centre 1, however, the camera showed some variability in the readings taken above 25°C: there was a tendency to under-read the blackbody at baseline, and yet over-read it at

Figure 4a-f:
Measurement error (difference between thermal camera and blackbody reading) at baseline (•) and follow-up (x), plotted against blackbody reading. Maximum permissible error specified by the manufacturer (----) is also included (not shown for centre 6: ± 5°C)

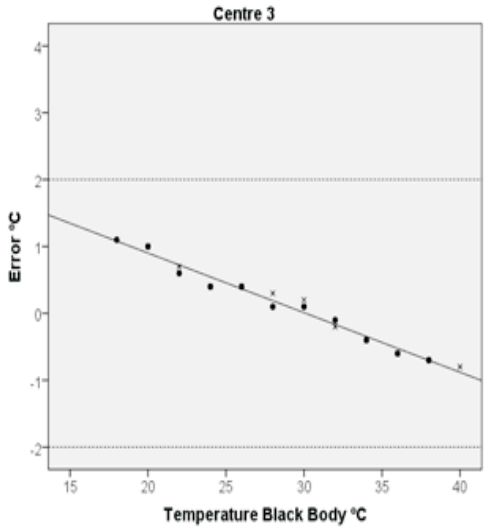
4a Measurement error acquired at centre 1



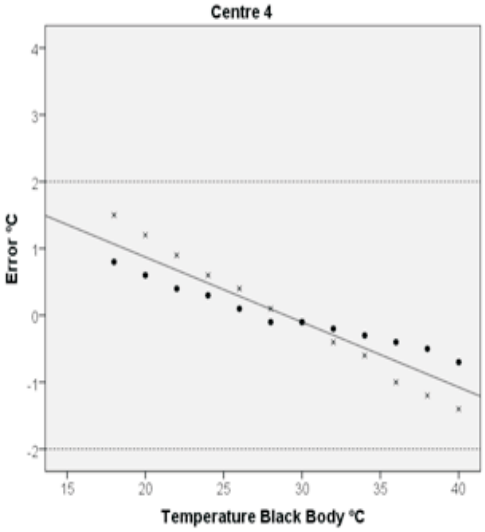
4b Measurement error acquired at centre 2



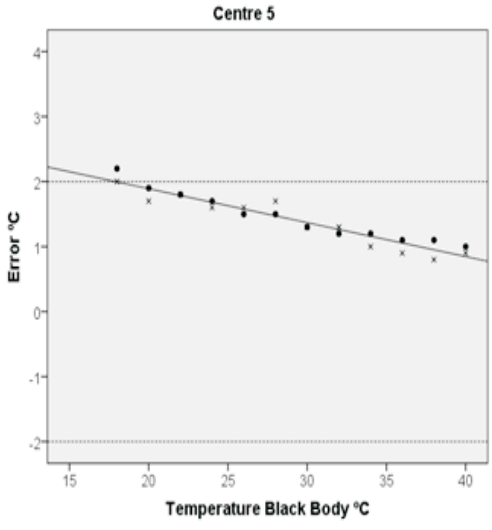
4c Measurement error acquired at centre 3



4d Measurement error acquired at centre 4



4e Measurement error acquired at centre 5



4f Measurement error acquired at centre 6

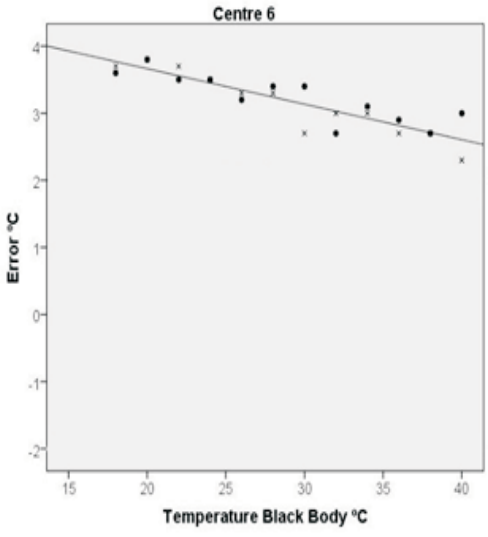


Figure 5
Bland and Altman plot of the mean of the thermal imaging camera and blackbody readings against the difference

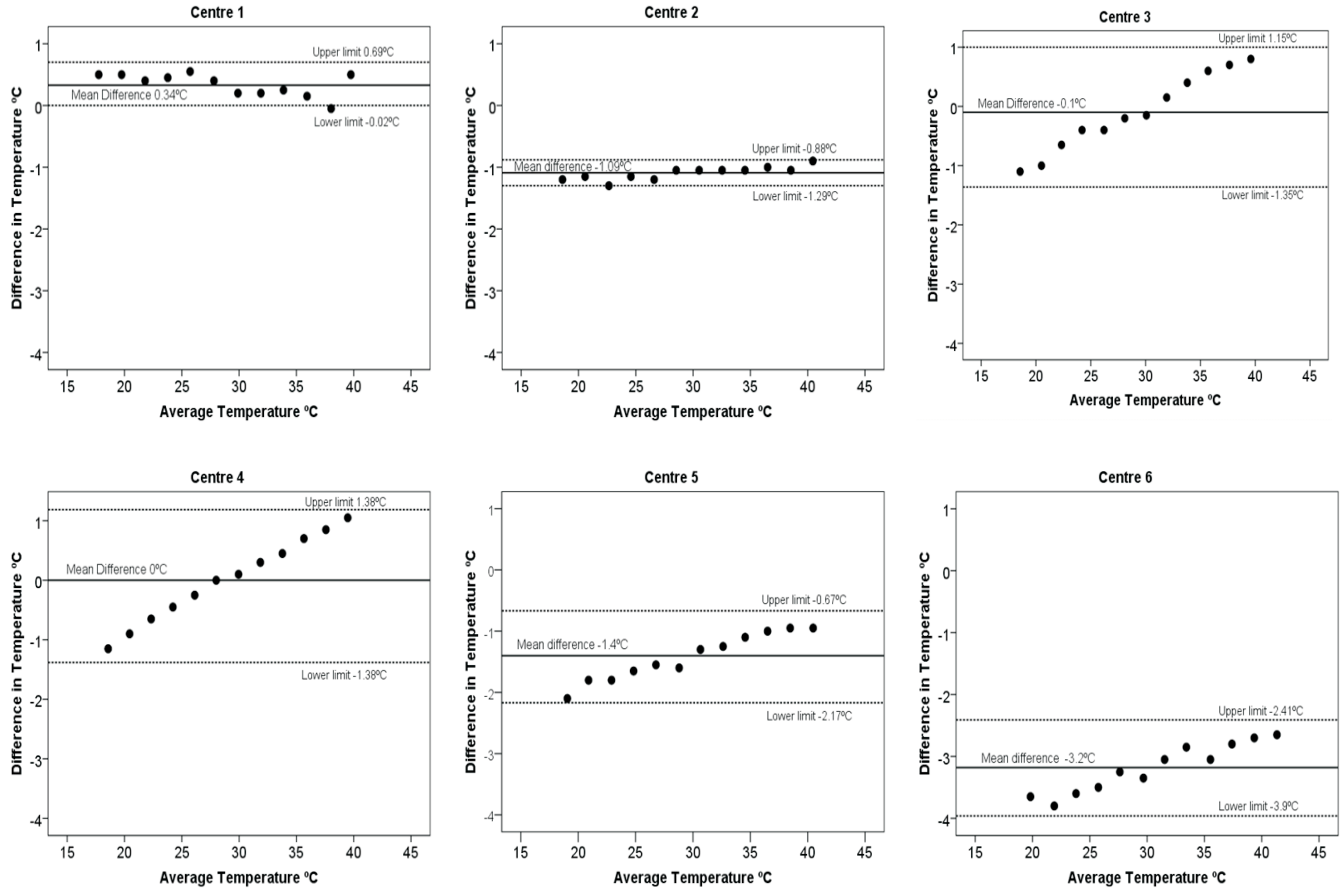


Table 2
Mean bias between blackbody and thermal camera

Centre	Mean bias (°C)	Limits of agreement (±1.96*SD (°C)
1	0.34	0.36 (-0.02 to 0.69)
2	-1.09	0.21 (-1.29 to -0.88)
3	-0.10	1.25 (-1.35 to 1.15)
4	0.00	1.38 (-1.38 to 1.38)
5	-1.42	0.75 (-2.17 to -0.67)
6	-3.18	0.77 (-3.95 to -2.41)

follow-up. Table 1 shows the repeatability (within-subject standard deviation) between the baseline and follow-up thermal imaging camera measurements.

Accuracy

Figure 5 shows the Bland and Altman plot of the mean of the thermal imaging camera and blackbody readings against the difference. For centres 1 and 2 the plots showed no relation between average temperature and the differences, but in the other four centres (which all employed FLIR "A" series cameras) the differences changed quite steadily with increasing average temperature. Table 2 shows the mean bias between the thermal camera and blackbody readings, with limits of agreements.

Discussion

In order to achieve reproducible results in thermal imaging investigations, one of the core requirements is to have equipment that is both reliable and repeatable. Up until quite recently cooled thermal imaging cameras were seen as the only option for acquiring reproducible high quality data. The cost of these devices meant that these were out of reach for most departments. However since the emergence of relatively low-cost, uncooled focal plane array thermal imager technology, a number of UK centres have now invested in such devices for investigating Raynaud's phenomenon, but there have always been questions about the reliability of the equipment being used. Validation of uncooled thermal imaging devices for assessing skin temperature therefore requires a traceable and simple-to-implement quality assurance protocol.

This is the first quality assurance study where a calibrated and externally verified blackbody source has been used to assess the performance of different thermal imaging cameras across a wide geographical region within the UK. In total six centres participated and results were acquired prior to, and following, a three-month clinical study (7) investigating reproducibility of a cold challenge in patients with Raynaud's phenomenon secondary to scleroderma.

The results from the warm-up tests showed that five of the six cameras were performing within the manufacturer's specifications of an accuracy of ±2°C when compared to a temperature setting of 30°C on the blackbody source. The thermal camera at Centre 6 was the only camera where the accuracy specification given by the manufacturer was ±5°C, and therefore this camera was within tolerance, even from switch-on. However, this camera showed by far the greatest variation in temperature reading during the warm-up period, and it was the only device where readings were continuing to change markedly at the end of the 30-minute evaluation. The thermal cameras at Centres 1-5 were all observed to be essentially stable at 30 minutes, and any systematic errors could be taken into account when analysing the results from clinical studies.

The repeatability experiment, examining the difference between the baseline and follow-up measurements, showed that for Centres 2 to 6 the within-subjects standard deviation was below 0.5°C. Only Centre 1 was outside this limit, and this may be related to the age of this thermal imaging camera (approximately 10 years old) whereas all the other devices were 6 years or younger. This is a reassuring result, as it confirms that there was insignificant drift in camera performance over the 3-month period of the VALIDS study, and therefore any limitations in clinical measurement repeatability could not be attributed to instrument factors.

The accuracy of the thermal imaging cameras as shown by Table 2 and the Bland and Altman plots shows that for Centres 1 to 5 the bias was well within ±2°C, which is acceptable for clinical imaging and within the manufacturers' specification. Indeed, amongst these five imagers, only one individual reading was found to be outside of the specification limit (for the camera at Centre 5 at a blackbody reading of 18°C). The thermal imaging camera at Centre 6 had a poorer bias of -3.2°C, which is however within the manufacturer's specification for this specific device. However, it is questionable if this is satisfactory for clinical use in a peripheral vascular setting, where the skin temperatures encountered only range across approximately 20°C.

All of the measurement errors were systematic and therefore could be taken into account and corrected when acquiring clinical imaging data. It is noted from the Bland Altman plots that the four FLIR cameras from the "A" range of devices all exhibited steadily rising biases across the range of blackbody temperatures studied, but with varying amounts of "offset" from the true blackbody value. The "SC" and "Agema" models showed a "flatter" response across the temperature range. These interesting findings probably reflect different approaches to the algorithm that calculates temperature from the detected infrared intensity signal within the collection of imagers evaluated.

A limitation of our Bland and Altman approach to analysing the agreement of the four "A" series cameras with the blackbody is that the mean of the differences is not constant throughout the temperature range: hence there is a likelihood that the overall limits of agreement are somewhat too wide for low temperatures, and too narrow for higher temperatures. A more robust approach for devices exhibiting varying means of the differences across the measurement range could be to analyse the logarithm of the measurements, or to find the 95% limits for the difference as a percentage of the mean (10).

Conclusion

In conclusion, we have demonstrated the practical application of a quality assurance programme for thermal imagers in medical use. This programme successfully validated thermography for a multi-centre clinical trial into cold-challenge reliability for Raynaud's phenomenon assessment.

Overall, the findings on the performance of modern uncooled focal plane array imagers were reassuring, but our results did confirm that the thermal imagers with less exacting accuracy specifications (as provided to Centre 6) do not match the performance of higher-specification devices.

It should be noted that an understanding of instrument performance is not the only prerequisite for reliable and repeatable clinical thermography. Rigorous adherence to patient preparation, image capture and analysis protocols is also vital, and guidelines on these topics are areas of ongoing international research (11).

Nonetheless, quality assurance based around a room-temperature blackbody device is vital to validate medical thermography, and we recommend a protocol similar to that which we have described for all clinical studies. The protocol outlined in this document could be used as part of acceptance testing for new thermal imaging cameras, monitoring the performance of these devices as they become older, and providing evidence to support the financial costs of replacement.

Acknowledgements

This study was funded by an Arthritis Research UK Clinical Studies Grant [20656] and an unrestricted educational grant from Actelion Pharmaceuticals Ltd. This study was supported by Arthritis Research UK grant number (Centre for Epidemiology) 20380 and the Manchester Academic Health Sciences Centre (MAHSC). The authors acknowledge the assistance given by IT Services and the use of the Computational Shared Facility at The University of Manchester.

The authors thank the following individuals for their help with data collection Sarah A Leggett, Tonia L Moore, John Allen, Marina E Anderson, Jason Britton, Maya H Buch, Francesco Del Galdo, Christopher P Denton, Graham Dinsdale, Bridgett Griffiths, Frances Hall, Audrey MacDonald, Neil J McHugh, Joanne B Manning, John D Pauling, Christopher Roberts, Jacqueline A Shipley, Ariane L Herrick.

References

1.Ring EF, Ammer K: Infrared thermal imaging in medicine. Physiological Measurement, 2012, 33:R33-46.

2.Plassmann P, Ring EFJ, Jones CD: Quality assurance of thermal imaging systems in medicine. Thermology International 2006, 16:10-15.

3.Preston-Thomas H: The international temperature scale of 1990 (ITS-90). Metrologia 1990, 27:3-10.

4.Howell KJ, Smith RE: Guidelines for specifying and testing a thermal camera for medical applications. Thermology International 2009, 19:5-12.

5.International Electrotechnical Commission. Particular requirements for basic safety and essential performance of screening thermographs for human febrile temperature screening. IEC 80601-2-59:2008.

6.Simpson R, Machin G, McEvoy H, Rusby R: Traceability and calibration in temperature measurement: a clinical necessity. Journal of Medical Engineering & Technology, 2006, 30: 212-217.

7.Murray A, Manning J, Moore T et. al.: A multicentre reliability study of laser speckle contrast imaging and thermography in patients with Raynaud's phenomenon secondary to systemic sclerosis. Abstract Number: 850. American College of Rheumatology Meeting 2017. Http://acrabstracts.org/abstract/a-multicentre-reliability-study-of-laser-speckle-contrast-imaging-and- thermography-in-patients-with-raynauds-phenomenon-secondary-to-systemic-sclerosis/ - accessed on 07/02/2018

8.Bland JM, Altman DG: Measurement Error; British Medical Journal 1996, 313: 21 September; 744

9.Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986, 8476:307-10.

10.Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. Ultrasound in obstetrics & gynecology 2003, 22:85-93.

11.Moreira DG, Costello JT, Brito CJ et al: Thermographic imaging in sports and exercise medicine: A Delphi study and consensus statement on the measurement of human skin temperature. Journal of Thermal Biology. 69:155-162. 2017.

Address for correspondence

Kevin Howell PHD ASIS FRPS MIPEM Csci
Clinical Scientist and Honorary Research Associate
Microvascular Diagnostics
UCL Institute of Immunity and Transplantation
Royal Free Hospital
Pond Street
London NW3 2QG
United Kingdom

Email: k.howell@l.ac.uk

(Received 04.02.2018, revision accepted 17.02.2018)