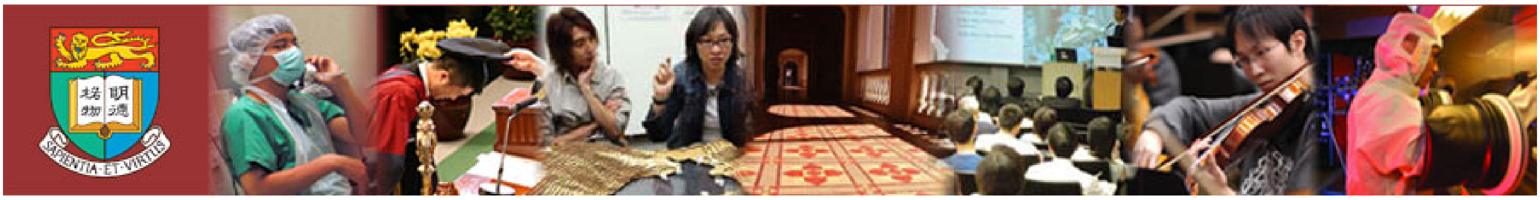

Title	Utility of infrared thermography for screening febrile subjects
Author(s)	Chan, LS; Lo, JLF; Kumana, CR; Cheung, BMY
Citation	Hong Kong Medical Journal, 2013, v. 19, p. 109-115
Issued Date	2013
URL	http://hdl.handle.net/10722/183107
Rights	Creative Commons: Attribution 3.0 Hong Kong License

O R I G I N A L A R T I C L E

Utility of infrared thermography for screening febrile subjects

LS Chan 陳龍生
Jessica LF Lo 盧玲芬
Cyrus R Kumana 顧崇仁
Bernard MY Cheung 張文勇

Objective To assess the utility of remote-sensing infrared thermography as a screening tool for fever.

Design Cross-sectional study comparing body temperatures measured by remote-sensing infrared thermography (maximum for frontal, forehead, or lateral views) with core temperatures measured by aural or oral methods.

Setting Accident and Emergency Department, Queen Mary Hospital, Hong Kong.

Participants A total of 1517 patients (747 men, 770 women) with or without fever; 34 of whom entered a substudy to measure the effects of distance on recorded temperature.

Main outcome measures The proportions of subjects with fever (core temperature of 38°C or above) detected by remote-sensing infrared thermography compared with the proportion detected by conventional thermometry.

Results The correlations between infrared thermography temperatures and core temperature were only moderate (r=0.36-0.44), albeit statistically significant. The temperature recorded by infrared thermography was inversely proportional to the distance from the camera. There were 113 (7.4%) subjects with a core temperature of 38°C or above. The areas under the receiver operating characteristic curves for the three infrared thermography measurements were around 0.8. However, the maximum sensitivity achieved at a low cut-off temperature of 35°C was only 0.87 (for frontal and lateral infrared thermography views), resulting in 13% of febrile subjects being missed. The maximum forehead temperature in general had the poorest performance among the three infrared thermography views.

Conclusions Forehead infrared thermography readings from a distance should be abandoned for fever screening. Although maximum lateral or frontal infrared thermography temperatures have reasonable correlations with core temperatures and areas under the receiver operating characteristic curves, the sensitivity-specificity combination might still not be high enough for screening febrile conditions, especially at border crossings with huge numbers of passengers.

Key words
Body temperature; Fever; Infrared rays; Sensitivity and specificity; Thermography

Hong Kong Med J 2013;19:109-15

The University of Hong Kong, Pokfulam, Hong Kong:
Department of Earth Sciences, Faculty of Science
LS Chan, PhD
Department of Medicine, Faculty of Medicine
JLF Lo, BNurs, MMedSc
CR Kumana, FRCP
BMY Cheung, PhD, FRCP

Correspondence to: Prof LS Chan
Email: chanls@hku.hk

New knowledge added by this study

- Although maximum lateral and/or frontal infrared thermography temperatures correlate with core temperatures, the sensitivity and specificity of these measurements might still not be high enough for screening febrile conditions.

Implications for clinical practice or policy

- Checking body temperature at border crossing using infrared thermography in its present form (on the face) is of questionable value and its continued use should be reviewed.

Introduction

Since the global outbreak of severe acute respiratory syndrome in 2003, infrared thermographic (IRT) instruments ranging from single-point infrared probes to full-image IRT cameras have been introduced at many airports and border crossings for screening travellers with elevated body temperatures. These systems are either IRT imaging cameras

紅外線體溫檢測作發燒篩檢的使用

目的 評估遙感紅外線測量儀作為一種發燒篩檢工具的效用。

設計 此橫斷面研究比較遙感紅外線體溫檢測（正面、前額及側向的最大值）所得的溫度與使用耳溫計或口腔探熱所得的中心溫度。

安排 香港瑪麗醫院急症室。

參與者 共1517人（747名男性，770名女性）參與本研究，當中包括有或無發燒的病人；其中34人進一步測量距離對於所量得溫度的影響。

主要結果測量 把使用遙感紅外線檢測到有發燒（中心溫度為38°C或以上）的人數比例，與使用傳統測溫儀所得的發燒人數比例作一比較。

結果 紅外線檢測到的溫度和中心溫度之間雖然達統計學上的顯著性，但其相關性只屬中等（r=0.36至0.44）。由紅外線測量儀所量度的溫度與儀器的距離成反比。共有113人（7.4%）的中心溫度為38°C或以上。三個遙感紅外線所量得的溫度的ROC曲線下之區域約為0.8。然而，如果用35°C為低截取值，最高敏感性只達0.87（正面和側面紅外線成像），意味着會有13%的發燒者未能被偵察到。在正面、前額及側向的三個角度中，前額溫度表現最差。

結論 應放棄使用前額溫度為遙感紅外線作發燒篩檢。雖然最大側向或正面的遙感紅外線成像溫度與體溫及ROC曲線下之區域有合理相關性，但其靈敏度及特異性可能還未足以作為發燒篩檢的條件，尤於有大量旅客人流的邊境口岸處為甚。

positioned at a passenger corridor and aimed at obtaining the maximum frontal view temperature of the passenger from a distance, or single-point probe sensors modified to capture the temperatures from a point or a line on the face of the passenger. Although claims have been made about the effectiveness of these methods, there is a lack of data from properly conducted studies.

A flurry of discussion articles on the IRT screening method appeared in 2004 and 2005, with some expressing reservations about its effectiveness,[1-4] and others advocating its extensive use as a screening measure.[5-7] In the studies that have been reported so far, factors affecting measurement results have not been thoroughly examined and the numbers of subjects studied were generally small, making subgroup comparisons difficult.

In this study, we aimed to measure core and IRT temperature in unselected subjects who presented to an accident and emergency (A&E) department. The large database generated enabled us to examine statistically the correlation and agreement between conventionally determined core temperatures and IRT temperatures for different age, gender, and ethnic groups.

Methods

During the period October 2005 to July 2006, convenience samples of patients attending the A&E Department of Queen Mary Hospital, Hong Kong, were approached for voluntary participation in the study. The study protocol and ethics were approved by the Institutional Review Board of the Hong Kong West Cluster. Verbal informed consent was obtained after explanation of the investigation's purpose and safety and that patient confidentiality would be maintained. Information on gender, age, ethnicity and skin complexion, as well as the wearing of a facemask, braces, or glasses was recorded. The study sample consisted of 747 males and 770 females.

All patients presenting to the A&E Department underwent routine body temperature measurement. This entailed conventional oral thermometry or an aural temperature measurement; if readings from both methods were available the higher of the two was used. Ambient temperature, barometric pressure, humidity, the amount of clothing warn by the subject, and the time of measurement were recorded. In this study, we defined fever as a core temperature of 38°C (100.4°F) or more.

The IRT images of the front and left side of the face were taken at a fixed distance of 1.5 m with a FLIR Systems ThermaCAM S40 infrared camera with a 24° lens made in the United States. The software program (ThermaCAM Researcher developed by FLIR Systems) was used to extract data from the thermogram temperatures of designated spots over the face. The IRT temperatures of the following areas were logged: forehead, temples, nose, mouth, cheeks, maximum temperature of entire face (with and without a mask), temple (from side view of face), ear, neck, and maximum temperature of side view of face. From these values, we derived three variables, namely AREAMAX for maximum frontal temperature, FOREHEAD for forehead temperature, and LATMAX for maximum lateral temperature.

The data were stratified by age and gender. Pearson correlation coefficients between IRT and core temperatures were determined. Independent t tests were used to examine the mean difference between males and females. The sensitivity, specificity, positive and negative predictive values of IRT at different cut-off points were determined for the three derived variables. We determined the receiver operating characteristics (ROC) by plotting sensitivity against 1-specificity. The likelihood ratio for a positive test describes the probability of having a positive test result in the true febrile group (sensitivity) to that of having a positive test in the non-febrile group (1-specificity), while the likelihood ratio for a negative test is the ratio of (1-sensitivity) to specificity. The areas under the ROC curve (AUROC) and the 95% confidence intervals (CIs) of the three variables were compared. In this regard, any statistically significant larger area indicated better performance.

We investigated the influence of distance from the camera on IRT temperature readings by using a simple linear regression. This analysis was based on the data from 31 voluntary subjects and their maximum recorded frontal temperatures. The maximum frontal temperatures were measured at distances between 1 m and 5 m.

Results

A total of 1517 subjects (747 men, 770 women) consented to participate in the study. The initial results were reported by Cheung et al in 2012.[8] Table 1 shows the number of subjects, their mean core temperature, IRT temperatures readings, and the correlation coefficients between the IRT and core temperatures for individual groups classified by age, gender, ethnicity, and febrile condition. There were 113 (7.4%) subjects with a core temperature of 38°C or above.

The ambient temperatures at the time of these measurements ranged from 20°C to 27°C. The regression (R^2) analysis revealed a slight dependency of the IRT temperature on ambient temperature at a rate of 0.196°C per °C of increase in ambient temperature. However, owing to such a small R^2 (0.093), it was deemed not justified to adjust for ambient temperature. Hence, all subsequent analyses were based on the original IRT temperature readings.

Figure 1 shows the linear relationship between IRT and core temperatures. Invariably, IRT temperatures were lower than core temperatures and highly variable, but showed significant (P<0.001) correlations (Table 1). The correlation between core and IRT temperatures depended on age, gender, and the actual core temperature. The correlation coefficients were generally higher in persons aged ≤20 years, males, and those who were febrile. The

TABLE 1. Mean core and infrared thermographic temperatures and correlation coefficient (*r*) of AREAMAX, FOREHEAD, and LATMAX with core temperature for different groups*

Group	Mean age (years)	No. of subjects	Core temp (°C)	AREAMAX		FOREHEAD		LATMAX	
				Mean temp (°C)	Correlation coefficient (*r*) with core temp	Mean temp (°C)	Correlation coefficient (*r*) with core temp	Mean temp (°C)	Correlation coefficient (*r*) with core temp
All subjects	45.8	1517	36.9	35.2	0.434	33.8	0.361	35.4	0.440
Age-groups (years)									
1-2	1.4	20	37.7	36.2	0.628	34.7	0.584	36.3	0.567
3-6	4.2	55	37.7	35.9	0.561	34.6	0.433	36.1	0.576
7-10	8.5	47	37.4	35.9	0.606	34.6	0.720	35.9	0.581
11-19	15.5	73	37.1	35.6	0.646	34.5	0.658	35.8	0.594
20-29	25.1	200	36.8	35.2	0.350	34.1	0.292	35.8	0.390
30-39	34.9	253	36.9	35.2	0.230	33.9	0.199	35.4	0.252
40-49	44.6	217	36.8	35.3	0.355	33.8	0.277	35.5	0.367
50-65	56.2	281	36.8	35.1	0.281	33.6	0.291	35.3	0.333
66-100	76.4	371	36.8	35.1	0.422	33.3	0.283	35.3	0.418
Gender									
Males	45.3	747	36.9	35.3	0.410	33.9	0.384	35.5	0.410
Females	46.2	770	36.9	35.2	0.286	33.7	0.252	35.4	0.274
Ethnicity									
Chinese	46.5	1439	36.9	35.2	0.337	33.8	0.306	35.4	0.333
Non-Chinese	33.3	78	36.9	35.3	0.484	33.9	0.422	35.4	0.388
Febrile condition									
Febrile	31.1	111	38.8	36.5	0.328	34.9	0.224	36.7	0.391
No-fever†	47.0	1400	36.7	35.1	0.273	33.7	0.241	35.3	0.265

* Core temp denotes core temperature, AREAMAX maximum frontal temperature, FOREHEAD forehead temperature, and LATMAX maximum lateral temperature

† Only subjects with a body temperature <38°C were included in the no-fever group, and missing data were not included in the analysis

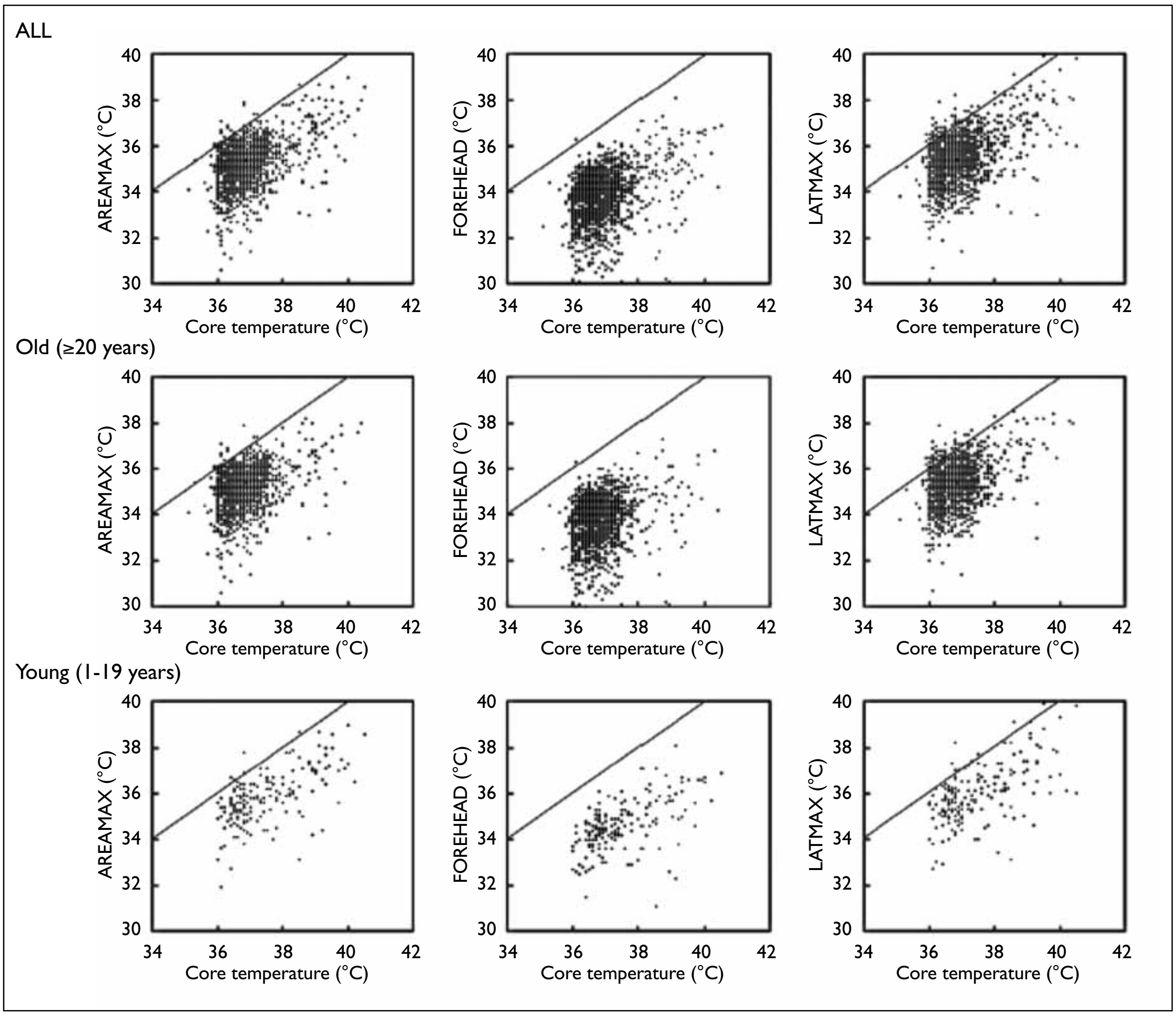

FIG 1. Scatter plots of infrared thermographic (IRT) temperature readings against core temperature readings for all samples: young (aged 1-19 years) and old (aged ≥20 years) groups. Solid line: reference line when two readings are equal

AREAMAX denotes maximum frontal temperature, FOREHEAD forehead temperature, and LATMAX maximum lateral temperature

respective correlation coefficients for AREAMAX, FOREHEAD and LATMAX were 0.410, 0.384, and 0.410 for males, and 0.286, 0.252, and 0.274 for females. While there was no significant difference between males and females with respect to mean core temperatures (P=0.287), the corresponding P values for the three IRT temperatures were 0.119, 0.005, and 0.007 for AREAMAX, FOREHEAD, and LATMAX. Whilst the sample comprised 1439 Chinese and 78 non-Chinese subjects, there were no obvious inter-ethnic differences in terms of correlations between IRT and core temperatures.

In general, higher correlations were observed in subjects with a core temperature of ≥38°C (Table 1). For subjects with lower core temperatures (<38°C), the correlation coefficients with IRT temperatures were in the range of 0.241 to 0.273. Moreover, differences between core and IRT temperatures were the greatest in febrile subjects. While the forehead temperature was on average of 3.0°C lower than the core temperature for 'non-febrile' subjects, the difference was 3.9°C for febrile subjects.

Since the IRT temperature would be used as a screening test for fever, a range of cut-off temperatures were examined for their utility in discriminating between febrile and non-febrile states. Figure 2 shows the ROC curves to detect a core temperature of 38°C or higher. The sensitivity, specificity, predictive values and likelihood ratios are shown in Table 2. The overall AUROC for AREAMAX, FOREHEAD and LATMAX were 0.812 (95% CI, 0.761-0.863), 0.780 (95% CI, 0.723-0.837), and 0.815 (95% CI, 0.763-0.867), there being no statistically significant difference regarding the three corresponding AUROC in the male (0.830, 95% CI, 0.764-0.896; 0.786, 95% CI, 0.707-0.865; 0.823, 95% CI, 0.750-0.895) and female subgroups (0.790, 95% CI, 0.711-0.869; 0.773, 95% CI, 0.691-0.855; 0.808, 95% CI, 0.735-0.881).

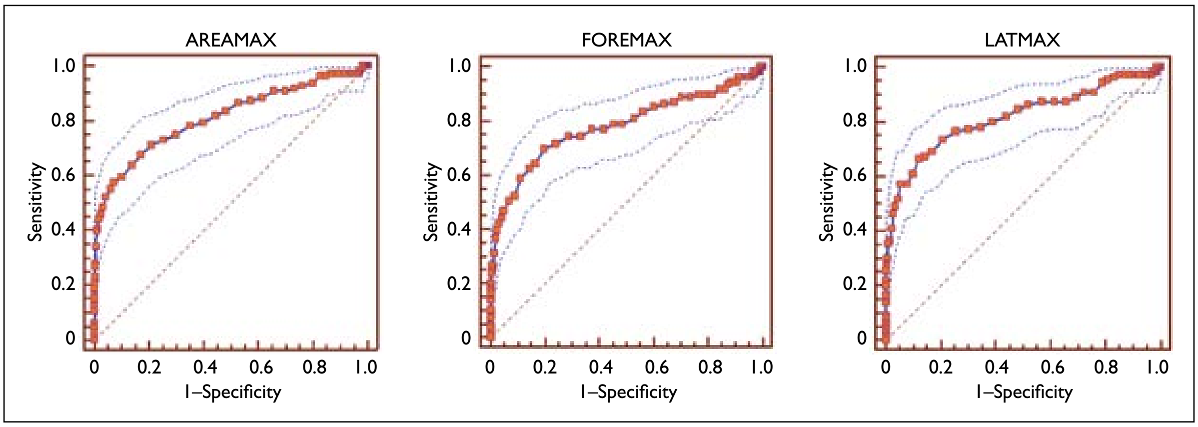

FIG 2. Receiver operating characteristics curve and the 95% confidence intervals (dotted curves) of infrared thermographic temperatures to detect fever (core temperature ≥38°C). The diagonals are lines of no-discrimination; data points falling above the diagonals are considered having better predictive values than random guesses
AREAMAX denotes maximum frontal temperature, FOREHEAD forehead temperature, and LATMAX maximum lateral temperature

TABLE 2. Performance parameters at different infrared thermographic threshold temperatures for the three variables*

Variables	AREAMAX				FOREMAX				LATMAX			
	Data				Data				Data			
Threshold temp (°C)	35	36	36.5	37	35	36	36.5	37	35	36	36.5	37
Sensitivity	0.87	0.64	0.49	0.34	0.50	0.24	0.14	0.04	0.87	0.74	0.57	0.41
Specificity	0.43	0.86	0.97	0.99	0.93	0.997	1.00	1.00	0.34	0.79	0.92	0.98
Positive predictive value	0.11	0.27	0.55	0.79	0.36	0.87	1.00	1.00	0.10	0.22	0.37	0.59
Negative predictive value	0.98	0.97	0.96	0.95	0.96	0.94	0.94	0.93	0.97	0.97	0.97	0.96
Positive likelihood	1.53	4.62	15.14	47.93	7.36	83.5	∞	∞	1.33	3.54	7.49	18.46
Negative likelihood	0.29	0.42	0.53	0.66	0.53	0.76	0.86	0.96	0.37	0.33	0.46	0.60
Test positive proportion†	0.58	0.17	0.07	0.03	0.10	0.02	0.01	0.00	0.63	0.25	0.11	0.05

* AREAMAX denotes maximum frontal temperature, FOREHEAD forehead temperature, and LATMAX maximum lateral temperature
† Test positive proportion: the proportion of persons with the defined infrared thermographic threshold temperature

Effects of distance on the infrared thermographic temperature recorded

As shown in Figure 3, the measured distance had a significant effect on IRT readings. Using the IRT measurement at 1 m as the reference, the IRT temperature was 0.35°C lower at 2 m, and 1.1°C lower at 5 m. The IRT temperature decreased on average by 0.3°C per metre increase in distance from the camera.

Discussion

In principle, the measured IRT temperature depends on the radiation efficiency of the object in question and the transmittance of the air between the camera and that object. Human skin temperature is known to vary substantially from one part of the body to another. If IRT is to be used for the screening of travellers with fever at ports and border crossings, consideration must be given to the specific locations on the face that should be targeted, and the possible influence of other variables on the body temperature.

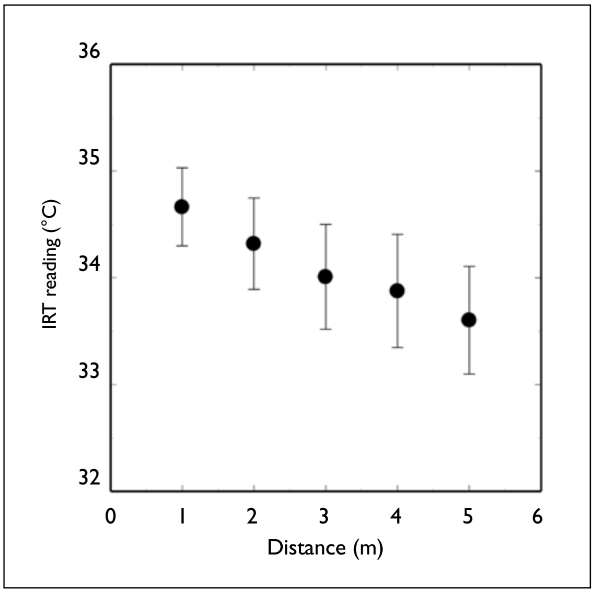

FIG 3. Relationship between infrared thermographic (IRT) temperature and distance of measurement
Regression: IRT temperature (°C) = 34.85 – 0.26 distance (m)

Most existing IRT screening methods compare the maximum temperature of the forehead or frontal view of the face to a threshold temperature. In a previous study based on 176 subjects, we raised concern over the reliance placed on using forehead temperature (as the screening criterion) commonly applied at border crossings.[1] Liu et al[9] also suggested using the auditory meatus temperature for the purpose of screening. Based on a sample size of 502 hospital patients, Ng et al[5] reported a correlation coefficient of 0.4974 and an AUROC of 0.979, when using the maximum forehead temperature as the test variable. This reported AUROC value was remarkably high, despite the correlation coefficient being only fair. However, in that study only 310 of the 502 were included in the regression and ROC analyses.

Our study revealed only very moderate, albeit statistically significant, correlation coefficients between IRT and core temperatures, and were generally <0.5 for the three types of measurement tested. The maximum forehead temperature was the least efficient method of screening, since it showed the lowest correlation with the core temperatures and had the smallest AUROC among the three variables. Furthermore, its mean among the study subjects deviated quite substantially from the mean core temperature (-3.10°C). This result sheds doubt on whether the single-point–type IR probes aimed at the forehead should continue to be used. The results also showed that three other factors (gender, age, and core temperature) had an effect on the utility of IRT temperature as a proxy for measuring body temperature.

In non-febrile subjects, the correlation coefficients between IRT and core temperatures were poor, rendering the use of IRT temperature almost useless as an approximation of the core temperature. In normal persons, there are multiple means of dissipating body heat, whereas in febrile persons, the dissipation of extra body heat manifests as elevated skin temperature.

It is evident that IRT is more applicable in younger age-groups, especially toddlers and teens.[10,11] Interestingly, mean temperatures appeared higher at younger ages, in part because a greater portion of younger persons visited the hospital because of fever, as indicated by their average core temperatures. Moreover, the core temperature is purported to be lower in the elderly, and their febrile response to infection may also be attenuated.[12]

The AUROCs were reasonable for the entire sample and there was no significant difference in the AUROC values among the three IRT measurements for male and female genders, despite the IRT temperatures of the forehead having low sensitivity as a performance measure. Nevertheless, the ROC curves did not reveal a cut-off temperature that had both very high sensitivity and specificity. Even at a low cut-off temperature of 35°C, the maximum sensitivity was only 0.87 (AREAMAX and LATMAX), which means that 13% of subjects with core temperature above 38°C would be missed. Furthermore, those testing positive would be >50% (requiring confirmatory temperature measurements), which was much higher than the actual proportion of 7.4% febrile subjects. This resulted in a high proportion of false positives and a low positive predictive value. A positive likelihood ratio of <2 also suggested poor ability to distinguish febrile from non-febrile patients. Increasing the cut-off temperature would increase the positive likelihood ratio and reduce the test-positive proportion (with a reduction in false positives and improvement in positive predictive value). However, this would be at the expense of reductions in sensitivity and missing higher proportions of subjects with fever. Although the actual proportion of subjects with fever at border crossings (including airports) would be much lower than in the study sample recruited from A&E department in the current study, the huge number of passengers that pass through such crossings every day could still be a nuisance, if a substantial proportion of them had to have a second test, even though the vast majority would have a negative result. The distance between the infrared camera and the subject had an important influence on the reading. Although the camera can be calibrated for different distances, this would be impractical at border crossings and airports.

Conclusion

Infrared thermographic temperature correlates only moderately with core temperature, but performs better in children, men, and among febrile subjects. The IRT temperature is inversely proportional to the distance from the camera. Although the study results suggested better test performances using either the maximum lateral or frontal temperature, their sensitivity might still not be high enough and the high number/proportion of false positives would be overwhelming. This property renders IRT unsuitable as a routine screening tool for febrile conditions, especially at border crossings with huge numbers of passengers. A single IRT measurement of the forehead from a distance should be replaced by a method with greater sensitivity and specificity.

Acknowledgements

This study was supported by the Research Fund for the Control of Infectious Diseases (Grant No. 03040232). A short preliminary version of this paper has been included in the Research Dissemination Reports of the Research Fund (Cheung BM, Chan LS, Lauder IJ, Kumana CR. Detection of body temperature with infrared thermography: accuracy in detection of fever. Hong Kong Med J 2012;18 Suppl 3:31-4). Maggie Chan contributed to the project as research assistant and performed some of the data analysis. Kouping Chen also assisted in undertaking the data analysis.

References

1. Chan LS, Cheung GT, Lauder IJ, Kumana CR. Screening for fever by remote-sensing infrared thermographic camera. J Travel Med 2004;11:273-9.
2. Peacock GR. Human radiation thermometry and screening for elevated body temperature in humans. Thermosense XXVI. Proceedings of SPIE 2004;5405:48-53.
3. St John RK, King A, de Jong D, Bodie-Collins M, Squires SG, Tam TW. Border screening for SARS. Emerg Infect Dis 2005;11:6-10.
4. Wu M. Stop outbreak of SARS with infrared cameras. Thermosense XXVI. Proceedings of SPIE 2004;5405:98-105.
5. Ng EY, Kaw GJ, Chang WM. Analysis of IR thermal imager for mass blind fever screening. Microvasc Res 2004;68:104-9.
6. Ng EY, Kaw G, Ng K. Infrared thermography in identification of human elevated temperature with biostatistical and ROC analysis. Thermosense XXVI. Proceedings of SPIE 2004;5405:88-97.
7. Tan YH. Teo CW. Ong E, Tan LB, Soo MJ. Development and deployment of infrared fever screening systems. Thermosense XXVI. Proceedings of SPIE 2004;5405:68-78.
8. Cheung BM, Chan LS, Lauder IJ, Kumana CR. Detection of body temperature with infrared thermography: accuracy in detection of fever. Hong Kong Med J 2012;18 Suppl 3:S31-4.
9. Liu CC, Chang RE, Chang WC. Limitations of forehead infrared body temperature detection for fever screening for severe acute respiratory syndrome. Infect Control Hosp Epidemiol 2004;25:1109-11.
10. Craig JV, Lancaster GA, Taylor S, Williamson PR, Smyth RL. Infrared ear thermography compared with rectal thermography in children: a systematic review. Lancet 2002;360:603-9.
11. Falzon A, Grech V, Caruana B, Magro A, Attard-Montalto S. How reliable is auxillary temperature measurement? Acta Paediatr 2003;92:309-13.
12. Norman DC. Fever in the elderly. Clin Infect Dis 2000;31:148-51.