BMY Cheung 張文勇
LS Chan 陳龍生
IJ Lauder
CR Kumana 顧崇仁

Detection of body temperature with infrared thermography: accuracy in detection of fever

Key Messages

1. Infrared thermography (IRT) for detecting body temperature is less accurate in women, elderly people, and those with fever.
2. The core temperature significantly but weakly correlates to the IRT temperatures obtained from frontal and lateral of the face, and the forehead.
3. Among the three areas, the forehead IRT temperature showed the largest discrepancy and poorest correlation with the core temperature.
4. If IRT is used, the lateral maximum temperature of the face should be used. A cut-off temperature of 36ºC gives 77% sensitivity and 74% specificity.
5. Owing to its weak correlation with the core temperature, IRT should not replace direct body temperature measurement in clinical situations.

Hong Kong Med J 2012;18(Suppl 3):S31-4

The University of Hong Kong:
Department of Medicine
BMY Cheung, CR Kumana
Department of Earth Sciences
LS Chan
Department of Statistics and Actuarial Science
IJ Lauder

RFCID project number: 03040232

Principal Investigator and corresponding author:
Prof Bernard MY Cheung
Department of Medicine, Queen Mary Hospital, Hong Kong SAR, China
Tel: (852) 2255 4347
Fax: (852) 2818 6474
Email: mycheung@hku.hk

Introduction

Since the outbreak of severe acute respiratory syndrome, infrared thermography (IRT) systems have been deployed at the airport and border crossings in Hong Kong for screening travellers. However, its use to identify people with elevated body temperature is limited. In a pilot study of 176 subjects,[1] temperatures measured by IRT might be used as a proxy for core temperature, but they are affected by a variety of factors, such as the part of the face measured. We aimed to investigate the effectiveness of IRT to identify people with fever.

Methods

This study was conducted from September 2005 to August 2006. The protocol was approved by the Institutional Review Board of the Hong Kong West Cluster of hospitals. Unselected patients attending the accident and emergency department of the Queen Mary Hospital were invited to participate. Patients on stretchers or needing immediate emergency treatment were excluded. Verbal informed consent was obtained from each subject.

The core temperature was defined as either the oral or aural temperature, or whichever was higher if both were available. At ports and border crossings, the maximum IRT temperatures obtained from the frontal (Areamax) or lateral (Latmax) of the face or the forehead temperature were used as proxies for the core temperature. Ambient temperature, barometric pressure, and humidity were also recorded. The degree of clothing and the time of measurement were noted.

For the study of the effect of distance on IRT readings, temperatures of 31 healthy (afebrile) volunteers were measured in a controlled laboratory setting with the subjects standing at 1, 2, 3, 4 and 5 m from the IRT camera.

The software program ThermaCAM Researcher was used to extract from the IRT temperatures of designated parts of the face. Data analysis was stratified by age and gender. Pearson correlation coefficients between IRT temperatures and oral/tympanic temperature were determined. The 95% confidence limits of agreement of IRT measurements with the reference method were calculated according to the method of Bland and Altman.[2] The standard error of the 95% limit of agreement is approximately $\sqrt{(3s^2/n)}$, where s is the standard deviation of the differences between measurements by the two methods, and n is the sample size.[2] The receiver operator characteristics were determined by plotting the sensitivity against 1-specificity. The sensitivity, specificity, false-positive, and false-negative rates of IRT were calculated. Likelihood ratios, which describe the odds of getting a positive or negative test result, were calculated from the sensitivity and specificity.

Results

A total of 747 men and 770 women consented to participate; 215 of them had a core temperature of ≥37.5ºC and were considered to have fever. The forehead IRT temperature showed the largest discrepancy from the core temperature and was on average 3.1ºC lower. The Latmax yielded the best correlation with the core temperature (r=0.441), whereas the forehead IRT temperature yielded the poorest correlation (r=0.361) [Table 1].

In all subgroups examined, forehead IRT temperature was consistently lower than Latmax or Areamax (Fig 1). The difference between core and IRT temperature was greatest in febrile subjects; the forehead IRT temperature was on average 3.0ºC and 3.7ºC lower than the core temperature in afebrile and febrile subjects, respectively (Fig 1).

In the Bland-Altman plots of the difference between the IRT and core temperatures against the mean of the IRT and core temperature, IRT temperatures were on average lower than the core temperature. The difference between IRT and core temperatures increased as core temperature decreased (Fig 2).

The subjects were divided into nine age groups (1-2, 3-6, 7-10, 11-19, 20-29, 30-39, 40-49, 50-65, 66-100 years). The best correlation of IRT temperatures with core temperature was seen in children (aged 3-18 years), followed by infants (aged 1-2 years). Male subjects showed better correlation between IRT and core temperatures. The respective correlation coefficients for the three variables of Areamax, Latmax, and Forehead were 0.496, 0.5, and 0.404 for males, and 0.369, 0.385, and 0.323 for females (Table 1). A better correlation was observed in subjects with a core temperature of ≥37.5ºC. For subjects with a normal body temperature, the correlation coefficients between the IRT and core temperatures tended to be <0.25.

Ambient temperature had a minor effect on IRT values. Each 1ºC change in ambient temperature changed the IRT values by 0.196ºC on average.

The sensitivity, specificity, type-I error, and type-II error at different IRT temperatures are tabulated in Table 2. At 36ºC, the positive and negative likelihood ratios were 3.97 and 0.39 for the Latmax, respectively.

Table 1. Mean infrared thermographic (IRT) temperatures for the frontal (Areamax) and lateral (Latmax) of the face and the forehead, and correlation coefficients (*r*) between IRT and core temperatures

Parameter	Areamax (n=1511)	Latmax (n=1513)	Forehead (n=1509)
Mean±SD IRT temperature (ºC)	35.23±0.99	35.43±1.03	33.79±1.15
Mean±SD difference from core temperature (ºC)	-1.67±0.93	-1.46±0.96	-3.10±1.11
Mean±SE lower limit of agreement	-3.49±0.04	-3.34±0.04	-5.28±0.04
Mean±SE upper limit of agreement	0.15±0.04	0.42±0.04	-0.92±0.04
r for all	0.434	0.441	0.361
r for males	0.496	0.500	0.404
r for females	0.369	0.385	0.323

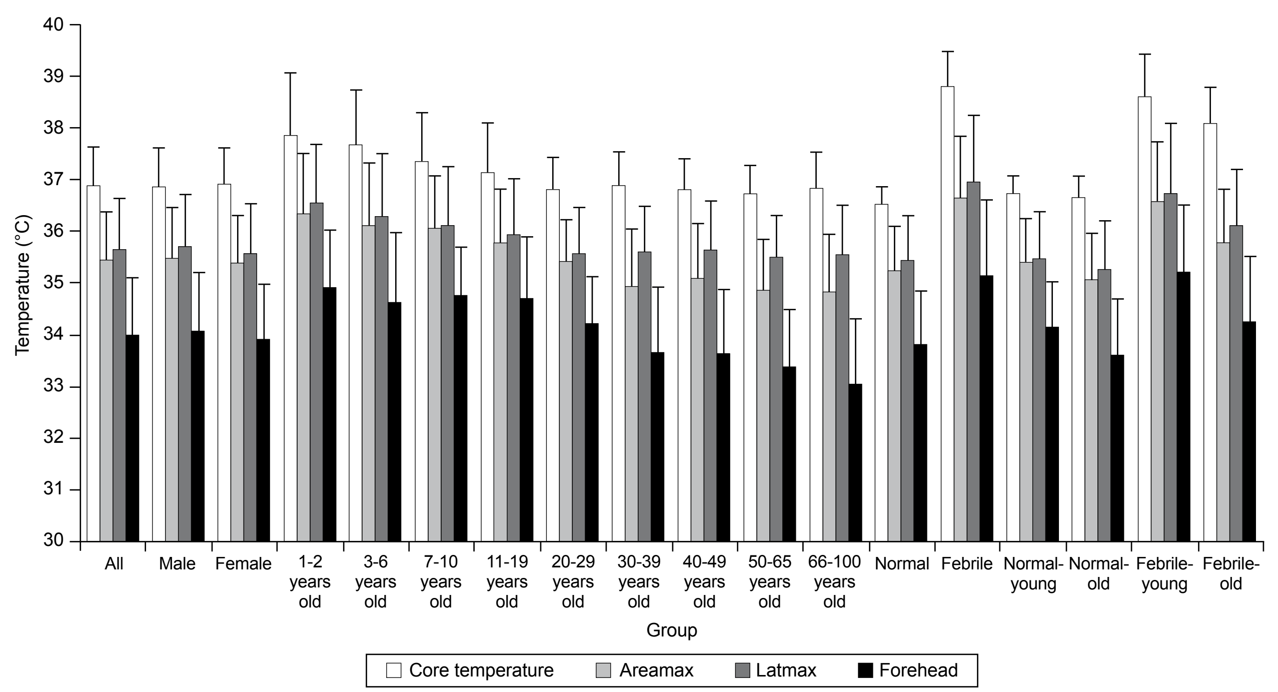

Fig 1. Mean and standard deviation of core and infrared thermography temperatures in different subgroups

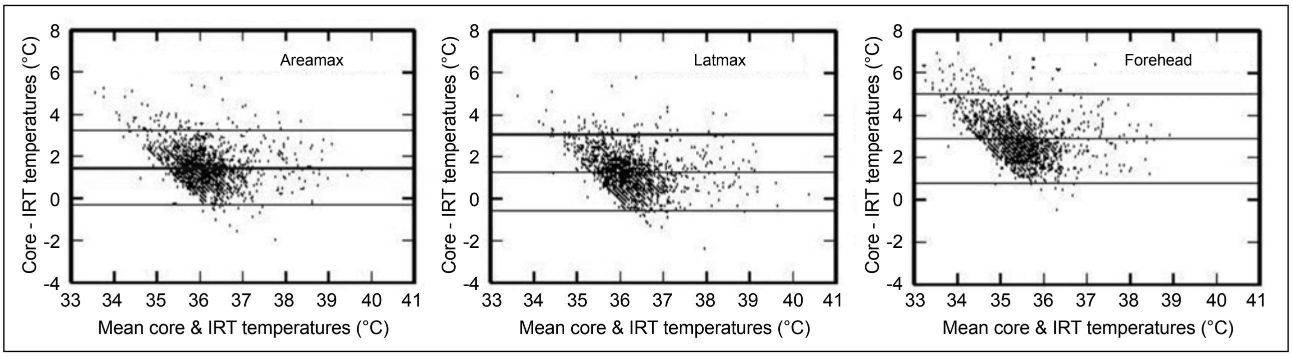

Fig 2. Bland-Altman plots of the difference between core and infrared thermography (IRT) temperatures of the frontal (Areamax) or lateral (Latmax) of the face or the forehead against the means of core and IRT temperatures

Table 2. Sensitivity and specificity of maximum frontal and lateral infrared thermographic (IRT) temperatures

Parameter	Cut-off temperature							
	34ºC	34.5ºC	35ºC	35.5ºC	36ºC	36.5ºC	37ºC	37.5ºC
Maximum frontal IRT temperature								
Sensitivity	0.97	0.94	0.88	0.79	0.68	0.52	0.40	0.21
Specificity	0.08	0.20	0.39	0.60	0.83	0.96	0.99	1.00
Type-II error, β	0.03	0.06	0.12	0.21	0.32	0.48	0.60	0.79
Type-I error, α	0.92	0.80	0.61	0.40	0.17	0.04	0.01	0.00
False negative rate	0.02	0.02	0.02	0.03	0.03	0.04	0.05	0.06
False positive rate	0.92	0.91	0.90	0.86	0.76	0.51	0.21	0.08
Failing %*	92.0	80.9	63.3	42.8	20.7	7.9	3.7	1.7
Maximum lateral IRT temperature								
Sensitivity	0.97	0.96	0.89	0.85	0.77	0.61	0.46	0.21
Specificity	0.07	0.17	0.31	0.52	0.74	0.90	0.97	1.00
Type-II error, β	0.03	0.04	0.11	0.15	0.23	0.39	0.54	0.79
Type-I error, α	0.93	0.83	0.69	0.48	0.26	0.10	0.03	0.00
False negative rate	0.03	0.02	0.03	0.02	0.02	0.03	0.04	0.06
False positive rate	0.92	0.92	0.91	0.88	0.81	0.67	0.43	0.23
Failing %*	92.9	84.0	70.9	51.0	29.4	13.5	6.1	2.0

* Total percentage of subjects tested positive

Distance between subject and IRT had a significant effect on IRT readings; IRT temperature decreased linearly with distance (p=0.001). Using 1 m as the reference, the IRT temperature was 0.35ºC lower at 2 m and 1.1ºC lower at 5 m. The IRT temperature decreased on average by 0.26ºC per meter of distance.

Discussion

The correlation of IRT temperatures with the core temperature was significant but weak (r<0.45). Gender, age, and core temperature influenced the accuracy of IRT temperature as a proxy for body temperature. Females showed a poorer correlation between IRT and core temperatures. It is not possible to rule out if this was due to cosmetics, as such data were only available on three subjects.

The IRT system seems more accurate in younger age groups, especially children and teenagers.[3-5] The core temperatures were higher in children than adults, perhaps because children with fever were more likely to attend hospital. The core temperatures in the elderly were lower, and their febrile response to infection could be attenuated.

The Bland-Altman analysis showed that IRT temperatures were lower than the core temperature, especially when the core temperature was low. This finding may be useful as it reduces the number of people with a normal core temperature being mistaken for having fever.

The use of forehead IRT temperature as a proxy for the body temperature is questionable.[6] The forehead IRT temperature was lowest among the three IRT temperatures of the face. Its correlation with the core temperature was also lowest. Based on the forehead IRT readings, if 37ºC was used as the cut-off temperature for screening, the sensitivity was exceedingly low (4%). Reducing the cut-off temperature to 36ºC and 35ºC increased the sensitivity to 25% and 52%, respectively. To achieve a sensitivity of about 79%, the cut-off temperature should be lowered to 34ºC. This, however, would yield a specificity of 55% and a false positive rate of 88% (88% of those tested positive would actually be afebrile). This would require an unacceptably high percentage (47.8%) of subjects to be retested. Thus, the forehead IRT temperatures are not effective in screening passengers with fever. This casts doubt on the efficacy of using a single-point IRT probe to detect passengers with fever.

When the maximum frontal IRT temperature was used as the screening temperature, a cut-off temperature of 36ºC would yield a sensitivity of 68% and would result in 22.4% of all subjects to fail the screening. This is much better than the forehead IRT temperature in terms of sensitivity and retesting rate. Reducing the cut-off temperature to 35.5ºC would yield a sensitivity of 79% and a specificity of 60%. However, 86% of those tested positive would actually be afebrile and the percentage of subjects failing the screening would increase to 51%.

When the maximum lateral IRT temperature was used as the screening temperature, the same cut-off temperature of 36ºC would yield a sensitivity of 77%, a specificity of 74%, and a false negative rate of 23%. This would be a reasonable setting in terms of sensitivity and false negative rate. However, it would require 29.4% of the subjects to be retested. If the percentage of subjects requiring retesting is a constraining factor, raising the cut-off temperature to 36.5ºC would reduce the percentage of subjects failing the screening to 13.5%. However, the sensitivity would be reduced to 61% and the false negative rate increased to 39%. This may be unacceptable during an epidemic.

The distance between the IRT camera and the subject is a limiting factor on the efficiency. Although the camera can be calibrated for different distances, it is impractical at border crossings and airports to do so. One particular mode of operation compares the maximum detected temperature of travellers passing in front of the camera with the temperature inside a control box kept at a constant temperature. Errors can arise if the subject and the control box are at different distances from the camera.

Conclusions

For the application of IRT in screening for travellers with elevated body temperature at airports and border crossings, the forehead IRT temperature differed substantially from the core temperature, and the maximum lateral IRT temperature should be used. The reading should also be taken at a defined distance from the camera. Overall the sensitivity of IRT in detecting fever is low unless the cut-off temperature is low. When the risk of an epidemic is high and high sensitivity is required, a low cut-off temperature (≤35.5ºC) should be chosen, although a large number of people will require a confirmatory temperature measurement. As IRT is relatively less accurate on women and older people, more sampling for aural measurement should be done on these individuals.

Acknowledgements

This study was supported by the Research Fund for the Control of Infectious Diseases, Food and Health Bureau, Hong Kong SAR Government (#03040232). We thank Ms Jessica Lo (research nurse) and Ms Maggie Chan (research assistant/data analysis).

References

1. Chan LS, Cheung GT, Lauder IJ, Kumana CR, Lauder IJ. Screening for fever by remote-sensing infrared thermographic camera. J Travel Med 2004;11:273-9.
2. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.
3. Falzon A, Grech V, Caruana B, Magro A, Attard-Montalto S. How reliable is axillary temperature measurement? Acta Paediatr 2003;92:309-13.
4. Craig JV, Lancaster GA, Taylor S, Williamson PR, Smyth RL. Infrared ear thermometry compared with rectal thermometry in children: a systematic review. Lancet 2002;360:603-9.
5. Ng DK, Chan CH, Lee RS, Leung LC. Non-contact infrared thermometry temperature measurement for screening fever in children. Ann Trop Paediatr 2005;25:267-75.
6. 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.