TEMPERATURE
https://doi.org/10.1080/23328940.2021.1899546

Taylor & Francis
Taylor & Francis Group

PRIORITY REVIEW

Non-contact infrared assessment of human body temperature: The journal *Temperature* toolbox

Josh Foster, Alex Bruce Lloyd, and George Havenith

Environmental Ergonomics Research Centre, School of Design and Creative Arts, Loughborough University, Loughborough, Leicestershire, UK

ABSTRACT

The assessment of human internal/core temperature (T_{core}) is relevant in many scientific disciplines, but also for public health authorities when attempting to identify individuals with fever. Direct assessment of T_{core} is often invasive, impractical on a large scale, and typically requires close contact between the observer and the target subject. Non-contact infrared thermometry (NCIT) represents a practical solution in which T_{core} can potentially be assessed from a safe distance and in mass screening scenarios, by measuring skin temperature at specific anatomical locations. However, the COVID-19 pandemic has clearly demonstrated that these devices are not being used correctly, despite expert guided specifications available in International Standard Organization (ISO) documents. In this review, we provide an overview of the most pertinent factors that should be considered by users of NCIT. This includes the most pertinent methodological and physiological factors, as well as an overview on the ability of NCIT to track human T_{core}. For practical use, we provide a checklist based on relevant ISO standards which are simple to follow and should be consulted prior to using NCIT for assessment of human T_{core}. Our intention is for users of NCIT to adopt this checklist, which may improve the performance of NCIT for its ability to track T_{core}.

ARTICLE HISTORY
Received 23 December 2020
Revised 2 March 2021
Accepted 3 March 2021

KEYWORDS
Fever; COVID-19; screening; guidelines; thermal camera; thermometer

Introduction

Assessment of human internal temperature is relevant in a wide variety of scientific disciplines. However, direct measurement of internal tissues can be invasive and logistically challenging, prompting the requirement for noninvasive methods. In this journal (*Temperature*), methods used in the clinical setting for local thermal tissue ablation and monitoring brain temperature have been comprehensively reviewed [1,2]. The present article focuses on non-contact infrared thermometry (NCIT) for the assessment of human internal temperature, which has primary applications for public health [3–5] and physiological research [6,7].

Non-contact infrared thermometry involves the assessment of skin surface temperature through measurement of its emitted radiation in the infrared waveband. Interest in NCIT for the assessment of human body temperature (core and skin) has surged in the wake of the COVID-19 pandemic, but was first widely adopted in the 2003 severe acute respiratory syndrome (SARS) epidemic [8]. NCIT devices have several applications but are adopted on a mass scale during pandemics to screen for elevated human body temperature which is associated with infection [9,10]. To date, using NCIT for border screening shows an extremely low ability to detect Ebola, Influenza, or SARS on a mass scale [9]. In some reports, the sensitivity of large-scale airport screening for Ebola and SARS was zero [9]. However, it is unclear if the lack of ability of NCIT to detect fever is based on i) poor operator practices, ii) because infected people were asymptomatic, or iii) because NCIT technology did not detect fever even with correct operator practices. Between 24 January and 17 February 2020, only 5.2% of the 271 total imported COVID-19 cases worldwide were detected through airport screening [11]. During the COVID-19 pandemic, the Thermology community have observed that the methods outlined by the International Standardization Organization (ISO) to ensure correct use of these devices are mostly not being followed [12]. They state that "*The minimum recommended requirements that the subjects*

CONTACT George Havenith g.Havenith@lboro.ac.uk Loughborough University, Environmental Ergonomics Research Center, Loughborough, UK

must be screened individually, facing the thermal camera, and with the face unobstructed by masks, spectacles or headwear, have simply not been implemented".

In this *Temperature* Toolbox article, NCIT is an umbrella term which can involve either a spot temperature measurement with an infrared thermometer, or the use of a thermal imaging camera, which visualizes the temperature distribution over a specified target plane. In part 1, we provide an overview of the *methodological factors* which must be considered when using NCIT to track internal/core temperature (T_{core}). In part 2, we report on the *physiological confounders* that may impact upon the ability of NCIT to track human body temperature. Controlling both the methodological and physiological confounders will improve the ability of NCIT to correctly screen for febrile body temperatures. The primary information in parts 1 and 2 was placed into a new checklist for users of NCIT, providing a simplified, digestible reference, nuanced against the clauses stipulated in ISO80,601 [13], ISO13154 [ISO13154, 14], and the general guidance provided in a consensus document on NCIT [7]. In part 3, we report on the ability of NCIT to track human T_{core} in both static (stable environment and with resting subjects) and dynamic (unstable environment and/or active subjects) conditions. We focus specifically on measurements taken at the forehead and the inner eye canthus. The ability of NCIT to track human T_{core} is fundamental for the technology to be an effective fever screening tool. Finally, we discuss future research direction which may further improve the performance of NCIT for noninvasive monitoring of human body temperatures.

As the primary focus of this review was the correct implementation of NCIT, an additional section is available as supplementary material. In this section, we produce an up-to-date perspective on the performance of NCIT in mass fever screening situations, focusing on metrics such as sensitivity, specificity, and positive and negative predictive values. This section is particularly useful for highlighting some of the probability issues associated with using NCIT in isolation to combat the spread of diseases.

Part 1. methodological considerations

Many of the issues which prohibit the utility of NCIT in physiological studies or in mass screening situations are related to users typically not following the correct protocols [12,15]. The International Standard documents [ISO13154, 13, 14] detail the technical specifications on the correct use of NCIT for fever screening. In this section, we describe the most pertinent of these methodological factors which should be considered, with the aim of improving the standard of NCIT use in medical imaging. We also provide a checklist in (Table 1), based on the literature and International Standards. It is our intention for the checklist to be applied prior to NCIT application.

Measurement site

A screening thermograph or spot measurement for the detection of fever should be measured at the face only. Whilst in news broadcast regarding the pandemic often alternative anatomical locations can be seen to be used (i.e. the wrist), these do not comply with international guidelines [ISO13154, 13, 14] and are considered too unreliable. When using a thermograph, the inner canthus of the eyelid provides a measurement site which, under appropriate conditions, correlates best with internal body temperature [16]. Using thermography, the highest temperature on the whole face region is adopted by some manufacturers, but this is not recommended in the standard [10,13]. When using a spot measurement, the center of the forehead is most adopted, and normally preferred over the inner eye-canthus [5,17]. Aside from the forehead being easier to take measurements on a mass scale, for those spot meter devices that are equipped with a laser pointer, even with low-power lasers, there is a risk of damage to the eye if it is directly exposed.

Room considerations

There are several elements which should be considered regarding the environment in which NCIT is used. Ring and Ammer [18] suggest a minimum room size of 2 × 3 m, but a bigger size is preferable since a 2 m distance (depending on the lens focal

Table 1. Checklist to ensure correct use of NCIT for tracking human core temperature.

	Purchasing any NCIT device	Tick if yes
1	The device has a stated accuracy of ± 0.3°C or lower	☐
2	The device has a minimum resolvable temperature difference (sensitivity) of 0.1°C. This is often stated as the "resolution" of spot infrared thermometers	☐
	Purchasing a thermal imaging camera	
3	The total resolution of the raw image is at-least 320 × 240 image pixels	☐
4	The device makes drift compensation (self-corrections) to keep the camera within the specified accuracy If no to above	☐
5	An external temperature reference source (ETRS, black body calibrator) is used with the device. The ETRS will have a combined stability and drift error of 0.1°C or less over a 14-day period. The ETRS will have an expanded uncertainty below ± 0.3°C	☐
	Measurement location	
6	Air temperature is stable, and between 18 and 24°C	☐
7	Air relative humidity is stable, and between 20 and 75%	☐
8	The room is free of forced air movement and sources of radiation (i.e., no reflective backgrounds, sunlight) that can impact the NCIT display temperature. A non-reflective cloth backdrop can be used to minimize sources of radiation.	☐
9	The room is at-least 2 × 3 meters in size	☐
	Device operation	
10	The user has consulted the manual to determine the correct measuring location and distance from the subject.	☐
11	The device is not overdue a calibration (see manufacturer guidelines)	☐
12	The device emissivity is set to 0.98 for long wave IR cameras or devices	☐
13	The device has warmed up to room temperature for at least 30 minutes	☐
14	The device is taking a measurement from a 90° angle and perpendicular to the front of the subjects' face	☐
15	For spot measurements, the target is specified as the center of the forehead	☐
16	For thermal imaging, the target is specified as the inner eye-canthus (i.e., inner corner of the eyelid)	☐
17	For thermal imaging, the rainbow color scale is used	☐
18	An absolute temperature threshold has been established through consultation with a healthcare professional or expert in the field of fever screening with NCIT	☐
19	Masks, eyeglasses, hats, and any other items obstructing the face have been removed	☐
	If using an ETRS	
20	The ETRS in focus and large enough to be easily discriminated from the background	☐
21	The ETRS is no larger than 10% of the size of subject's face (in a face only image)	☐
22	The ETRS is set to a temperature close to the threshold temperature (i.e., 35–38°C)	☐
	Reporting	
23	Model/manufacturer of camera; accuracy, resolution, model of ETRS, lens type, software & emissivity used	☐
24	Location, environment (air temperature & humidity), Reflected temperature	☐
25	Any pre- 'treatment of subjects scanned (i.e., acclimation to conditions)	☐

length used) is often required between the operator and the subject. IEC 80,601-2-59 (2019) stipulates that the ambient temperature should be maintained between 18°C and 24°C and relative humidity between 10% and 75%. Airflow from ventilation ducts should be deflected (i.e., with screens) to minimize forced cooling or heating of the target. The area chosen should also ensure that no source of infrared radiation (e.g., incandescent and halogen lightings) surrounds the experimental setup. For these reasons, it is therefore not advisable to conduct NCIT screening outdoors due to the transient and unexpected nature of the environmental conditions.

Emissivity

The emissivity defines the ratio of the emitted thermal energy relative to that of a perfect emitter, i.e., a black body, at the same temperature and wavelength and under the same viewing conditions [19]. Different materials of the same temperature emit infrared energy at different rates, and for long wave infrared devices, the emissivity for human skin should be set at 0.98 (0.91 for less common medium wave cameras). Setting the emissivity incorrectly can increase device error due to increased impact of environmental conditions on the calculations.

Spot distance ratio

NCIT devices that use a spot measurement (i.e., not a thermal image) can provide different results depending on the spot distance ratio and angle of the device. The spot distance ratio, or optics ratio, defines the size of the area measured relative to the distance of the object to the device. The greater the distance, the larger the measured area, but a unit with a high optics ratio will, for the same distance, provide a reading for a smaller area, which is beneficial for accuracy purposes. The optics area and the actual area over which the measurement is taken are often not considered by users, who tend to think that the laser point provided by some devices represents the measurement area. As these devices will return a value which is the *average* temperature value over a given area, this

introduces a source of error. The optimum distance is specific to each device and depends on the spot distance ratio. Finally, it is advocated to keep the device at a 90° angle when taking temperature measurements, perpendicular to the forehead (https://www.fda.gov/medical-devices/general-hospital-devices-and-supplies/non-contact-infrared-thermometers).

Pixel count/resolution

When a thermal camera is used for NCIT measurement, especially when trying to locate the small inner canthus temperature, it is important to consider the number of pixels that cover the measurement site. 20,advise against the use of single pixels in comparing data to contact sensors, while 21,have shown that temperatures measurements including more pixels tend to differ from those with a small number of pixels. A minimum size of 3 × 3 pixels is advised for any thermography measurement [22].

Assuming a camera resolution of 640 by 480 pixels, a thermogram like that in Figure 1, with a field of view estimated at 300 mm wide, roughly has a size of 0.5 mm per pixel (2 pixels/mm). This meets the criteria set in IEC 80601-2-59 (2019) which suggests that for optimal analysis there should be at least 1 pixel per mm. This implies that several pixels will cover the inner(medial) canthus, as required, providing a reliable measurement for this area. However, when considering situations observed, e.g., at airports, cameras are often aimed at a stream of people rather than an individual's face and cover a field of view with a width of several meters. Taking an example of 3 meters with the same camera and lens, each pixel covers around 4.5 mm (0.22 pixels/mm). Thus, no individual pixel will be representing *only* the inner eye-canthus, making the measurement more error

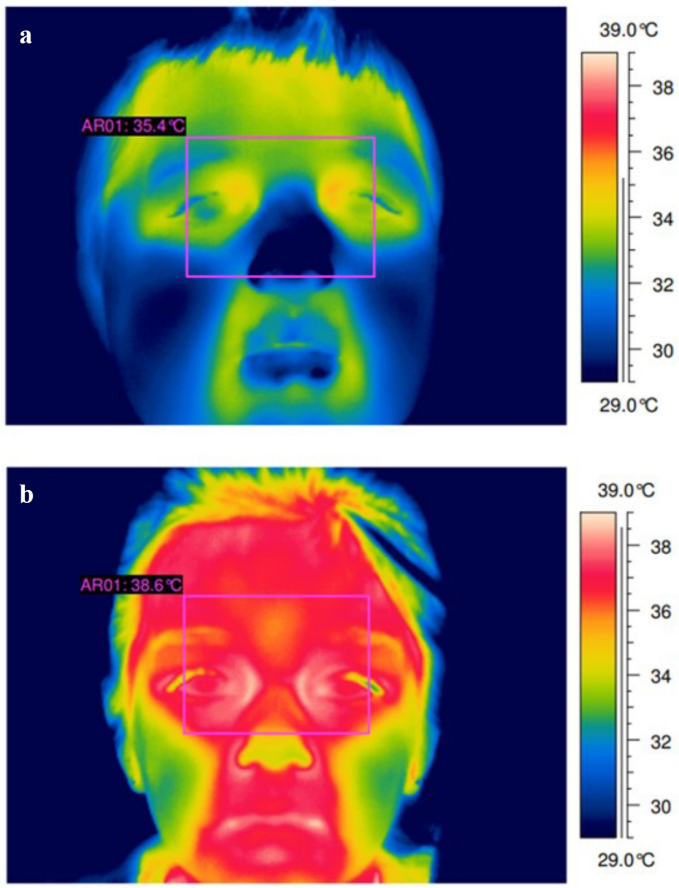

Figure 1. Two 10-year-old males, the top thermogram (a) **shows a healthy (non-febrile) case. The bottom thermogram** (b) **shows a case of fever. Thermogram A shows a max temperature of 35°C at the inner eye-canthus. Thermogram B shows a max temperature of 38.6°C at the inner eye-canthus. Figure adapted with permission of QIRT Council** [24].

prone. Budzan and Wyzgolik [23] observed a reduction of 1.6°C in their assessment of inner canthus temperature going from 1 to 3 meters distance (384 × 288-pixel uncooled FPA microbolometer camera). Lens angle is not provided, but field of view is estimated at 70 cm at 1 m (0.55 pixels/mm) to 140 cm at 2 m (0.28 pixels/mm). Note that neither of the conditions in this experiment meets the IEC advised resolution.

External temperature reference source

The standard [13] defines the external temperature reference source (ETRS) as part of the screening thermograph that is used to ensure accurate operation between calibrations of an NCIT device. Many systems now include an internal reference temperature, with some manufacturers suggesting that external checks are not required. However, unless frequent servicing is obtained, it is still advisable to use an ETRS where possible, particularly to check for drift in the temperature sensitivity of the camera over time [18]. An ETRS, a blackbody calibrator placed in the field of view of the camera, allows the operator to make frequent checks on the camera and infrared thermometer. Software integrated with thermal cameras can also make self-corrections based on the difference between measured and true ETRS temperature, to ensure stability and minimize drift over a given time-period. As stated in clause 201.101.3 of the standard [13], the size of the ETRS in the image should be ≥20 × 20 pixels and it should provide a stability and drift error of no more than ±0.1° C across an assessment interval of 14 days to be useable as reference measurement. The size of the ETRS in the screening thermograph should be sufficiently large so that the screening thermograph's measurement is not affected by its small size and to allow a clear identification of the ETRS area within the target plane. The ETRS should not be larger than 10% of the face to not adversely affect the infrared camera, though it has been suggested that sizes up to 20% may still work ok [24]. The ETRS should be set at a temperature close to the threshold for detecting fever (i.e., ~35°C).

Accuracy

The accuracy of an NCIT device defines the potential difference between the measured temperature and the true temperature and is expressed in degrees or as a percentage error, or both. The standard [13] advocate that NCIT devices should have an accuracy specification of ± 0.5°C or lower over the range of 34°C to 39°C. Device accuracy can be maintained by ensuring stability in the environmental temperature surrounding the NCIT device, correct use of an ETRS, and allowing the device to warm up to ambient conditions before use. If the optics or the inside of the camera body are changing temperature, this can offset the temperature measurement. Many NCIT devices are advertised to have an accuracy of ± 1°C or 2° C, which does not meet the specifications advocated in the standard. However, some devices designed for human temperature screening are calibrated only within the relevant T_{skin} ranges (i.e., between 30°C and 40°C), which improves the accuracy to within 0.3°C. However, this error is the best-case scenario and can be inflated by poor operating procedures. During the COVID-19 pandemic, many devices not specifically designed for fever screening have been advertised and sold for this purpose [12]. Often these devices are developed for the building industry with insufficient specifications for fever screening (https://www.gov.uk/government/news/dont-rely-on-temperature-screening-products-for-detection-of-coronavirus-covid-19-says-mhra).

Device

Thermal imaging (as opposed to a spot measurement) is the recommended approach for the non-contact assessment of internal temperature [ISO13154, 13, 14]. The main reason this approach is preferred over a spot measurement is the ease in which the inner canthus temperature can be assessed (see above section on "measurement site"). Accurately determining inner canthus temperature using a spot device is problematic due to the requirement to be in very close proximity to the test subject to avoid issues with spot-distance ratios (see above section on "spot distance ratio"). Using a thermograph, the

inner canthus can be measured relatively easily but should meet the resolution requirements set forth in the ISO standard [13] and noted above. The financial implications of using a suitable thermal imaging camera are significant to most organizations, hence why spot measurements are often utilized for mass screening. If a spot measurement is to be used, the center of the forehead is the most suitable measurement site.

Some devices are advertised to be used specifically for fever screening based on elevated T_{skin}. One of the main advantages is that these devices are calibrated within a narrower range, which can improve device accuracy to within specifications noted in the ISO standard of ± 0.3°C [13] (see above section on "accuracy").

Threshold temperature

The threshold temperature is the temperature setting above which the target is potentially febrile [ISO13154, 14]. The ISO standard does not stipulate a precise threshold temperature and instead states that *"the responsible organization should consult their medical advisor on the setting of the threshold temperature"*. One difficulty with advising a threshold temperature is that a single temperature will yield different performance characteristics. If the threshold temperature is too low, it will result in many false positives. If it is set too high, it may result in many false negatives, contributing to the spread of infections. In a clinical setting, a forehead temperature threshold of 35.1°C provides a good balance of false positive and false negatives [5]. For the inner canthus temperature, a threshold temperature of 37.5°C is recommended in a clinical setting [25]. Given the limited accuracy of many NCIT devices in use in the field, a fixed threshold may represent quite different temperatures in reality. Some devices operate in an "adjusted mode", which means the T_{skin} is converted to an internal body temperature with a proprietary algorithm. In that case, fever is defined as a core temperature exceeding 38°C.

Relative temperature screening

Due to potential accuracy issues with NCIT as well as impact of local conditions (e.g., climate), several manufacturers of NCIT screening systems (e.g. FLIR and Testo) advocate using a relative rather than an absolute temperature threshold approach (see https://www.flir.co.uk/discover/public-safety/flir-screen-est-how-flir-screening-solutions-provide-easy-efficient-and-accurate-measurement/or https://www.testo.com/en-UK/products/thermography-fever-detection). Using a relative approach means that a group of healthy individuals' face/inner canthus temperature is used as a baseline reference/control temperature, circumventing the device's accuracy issue with an *absolute* temperature measurement. Thereafter, if another individual is, e.g., more than 1° C above the group mean reference, the camera's alarm will indicate a T_{skin} that is elevated above normal. The user can specify the acceptable deviation above the reference value. FLIR recommends updating the group mean reference temperature with healthy samples every 15 minutes, presumably to account for diurnal variation in body temperature, and potential changes to the environment. An example of how this technique works is shown in (Figure 2) using simulated data from 8am to 6pm.

To our knowledge, while the reasons for a relative approach are clear, there has been no study investigating the performance of NCIT for detecting fever using a relative approach. A comparative study investigating the screening efficacy of NCIT when using absolute vs relative thresholds would clearly be beneficial. Limitations of using a relative approach include i) the requirement to regularly update the average with healthy individuals (though if prevalence is low, infected individuals appearing in the average would have low impact, i.e., a running average could be used of all people tested, if infection rates are low) ii) determining what constitutes an individual as "healthy" without a pathogen-specific diagnostic test, and iii) the requirement of using an expensive thermal imaging camera.

Part 2. physiological confounders

Thermoregulatory confounders in fever

There are important thermophysiological considerations which may limit the utility of fever screening with NCIT. Fever is a highly conserved mechanism through which many organisms (including humans) combat infection from invading pathogens. Part of the febrile syndrome involves a rise in T_{core} >38°C,

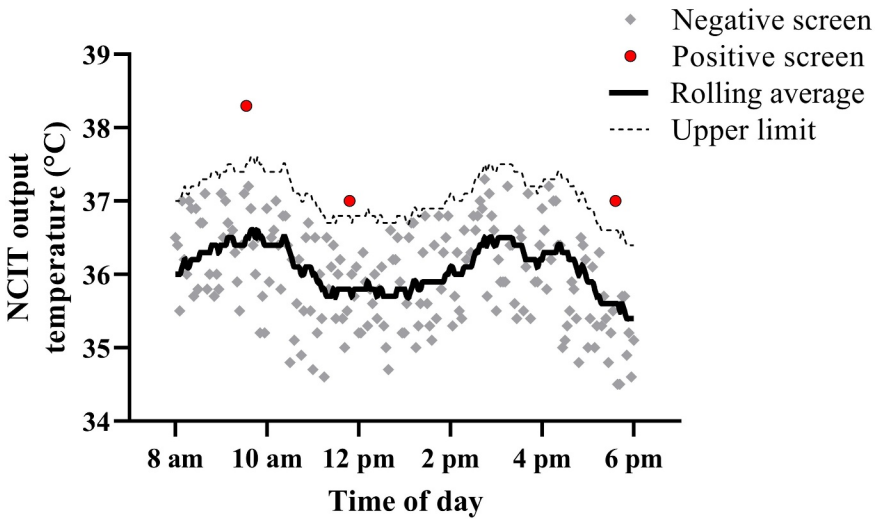

Figure 2. Simulated data to show the relative temperature screening method. A rolling average (solid black line) of all data points is used as the reference temperature. A temperature threshold of 1°C above this rolling average (dashed line) would result in a positive temperature screen, as shown by the red circular symbols. All data points below this threshold (gray symbols) would screen negative.

which decreases survivability and function of bacteria and reduces viral replication [26]. Consequently, detecting an increased T_{core} by NCIT screening is an appealing method to detect those who may be fighting an infection.

Core temperature

NCIT screening for the detection of infection assumes a constant, raised T_{core} which is also reflected by a raised T_{skin}. There are several issues with this assumption. Firstly, the T_{core} response to infection depends on the type of pathogen in question and the time since infection. For example, COVID-19 typically has a 5-day incubation period before any symptoms [i.e., fever) become apparent. A raised T_{core} will therefore not be detected in individuals during the incubation period, despite these individuals being infectious already, since the immune/febrile response has not yet been activated. Second, the T_{core} pattern during immune activation is not consistent across pathogen types and infection loads. Roth and Blatteis [27] eloquently demonstrate a heterogeneous T_{core} pattern in response to various types of pathogens. For example, Typhus infection produces a sustained rise in T_{core} over several days, Malaria produces a biphasic response, pleuritis and sepsis produces a triphasic response, and Tuberculosis seems to produce a sinusoidal like T_{core} response. Moreover, the T_{core} pattern to *the same* pathogen seems to depend greatly on the load/dose, where a very high dose of lipopolysaccharide (LPS] produced a strong *hypothermic* response [26,28]. The T_{core} response (T_{core} peak and time course) to COVID-19 infection to our knowledge has not been documented and data are urgently required.

Skin temperature

An important question regarding the current paper and utility of NCIT screening is the T_{skin} response to fever, which is seldom documented in humans. For maximum utility of NCIT, the T_{skin} pattern should broadly follow the T_{core} pattern, or at least experience a sustained rise throughout some stage of the febrile response. In a normothermic environment, resting T_{skin} is primarily controlled by the rate of cutaneous/skin blood flow, where a higher SkBF (and therefore T_{skin}) mediates a greater rate of dry heat loss to the environment. In a warm environment and during exercise, skin blood flow is elevated to ensure increased rates of core to skin heat transport, pushing more heat to the skin. However, when sweating is initiated, being a very powerful cooling mechanism when evaporating, the net effect is that typically skin temperature is lowered. Skin blood flow and skin temperature are reduced in a cold environment (decreasing T_{skin}) to reduce heat loss from the skin. The aim of the febrile response is to

increase T_{core}, and as such, it seems intuitive that skin blood flow, and thus T_{skin}, would *decrease* to limit heat loss to the environment. Such findings are supported in animal models of experimental fever, such as a reduction in ear temperature of the rabbit or tail temperature of the rat and mouse in the early phase where T_{core} exhibits its first peak [29–31]. However, a reduction in *human* T_{skin} during fever (which would severely limit the utility of NCIT scanning) does not seem to occur in all sites. The human body temperature responses to infection were documented after administration of the US Army triple typhoid vaccine [32]. The severity of fever varied between participants, but unlike the data in mice, there was no fall in mean T_{skin}, which rose continuously and peaked at 36.7°C in the 29°C environment, 2 hours after administration. T_{core} rose in parallel, peaking close to 40°C. Thereafter, T_{skin} and T_{core} fell to baseline values within around 6 hours and then plateaued. It seems that in humans, the rise in T_{core} is mediated primarily by increased metabolic activity, whereas in mice, more marked reductions in skin blood flow are required. Although T_{skin} (measured over 10-sites) during severe fever did not reach the typical 37.5°C threshold as advised in mass screening studies [33], a single site that better reflects T_{core} (i.e. inner eye-canthus) would likely, albeit speculatively, have surpassed such a threshold in this case. Additional studies in humans generally find a reduction in hand blood flow in response to systemic endotoxin administration, with minimal T_{skin} change at other sites [33, 34].

Taken together, data in febrile humans [25,27, 32,34,35] indicate that core and skin temperature responses vary between participants, environments, and pathogen. Hand skin temperature is likely to decrease during fever development, but the response at other sites is difficult to predict with any confidence. Data using NCIT for fever screening in children indicate that facial temperatures (i.e forehead and inner canthus) can provide a good estimate of a raised T_{core} under well-controlled conditions. However, data is otherwise severely limited on the facial temperature response to fever, especially related to its time course across specific infections and the impact of antipyretic drugs on the temporal pattern.

Activity

Any individual screened for elevated body temperature should be in as close to a resting state as possible. There are two reasons why this is important. Firstly, physical activity requires an increase in metabolic rate, which in turn increases heat production in the body. Moderate to heavy physical activity can increase T_{core} to levels similar to that of fever [36], even in cool environments. The required activity level to reach these values is further reduced with high clothing coverage (such as with a business suit or insulating jacket). A recent review published in this journal highlights the need to separate fever from hyperthermia when screening for fever using NCIT [37]. They suggest using a 30-minute resting period following cessation of work or exercise before screening for fever, due to the impact of activity on T_{core}. Secondly, moderate to heavy activity can induce a sweating response. When sweat evaporates from the skin surface it provides a local cooling effect, and the T_{skin} at a specific region of interest will begin to deviate from the core temperature. Sweating has been shown to diminish the ability of NCIT to track forehead [38] and inner canthus temperature [39].

Makeup and eyeglasses

Heavy face and eye makeup may change the thermal properties of forehead and inner canthus. Similarly, eyeglasses are likely to create a microclimate between the inner canthus and the lens, which may impact the output temperature even if they are temporarily removed. To our knowledge, however, this has not been specifically investigated and the implications for NCIT are unknown.

Age

Age has been shown to significantly impact output temperatures from NCIT [10]. T_{skin} is dictated by the environment, core temperature and rate of skin blood flow [40]. Healthy aging has been shown to cause alterations in skin blood flow in the resting state and during exercise or

thermal stress. During rest, skin blood flow in older participants is increased compared with their younger counterparts, partly due to reductions in skin sympathetic nerve activity [41], resulting in warmer skin at rest. That older individuals typically present with warmer skin in the resting state means that i) they are more likely to screen positive for high T_{core} in the absence of fever, and ii) their relationship between T_{core} and T_{skin} is not the same as younger individuals.

Immunosenescence refers to the gradual deterioration of the immune system brought on by natural aging [42]. In the context of fever, aging decreases endotoxin-induced production of prostaglandin E_2 (PGE_2) in the hypothalamus, and PGE_2 initiates the thermoregulatory response to infection [43]. Consequently, aging (typically over 65 years) impacts the ability of humans to mount a febrile response to infection, which reduces the peak T_{core} attained [44–46]. Recent evidence supports lower T_{core} thresholds for COVID-19 screening in nursing homes [47], suggesting that a "one size fits all" threshold temperature is not optimal for all age groups. However, there exists a lack of evidence relating to how the threshold temperature varies with aging i.e., is there a linear or exponential reduction in maximum febrile T_{core}, and how does this vary based on the pathogen and dose. More research is required to generate an age adjusted temperature threshold but is an important avenue for future work.

Antipyretic drugs

Antipyretic drugs are a class of medications used to decrease core body temperature during fever, and therefore have the potential to mask elevated T_{core} in otherwise symptomatic individuals. In both children and adults, treatment with acetaminophen and/or ibuprofen lowers core temperature to close to normal physiological values during fever [48,49]. There also exists evidence that acetaminophen mildly reduces T_{core} in non-febrile individuals [50–52], but the effect is seen primarily in cool/cold climates [53]. In the climate zones advocated in (Table 1), acetaminophen is unlikely to exert a meaningful effect on T_{core} but should be accounted for in physiological studies which attempt to minimize daily variation in T_{core}.

Part 3. can non-contact readings track human internal/core temperature?

Measurement of resting T_{core} is required to diagnose fever. In the clinical setting, a resting T_{core} > 38°C is typically used to screen for fever, ideally supported with additional assessment of white blood cell counts to confirm an active immune response [52]. The gold standard measurement tissues for assessing T_{core} include rectal and esophageal temperature but these measurements are typically not practical outside of a laboratory environment. Consequently, tympanic or oral temperature is more often adopted in either a clinical or public setting. Despite high device accuracy, tympanic and sublingual (oral) temperature measurements are prone to ambient influences, such as air temperature and wind [54,55]. In the context of fever screening, these diagnostic tools are used to help *confirm* febrile status in individuals who screen positive for elevated temperature with NCIT.

Attempts to assess T_{core} from the T_{skin} *with contact* have had limited success, with researchers concluding *"the idea that it is possible to create a universally usable non-invasive heat strain monitor may be unachievable"* [56]. However, this work excluded measurements at the head, such that these conclusions may not necessarily apply to a range of possible temperature measurement sites that could be used for fever screening. In the following section, we discuss whether NCIT measurements on the forehead ($T_{forehead}$) and inner eye-canthus ($T_{canthus}$) have utility for tracking T_{core}.

Forehead temperature

$T_{forehead}$ is a commonly utilized site to identify a fever. It is an attractive site due to its easy access, minimal interaction with clothing, and very close proximity to brain tissue. Parents all over the world commonly rely on $T_{forehead}$ (assessed by palpation) as a first indication whether their child has a fever, and NCIT of this site is common practice for fever screening [57]. Furthermore, in a resting state and in normothermic conditions, the heat content of the forehead is impacted by the blood flow and metabolic rate of the cerebrum [58]. A differentiation is required between static and dynamic conditions. Here, we define a static

condition as one where ambient temperature is stable, and the participants tested are well rested. A dynamic condition can involve sub-optimal environments (heat or cold), and/or active participants. For example, an indoor, hospital environment is generally of a static nature, whereas an airport or outdoor environment is dynamic. Below will detail why this distinction is necessary.

Agreement between NCIT of $T_{forehead}$ and a reference T_{core} (tympanic) has been assessed in the static hospital environment [5,59–62]. Typically, $T_{forehead}$ cannot be used to accurately predict T_{core} in non-febrile patients due to wide variations in $T_{forehead}$ when T_{core} was between 36.5°C and 37.5°C [5]. However, when tympanic temperature > 38°C (i.e. in the febrile range), $T_{forehead}$ correlates well with tympanic temperature and is suitable for fever screening purposes [5,59,61]. Therefore, the consensus among hospital-based studies is that $T_{forehead}$ does not *always* track T_{core} and therefore cannot replace contact methods for an accurate assessment, but it appears that $T_{forehead}$ is useful for *screening* purposes, preceding a more invasive contact assessment of T_{core} [5,61]. As mentioned previously, a threshold temperature of 35.1°C was suggested for fever screening [5].

For NCIT of the forehead to be effective *outside* of the hospital environment, it must adequately follow T_{core} under dynamic conditions. In the relevant scenarios for the current COVID-19 pandemic, mass screening takes place in environments such as airports, harbors, restaurants, retail outlets, gyms, etc. In these mass screening situations, people may arrive from the cold or heat, and are unlikely to be in a well-rested, inactive state. These conditions thus differ from the relatively stable and temperate conditions expected in a clinical setting. Assessing $T_{forehead}$ in an outdoor/uncontrolled setting yields poor performance metrics and is not suitable for fever screening or tracking T_{core} marked by a high false negative rate (low sensitivity) [63]. These findings are supported by 17, who found outdoor temperature to be a significant confounder which affects the relationship between $T_{forehead}$ and tympanic temperature. That $T_{forehead}$ has a high false negative rate when used outdoors is alarming from a public health perspective since it is less likely to identify febrile individuals. The advice in Table 1 should be consulted to ensure optimal room conditions.

Using an exposure involving resting and exercise conditions, Kistemaker and colleagues [64] demonstrated that $T_{forehead}$ deviates significantly from T_{core} during activity, but follows it well during resting conditions. During passive heat stress where T_{core} was linearly elevated over 1 hour up to 38°C, $T_{forehead}$ instead *decreased* throughout the exposure, clearly failing to track T_{core} [37]. Furthermore, NCIT assessment of $T_{forehead}$ failed to detect hyperthermia in marathon runners [65] and during occupational heat exposure with heavy protective clothing [66]. Hence, these data support the notion that during activity or environmental heat or cold, $T_{forehead}$ cannot predict T_{core} and therefore may be unsuitable for fever screening. However, it is unclear if the $T_{forehead}$ would show better correlation with T_{core} in these dynamic situations in febrile individuals, since those mentioned in the preceding studies involving activity [66] and passive heat exposure [37] were from a healthy cohort. Although replicating a fever situation was not the intention of these studies, it is worth noting that exercise and passive heat stress are not good models for replicating fever, due to differences in blood flow distribution and sweat rates [26,67]. Overall, in a static environment, the correlation is between $T_{forehead}$ and T_{core} is poor in healthy people, but the relationship improves when individuals become febrile [5]. Therefore, in well-controlled settings, $T_{forehead}$ appears to be a suitable tool to screen for fever, but not to reliably track T_{core} across both the healthy and febrile range.

Inner eye-canthus temperature

The inner canthus of the eye typically represents the warmest spot on the face and is therefore considered the most suitable site for tracking T_{core} [4]. Assessment of inner canthus temperature ($T_{canthus}$) requires a *thermal imaging camera*, unlike $T_{forehead}$ which only requires a spot measurement. In a study involving 191 children (18 febrile), $T_{canthus}$ was the best predictor of fever compared with $T_{forehead}$ and tympanic temperature [25]. In that study, fever was diagnosed with an axillary temperature > 37.6°C, a site

which is relatively stable in varying ambient conditions [16]. In support, data from facial temperatures of nonfebrile (healthy) children show that $T_{canthus}$ is the least variable skin site compared with the $T_{forehead}$ and nose, which supports its use for T_{core} tracking [68]. The International Standardization Committee (IEC 80601-2-59, 2020 [13]) suggest using this site, based on a study comparing several T_{core} and T_{skin} measurement sites under different ambient conditions [16]. In that study, $T_{canthus}$ was stable between 21°C and 26°C ambient temperature, but dropped significantly at 15°C, highlighting the importance of using a thermo neutral environment during assessment. The optimal room conditions are highlighted in Table 1.

Using the absolute temperature of this region for fever screening carries the same issues as assessing $T_{forehead}$ in a dynamic setting. [38],compared temperature responses of $T_{canthus}$ to T_{core} during one exposure involving rest, exercise, recovery, and passive heating. They found T_{core} and $T_{canthus}$ to differ significantly in all conditions by >1°C, with differences between the two measurements exceeding 3°C during exercise. Such findings have been supported by more recent studies using a similar design [69,70], both of which suggest that $T_{canthus}$ assessed by NCIT is a poor predictor of T_{core} in a dynamic situation.

Overall, $T_{canthus}$ is a poor predictor of absolute T_{core} in dynamic conditions, but is more stable than $T_{forehead}$ in static, non-exercise conditions. Users should consult the checklist in Table 1 to ensure correct operation of the camera and ensuring room conditions are adequate for an accurate measurement $T_{canthus}$ measurement. An absolute threshold of 37.5°C has been suggested for fever screening [25], though this needs to be considered together with technical limitations of thermal cameras, as discussed in Part 1.

Summary on absolute T_{core} prediction from NCIT monitors

Overall, there is good evidence to suggest that under well controlled, stable conditions (both environmental and physiological), there is some utility for NCIT in the screening of T_{core}. If absolute thresholds are to be adopted, the data suggest that 35.1°C and 37.5°C are appropriate for the forehead and inner eye-canthus, respectively. That NCIT performs very poorly under more dynamic conditions is not a trivial issue. For almost all cases where mass fever screening is used, the logistics required to obtain such a stable measurement will strongly prohibit its utility. $T_{forehead}$ and $T_{canthus}$ are strongly impacted by the environment, sweat, activity level, and time of day, and based on the evidence, are not useful for tracking T_{core} in the conditions which they are currently used (i.e., non-stable environments and/or non-rested humans). Whether relative temperature screening is effective for reducing the confounding effect of temperature fluctuations and time of day remains to be determined, but it cannot solve the impact of activity issue. Moreover, it is unclear to what extent common factors such as makeup, eyeglasses, face masks, and menstrual cycle phase have on the temperature received from NCIT.

Future directions

A global initiative with the objective to improve body temperature measurement, primarily by infrared methods, has been started from the Consultative Committee of Thermometry (CCT; International Committee of Weights and Measures) [71]. The CCT has established a Task Group for Body Temperature Measurement whose objective is to establish reliable clinical thermometry on a global basis and whose initial focus will be to improve infrared methods of body temperature measurement (ear, forehead, thermal imaging). The task group will aim to achieve this objective through the following five actions: 1) Lead a global comparison of calibrators for a range of infrared body temperature thermometers, including aural, forehead and large area thermal imagers, 2) identify current best practice and develop recommendation and guidelines for use of body temperature thermal imaging in a) health services b) airport and other screening situations around the world, 3) identify current best practice of infrared body temperature measurement (ear, forehead), and develop best practice recommendations, 4) Review standards and collaborative work with appropriate standardization bodies (e.g. ISO/IEC) concerned with producing standards for body temperature measurement devices, and 5) Establish metrology, medical and manufacturer forums within the metrology regions to identify the problems with

the current approaches to body temperature measurement and develop practical solutions and establish appropriate links to the World Health Organization. Based specifically on our report, there is i) a need to determine the performance of NCIT for fever screening using relative vs absolute temperature thresholds, ii) a need to model how optimal temperature thresholds change as a function of aging, and iii) a need to determine the impact of eyeglasses and makeup on the ability of NCIT to track human T_{core}.

Conclusions

Skin temperature assessment with non-contact infrared thermometry can sufficiently track core body temperature, but only with appropriate technology and under standardized conditions. At present, non-contact infrared thermometry has performed poorly for mass fever screening at border crossings, and may be due to poor adoption of the international standard guidelines [ISO13154, 13, 14]. Under standardized conditions, NCIT assessment of either the forehead or inner eye-canthus has utility for fever screening but cannot replace conventional methods of internal temperature assessment. We recommend using the checklist provided in Table 1 before using NCIT for fever screening or skin temperature measurement.

In addition to the discussion in the main text, we refer the reader to the supplementary online material with this paper, specifically discussing the performance of NCIT for mass fever screening. The performance of NCIT is based on metrics such as sensitivity, specificity, and the positive and negative predictive values. All of which are explained (including their calculation) in the supplementary file.

Abbreviations

CCT	Consultative Committee of Thermometry
ETRS	external temperature reference source
IEC	International Electrotechnical Commission
ISO	International Standardization Organization
NCIT	non-contact infrared thermometry
PGE_2	prostaglandin E_2,
SARS	severe acute respiratory syndrome
$T_{canthus}$	inner eye canthus temperature
T_{core}	core body temperature
$T_{forehead}$	forehead temperature
T_{skin}	skin temperature

Acknowledgments

We would like to thank the reviewers for their helpful comments which improved the quality of the paper. Funding was provided by 'HEAT-SHIELD', European Union's Horizon 2020 research and innovation programme under grant agreement no. 668786.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the H2020 European Research Council [668786].

ORCID

Josh Foster http://orcid.org/0000-0001-6232-9336
Alex Bruce Lloyd http://orcid.org/0000-0003-0657-3582
George Havenith http://orcid.org/0000-0001-6223-4265

References

[1] Bakhsheshi MF, Lee TY. Non-invasive monitoring of brain temperature by near-infrared spectroscopy. Temperature. 2015;2(1):31–32. doi:10.4161/23328940.2014.967156
[2] Raiko J, Koskensalo K, Sainio T. Imaging-based internal body temperature measurements: The journal Temperature toolbox. Temperature. 2020;7(4):363–388. doi:10.1080/23328940.2020.1769006
[3] Dong D. Airport fever screening: useful or futile? Correctly used infrared thermography is needed. Sci. 2020;80.
[4] Mercer JB, Ring EFJ. Fever screening and infrared thermal imaging: concerns and guidelines. Thermol Int. 2009;19:67–69.
[5] Ng DK, Chan CH, Lee RS, et al. Non-contact infrared thermometry temperature measurement for screening fever in children. Ann Trop Paediatr. 2005;25(4):267–275.
[6] Fournet D, Ross L, Voelcker T, et al. Body mapping of thermoregulatory and perceptual responses of males and females running in the cold. J Therm Biol. 2013;38(6):339–344.
[7] 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. J Therm Biol. 2017;69:155–162.
[8] St. John RK, King A, De Jong D, et al. Border screening for SARS. Emerg Infect Dis. 2005;11(1):6–10.

[9] Mouchtouri VA, Christoforidou EP, Der Heiden MA, et al. Exit and entry screening practices for infectious diseases among travelers at points of entry: looking for evidence on public health impact. Int J Environ Res Public Health. 2019;16(23):1–53.
[10] Nguyen AV, Cohen NJ, Lipman H, et al. Comparison of 3 infrared thermal detection systems and self-report for mass fever screening. Emerg Infect Dis. 2010;16 (11):1710–1717.
[11] Mouchtouri VA, Bogogiannidou Z, Dirksen-Fischer M, et al. Detection of imported COVID-19 cases worldwide: early assessment of airport entry screening, 24 January until 17 February 2020. Trop Med Health. 2020;48(1):79.
[12] Howell KJ, Mercer JB, Smith RE. Infrared thermography for mass fever screening: repeating the mistakes of the past? Thermol Int. 2020;30:1–5.
[13] IEC 80601-2-59. International standardisation organisation. medical electrical equipment — part 2–59: particular requirements for basic safety and essential performance of screening thermographs for human febrile temperature screening. (2020)
[14] ISO13154. International standardisation organisation. medical electrical equipment – deployment, implementation and operational guidelines for identifying febrile humans using a screening thermograph. 2017.
[15] Havenith G, Lloyd AB. Counterpoint to "Infrared cameras overestimate skin temperature during rewarming from cold exposure. J Therm Biol. 2020;92:102663.
[16] Pascoe DD, Fisher G. Comparison of measuring sites for the assessment of body temperature. Thermol Int. 2009;19:35–42.
[17] Hausfater P, Zhao Y, Defrenne S, et al. Cutaneous infrared thermometry for detecting febrile patients. Emerg Infect Dis. 2008;14(8):1255–1258.
[18] Ring EFJ, Ammer K. The technique of infrared imaging in medicine. Thermol Int. 2000;10:7–14.
[19] Ring E, Thomas R, Howell K. Sensors for medical thermography and infrared radiation measurements. In: Jones D, editor. Biomedical Sensors. New York: Momentum Press; 2009;1-10.
[20] Maniar N, Bach AJE, Stewart IB, et al. The effect of using different regions of interest on local and mean skin temperature. J Therm Biol. 2015;49:33–38.
[21] Priego Quesada JI, Martínez Guillamón N, De Anda RMCO, et al. Effect of perspiration on skin temperature measurements by infrared thermography and contact thermometry during aerobic cycling. Infrared Phys Technol. 2015;72:68–76.
[22] Ring EFJ, Ammer K. Infrared thermal imaging in medicine. Physiol Meas. 2012;33(3):R33.
[23] Budzan S, Wyzgolik R. Face and eyes localization algorithm in thermal images for temperature measurement of the inner canthus of the eyes. Infrared Phys Technol. 2013;60:225–234.
[24] Ghassemi P, Joshua Pfefer T, Casamento JP, et al. Best practices for standardized performance testing of infrared thermographs intended for fever screening. Plos One. (2018).
[25] Ring EFJ, Jung A, Zuber J, et al. Detecting fever in polish children by infrared thermography. In 9th International Conference on Quantitative InfraRed Thermography; 2008; Krakow, Poland
[26] Garami A, Steiner AA, Romanovsky AA Fever and hypothermia in systemic inflammation. In: Romanovsky AA, editor. Handbook of clinical neurology. Elsevier B.V; 2018. p. 565–597.
[27] Roth J, Blatteis CM. Mechanisms of fever production and lysis: lessons from experimental LPS fever. Compr Physiol. 2014;4:1563–1604.
[28] Romanovsky AA, Kulchitsky VA, Akulich NV, et al. First and second phases of biphasic fever: two sequential stages of the sickness syndrome? Am J Physiol - Regul Integr Comp Physiol. 1996;271(1):244–253.
[29] Fabricio ASC, Rae GA, D'Orléans-Juste P, et al. Endothelin-1 as a central mediator of LPS-induced fever in rats. Brain Res. 2005;1066(1–2):92–100.
[30] Ivanov AI, Patel S, Kulchitsky VA, et al. Platelet-activating factor: a previously unrecognized mediator of fever. J Physiol. 2003;553(1):221–228.
[31] Riedel W. Mechanics of fever. J Basic Clin Physiol Pharmacol. 1990;1(1–4):291–322.
[32] Palmes ED, Park CR. The regulation of body temperature during fever. Arch Environ Health. 1965;11 (6):749–759.
[33] Bitar D, Goubar A, Desenclos JC. International travels and fever screening during epidemics: a literature review on the effectiveness and potential use of non-contact infrared thermometers. Euro Surveill. 2009;14:1–5.
[34] Bryce-Smith R, Coles DR, Cooper KE, et al. The effects of intravenous pyrogen upon the radiant heat induced vasodilatation in man. J Physiol. 1959;145(1):77–84.
[35] Buskirk ER, Thompson RH, Rubenstein M, et al. Heat exchange in men and women following intravenous injection of endotoxin. J Appl Physiol. 1964;19 (5):907–913.
[36] Lind AR. A physiological criterion for setting thermal environmental limits for everyday work. J Appl Physiol. 1963;18(1):51–56.
[37] Daanen H, Bose-O'Reilly S, Brearley M, et al. COVID-19 and thermoregulation-related problems: practical recommendations. Temperature. 2021;8(1):1–11. doi:10.1080/23328940.2020.1790971
[38] Low DA, Vu A, Brown M, et al. Temporal thermometry fails to track body core temperature during heat stress. Med Sci Sports Exerc. 2007;39(7):1029–1035.
[39] Teunissen LPJ, Daanen HAM. Infrared thermal imaging of the inner canthus of the eye as an estimator of body core temperature. J Med Eng Technol. 2011;35(3–4):134–138.
[40] Havenith G. Individualized model of human thermoregulation for the simulation of heat stress response. J Appl Physiol. 2001;90(5):1943–1954.

[41] Grassi G, Seravalle G, Turri C, et al. Impairment of thermoregulatory control of skin sympathetic nerve traffic in the elderly. Circulation. 2003;108(6):729–735.
[42] Makinodan T, Good RA, Kay MMB Cellular Basis of Immunosenescence. In: Makinodan T and Yunis E, editor. Immunology and Aging. Springer US; 1977. p. 9–22.
[43] Bibby DC, Grimble RF. Effect of age on hypothalamic prostaglandin E2 production and fever in response to tumour necrosis factor (cachectin) and endotoxin in rats. Clin Sci. 1991;81(3):313–317.
[44] Bradley SF, Kauffman CA. Aging and the response to salmonella infection. Exp Gerontol. 1990;25(1):75–80.
[45] Lu SH, Chen YC, Chang YC, et al. Effect of age on febrile response in patients with healthcare-associated bloodstream infection. Geriatr Nurs (Minneap). 2013;34(5):366–372.
[46] Marco CA, Schoenfeld CN, Hansen KN, et al. Fever in geriatric emergency patients: clinical features associated with serious illness. Ann Emerg Med. 1995;26(1):18–24.
[47] McConeghy KW, White E, Panagiotou OA, et al. Temperature screening for SARS-CoV-2 in nursing homes: evidence from two national cohorts. J Am Geriatr Soc. 2020;68(12):2716–2720.
[48] Mds E-L, Coppens K, Hunt LP, et al. Randomised controlled trial of combined paracetamol and ibuprofen for fever. Arch Dis Child. 2006;91(5):414–416.
[49] Sarrell EM, Wielunsky E, Cohen HA. Antipyretic treatment in young children with fever: acetaminophen, ibuprofen, or both alternating in a randomized, double-blind study. Arch Pediatr Adolesc Med. 2006;160(2):197–202.
[50] Ayoub SS. Paracetamol (acetaminophen): A familiar drug with an unexplained mechanism of action. Temperature. 2021. doi:10.1080/23328940.2021.1886392.
[51] Foster J, Mauger A, Thomasson K, et al. Effect of acetaminophen ingestion on thermoregulation of normothermic, non-febrile humans. Front Pharmacol. 2016;7:54.
[52] Kasner SE, Wein T, Piriyawat P, et al. Acetaminophen for altering body temperature in acute stroke: a randomized clinical trial. Stroke. 2002;33(1):130–134.
[53] Foster J, Mauger AR, Govus A, et al. Acetaminophen (Paracetamol) induces hypothermia during acute cold stress. Clin Drug Investig. 2017;37(11):1055–1065.
[54] Doyle F, Zehner WJ, Terndrup TE. The effect of ambient temperature extremes on tympanic and oral temperatures. Am J Emerg Med. 1992;10(4):285–289.
[55] Morán-Navarro R, Courel-Ibáñez J, Martínez-Cava A, et al. Validity of skin, oral and tympanic temperatures during exercise in the heat: effects of wind and sweat. Ann Biomed Eng. 2019;47(1):317–331.
[56] Richmond VL, Davey S, Griggs K, et al. Prediction of core body temperature from multiple variables. Ann Occup Hyg. 2014;59(9):1168–1178.
[57] Romanovsky AA. Skin temperature: its role in thermoregulation. Acta Physiol. 2014;210(3):498–507.
[58] Sessler DI, Sessler AM. Experimental determination of heat flow parameters during induction of general anesthesia. Anesthesiology. 1998;89(3):657–665.
[59] Harioka T, Matsukawa T, Ozaki M, et al. "Deep-forehead" temperature correlates well with blood temperature. Can J Anesth. 2000;47(10):980–983.
[60] Kirk D, Rainey T, Vail A, et al. Infra-red thermometry: the reliability of tympanic and temporal artery readings for predicting brain temperature after severe traumatic brain injury. Crit Care. 2009;13(3):R81–R81.
[61] Mogensen CB, Wittenhoff L, Fruerhøj G, et al. Forehead or ear temperature measurement cannot replace rectal measurements, except for screening purposes. BMC Pediatr. 2018;18(1):1–6.
[62] Myny D, De Waele J, Defloor T, et al. Temporal scanner thermometry: a new method of core temperature estimation in ICU patients. Scott Med J. 2005;50(1):15–18.
[63] Liu -C-C, Chang R-E, Chang W-C. Limitations of forehead infrared body temperature detection for fever screening for severe acute respiratory syndrome. Infect Control Hosp Epidemiol. 2004;25(12):1109–1111.
[64] Kistemaker J, Den Hartog E, Daanen H. Reliability of an infrared forehead skin thermometer for core temperature measurements. J Med Eng Technol. 2006;30(4):252–261.
[65] Ronneberg K, Roberts WO, McBean AD, et al. Temporal artery temperature measurements do not detect hyperthermic marathon runners. Med Sci Sports Exerc. 2008;40(8):1373–1375.
[66] Pryor RR, Seitz JR, Morley J, et al. Estimating core temperature with external devices after exertional heat stress in thermal protective clothing. Prehospital Emerg Care. 2012;16(1):136–141.
[67] Romanovsky AA The thermoregulation system and how it works. In: Romanovsky AA, editor. Handbook of Clinical Neurology. Elsevier B.V; 2018; 156:3–43.
[68] Vardasca R, Magalhaes C, Marques D, et al. Bilateral assessment of body core temperature through axillar, tympanic and inner canthi thermometers in a young population. Physiol Meas. 2019;40(9):094001.
[69] Fernandes AA, Moreira DG, Brito CJ, et al. Validity of inner canthus temperature recorded by infrared thermography as a non-invasive surrogate measure for core temperature at rest, during exercise and recovery. J Therm Biol. 2016;62:50–55.
[70] Towey C, Easton C, Simpson R, et al. Conventional and novel body temperature measurement during rest and exercise induced hyperthermia. J Therm Biol. 2017;63:124–130.
[71] Machin G, Xiaofeng L, Del Campo D, et al. Global initiative to improve infra-red based body temperature measurements. Thermol Int. 2020;20:97.

Supplementary information to:

Non-contact infrared assessment of human body temperature: The journal Temperature toolbox,

Josh Foster, Alex Bruce Lloyd & George Havenith (2021)

Temperature, DOI: 10.1080/23328940.2021.1899546

Mass skin temperature screening for fever detection

Data from past research investigating the correlation between forehead and tympanic temperature in 500 individuals (Ng et al., 2005) will be used to clarify the meaning and calculation of sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Figure S1 shows the correlation which is used in this example. The black horizontal dashed line is the threshold $T_{forehead}$ (assessed with NCIT) in which an individual would *screen* positive for fever. The vertical black line is the clinical threshold temperature which defines an individual as being *truly* healthy or *truly* febrile.

In this example, all those who have a $T_{forehead} > 35.1°C$ (dashed line) would screen positive for fever and therefore would require a follow up tympanic or other febrile assessment. All those below the dashed line would be deemed as healthy and not feverish, not requiring a follow up diagnostic test. All those to the left of the vertical line (tympanic temperature 38°C) are *truly* healthy and do not have a fever, whereas all those to the right of the vertical black line *truly* do have a fever and would be considered 'infected'. We can then separate the data into four distinct zones (1-4). Zone 1 shows all individuals who are *truly* healthy but are screened incorrectly as having fever (false positive). Zone 2 shows all individuals who are *truly* healthy and are screened correctly (true negative). Zone 3 shows all individuals who *truly* have fever and are screened correctly (true positive). Zone 4 shows all individuals who *truly* have fever but are screened incorrectly (false negative).

Clarification of terms

The definitions below are written in the context of using NCIT for fever screening only. The "zones" are related to those described above and are shown in Figure S1.

Sensitivity

Sensitivity is the ability of a test to correctly identify febrile individuals. It is calculated from the proportion of true positives (Zone 3) to false negatives (Zone 4):

$$Sensitivity\ (\%) = \frac{True\ positives}{(True\ positives + False\ negatives)} \times 100$$

Specificity

Specificity is the ability of a test to correctly identify healthy, non-febrile individuals. It is calculated from the proportion of false positives (Zone 1) to true negatives (Zone 2):

$$Specificity\ (\%) = \frac{True\ negatives}{(True\ negatives + False\ positives)} \times 100$$

Positive predictive value (PPV)

PPV is the probability that an individual who screens positive for being febrile is *truly* febrile. It is calculated from the proportion of false positives (Zone 1) to true positives (Zone 3):

$$PPV\ (\%) = \frac{True\ positives}{(True\ positives + False\ positives)} \times 100$$

Negative predictive value (NPV)

NPV is the probability that an individual who screens negative for being febrile is *truly* healthy. It is calculated from the proportion of false negatives (Zone 3) to true negatives (Zone 4):

$$NPV\ (\%) = \frac{True\ negatives}{(True\ negatives + False\ negatives)} \times 100$$

Receiver operator characteristic (ROC) curve

An ROC curve is a graphical tool used to determine the overall utility of NCIT for screening purposes. It shows how the relationship between sensitivity and specificity changes as a function of different threshold temperatures with NCIT. The choice of threshold temperature is typically that which produces a simultaneously high sensitivity and specificity, such that the greater the area under the ROC curve, the greater efficacy NCIT has for fever screening. Figure S1B shows an example using the dataset previously described (Ng et al., 2005). In that work, a threshold temperature of 35.1°C produced a sensitivity of 89%, and a specificity of 75%.

Thermal imaging cameras were installed at six major airports in Canada during the SARS epidemic. Of the 467 870 people screened, 95 screened positive for fever, but none had a fever confirmed subsequently from tympanic temperature. The cost of this unsuccessful programme was $CA17m (£9m; €12m; $15m), and it was concluded that NCIT was futile in this instance because the prevalence of SARS was extremely low, yielding a PPV of effectively 0 (St. John et al., 2005). The efficacy of thermal imaging for fever detection at airports was assessed in 1275 airline passengers during the 2009 Influenza A pandemic (Priest et al., 2011). In that study, NCIT yielded a sensitivity and specificity of 86 and 71%, respectively. However, only 0.5% of the study population were febrile (tympanic temperature ≥ 37.8°C, such that PPV was only 1.5% i.e. 98.5% of individuals who screened positive with NCIT were not positive in a tympanic measurement. The authors' concluded that *"NCIT is unlikely to be effective for entry screening of travellers to detect influenza infection with the intension of preventing entry of the virus into a county"*. Bitar et al. (2009) conducted a systematic review involving six fever screening studies (primarily at hospitals), to determine the performance of NCIT for fever screening. All studies used the forehead as the target area, while 3 included additional measures of either the inner canthus, or aural (tympanic) temperature. The striking observation from the analysis was the variability in the results between studies. The sensitivity varied between studies from 4.0 to 89.6%, while the specificity remained high, between 75.4 to 99.6%. The PPV ranged from 0.9 to 81.4%. The source of such variability between studies does not seem to be the measurement site, since in one study, $T_{forehead}$ and $T_{canthus}$ performed equally well for fever detection (Ng et al., 2004). In many of the studies, the ambient temperature is not reported, nor was the prior activity status of the individuals (see methodological and physiological considerations sections above). Moreover, when the

prevalence of fever was artificially set to 1% between all studies (which best reflects the early stages of a pandemic), PPVs were generally below 10% (Bitar et al., 2009), meaning that a 90% individuals will be mistakenly classified as infected. One of the major problems with mass screening is simply related to probability. Mabey et al., (2014) modelled the likelihood of infected individuals being screened correctly for Ebola with NCIT. The incubation period of Ebola is ~9 days, reducing the probability of positive screening even if NCIT was 100% sensitive. Their model, based on a 6.42-hour flight from Freetown to London, showed that if everyone with symptoms were denied entry upon arrival at airports, only 7% of individuals infected with Ebola would be identified. For COVID-19, and depending on the incubation period, it is estimated that 46% of travellers would not be detected using NCIT (Quilty et al., 2020). These calculations do not consider technical issues with NCIT and assume the technique in itself would be adequate.

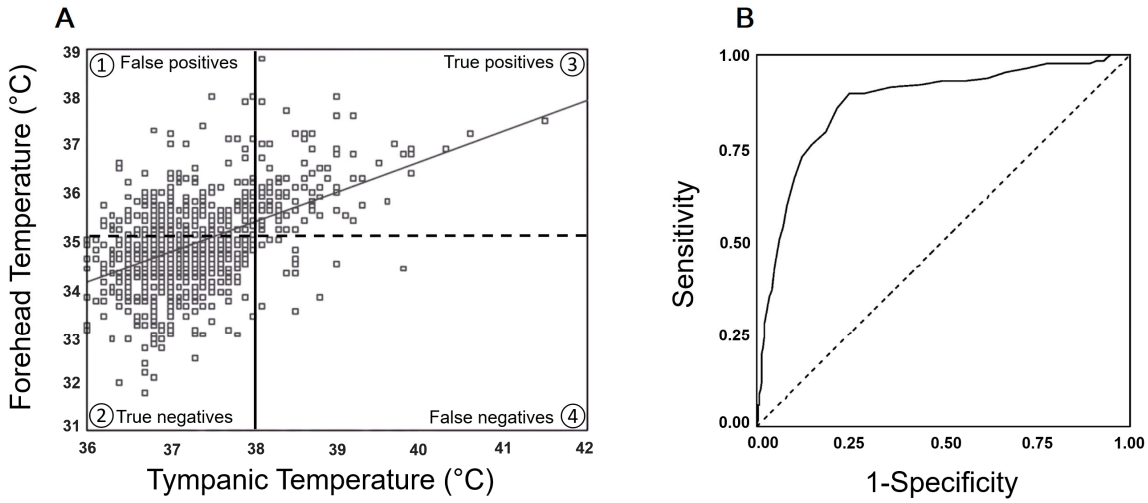

Figure S1. Panel A shows the correlation between forehead temperature and tympanic temperature in a clinical setting. The horizontal (dashed) line is the forehead temperature threshold in which an individual would screen positive for being febrile. The vertical line shows the tympanic temperature threshold where an individual would be deemed truly febrile if they exceeded this value. Data from zones 1 and 2 are used to calculate specificity. Data from zones 3 and 4 are used to calculate sensitivity. Data from zones 1 and 3 are used to calculate PPV. Data from zones 2 and 4 are used to calculate NPV. Panel B shows the associated ROC curve. Figures adapted with permission from (Ng et al., 2005).

References

Bitar, D., Goubar, A., and Desenclos, J. C. (2009). International travels and fever screening during epidemics: a literature review on the effectiveness and potential use of non-contact infrared thermometers. *Euro Surveill.* 14, 1–5. doi:10.2807/ese.14.06.19115-en.

Bryce-Smith, R., Coles, D. R., Cooper, K. E., Cranston, W. I., and Goodale, F. (1959). The effects of intravenous pyrogen upon the radiant heat induced vasodilatation in man. *J. Physiol.* 145, 77–84. doi:10.1113/jphysiol.1959.sp006128.

Buskirk, E. R., Thompson, R. H., Rubenstein, M., and Wolff, S. M. (1964). Heat exchange in men and women following intravenous injection of endotoxin. *J. Appl. Physiol.* 19, 907–913. doi:10.1152/jappl.1964.19.5.907.

Fabricio, A. S. C., Rae, G. A., D'Orléans-Juste, P., and Souza, G. E. P. (2005). Endothelin-1 as a central mediator of LPS-induced fever in rats. *Brain Res.* 1066, 92–100. doi:10.1016/j.brainres.2005.10.037.

Garami, A., Steiner, A. A., and Romanovsky, A. A. (2018). "Fever and hypothermia in systemic inflammation," in *Handbook of Clinical Neurology* (Elsevier B.V.), 565–597. doi:10.1016/B978-0-444-64074-1.00034-3.

Ivanov, A. I., Patel, S., Kulchitsky, V. A., and Romanovsky, A. A. (2003). Platelet-activating factor: A previously unrecognized mediator of fever. *J. Physiol.* 553, 221–228. doi:10.1113/jphysiol.2003.055616.

Mabey, D., Flasche, S., and Edmunds, W. J. (2014). Airport screening for Ebola: Will it make a difference? *BMJ* 349. doi:10.1136/bmj.g6202.

Ng, D. K., Chan, C. H., Lee, R. S., and Leung, L. C. (2005). Non-contact infrared thermometry temperature measurement for screening fever in children. *Ann. Trop. Paediatr.* 25, 267–275. doi:10.1179/146532805X72412.

Ng, E., Kaw, G., and Chang, W. (2004). Analysis of IR thermal imager for mass blind fever screening. *Microvasc. Res.* 68.

Palmes, E. D., and Park, C. R. (1965). The regulation of body temperature during fever. *Arch. Environ. Health* 11, 749–759. doi:10.1080/00039896.1965.10664295.

Priest, P. C., Duncan, A. R., Jennings, L. C., and Baker, M. G. (2011). Thermal Image Scanning for Influenza Border Screening: Results of an Airport Screening Study. *PLoS One* 6, e14490. doi:10.1371/journal.pone.0014490.

Quilty, B. J., Clifford, S., Flasche, S., and Eggo, R. M. (2020). Effectiveness of airport screening at detecting travellers infected with novel coronavirus (2019-nCoV). *Eurosurveillance*. doi:10.2807/1560-7917.ES.2020.25.5.2000080.

Riedel, W. (1990). Mechanics of fever. *J. Basic Clin. Physiol. Pharmacol.* 1, 291–322. doi:10.1515/JBCPP.1990.1.1-4.291.

Ring, E. F. J., Jung, A., Zuber, J., Rutkowski, P., Kalicki, B., and Bajwa, U. (2008). Detecting Fever in Polish Children by Infrared Thermography. in *9th International Conference on Quantitative InfraRed Thermography* doi:10.21611/qirt.2008.03_07_17.

Romanovsky, A. A., Kulchitsky, V. A., Akulich, N. V., Koulchitsky, S. V., Simons, C. T., Sessler, D. I., et al. (1996). First and second phases of biphasic fever: Two sequential stages of the sickness syndrome? *Am. J. Physiol. - Regul. Integr. Comp. Physiol.* 271, 244–253. doi:10.1152/ajpregu.1996.271.1.r244.

Roth, J., and Blatteis, C. M. (2014). Mechanisms of fever production and lysis: Lessons from experimental LPS fever. *Compr. Physiol.* 4, 1563–1604. doi:10.1002/cphy.c130033.

St. John, R. K., King, A., De Jong, D., Bodie-Collins, M., Squires, S. G., and Tam, T. W. S. (2005). Border screening for SARS. *Emerg. Infect. Dis.* 11, 6–10. doi:10.3201/eid1101.040835.