LIMITATIONS OF TEMPORAL (FOREHEAD) TEMPERATURE READINGS AS A SCREENING METHOD FOR COVID-19

Luis Fernando Aragón-Vargas, Ph.D., FACSM
Editor-in-Chief, Pensar en Movimiento
Director, Human Movement Science Research Center
Universidad de Costa Rica, Costa Rica

Published: 2020-06-09
DOI: https://doi.org/10.15517/pensarmov.v18i1.42241

ABSTRACT

As the Covid-19 pandemic continues to evolve, many countries are in the process of reopening schools, churches, and businesses. In an effort to screen for potentially infected individuals, temporal artery (forehead) temperature is measured in thousands of people each day. This editorial discusses some of the major limitations of this procedure as a screening method for Covid-19 and warns against the danger of the false sense of security that this practice may cause.

Keywords: pandemic control measures, fever, exercise, thermoregulation.

At the end of May 2020, the Covid-19 pandemic had resulted in over 6 million confirmed cases and more than 350,000 deaths worldwide (Total confirmed COVID-19 deaths, n. d.). Multiple public-health strategies have been attempted in different regions of the world as authorities try to control spread of the disease, while the SARS-CoV-2 virus responsible for the pandemic proves to be extremely contagious. As many countries attempt the return to business

1 Original version in English. Also available in Spanish in https://revistas.ucr.ac.cr/index.php/pem/article/view/42291
as usual, in what has been called “a new normality”, a key element in the prevention of contagion is reasonable screening for infected people.

A common, generalized screening method involves measuring body temperature to detect fever. While fever is not specific to Covid-19, it is regarded as a method with good sensitivity for this disease (de Oliveira Neto, de Oliveira Tavares, Schuch, & Lima, 2020). Body temperature testing before admission is currently recommended in a wide variety of health care scenarios (Krengli, Ferrara, Mastroleo, Brambilla, & Ricardi, 2020; Rombolà et al., 2020; Sainati & Biffi, 2020); this type of screening is also commonly observed in TV reports before people are admitted to health care clinics, restaurants, worksites, or public transportation. Finally, screening for fever is part of the World Health Organization recommendations for management of ill travelers at international borders (World Health Organization [WHO], 2020).

Apparently, the method of choice for this fever screening is to measure temporal artery (forehead) temperature with different infrared, contactless devices which provide for a practical, reasonably safe, and quick reading. However, while any effort to detect potentially ill individuals is commendable, the accuracy, reliability and validity of such readings have not been evaluated on the field. In this editorial, I intend to briefly review some basic physiology and epidemiology, together with published field studies with exercising humans, to explore some of the limitations associated with the use of field temporal temperature readings (T_temp) as a screening method for Covid-19.

Thermoregulation basics: core temperature and peripheral temperature

Humans are hot-blooded animals. We regulate body temperature through a series of responses coordinated by the preoptic area of the anterior hypothalamus, by conserving and dissipating generated heat at different rates, always attempting to maintain core (central) temperature within a rather narrow range (Castellani, 2003). Peripheral temperature, in turn, is more variable, susceptible to environmental conditions such as radiation, wind velocity, and air temperature and humidity.

At the end of an intense exercise session in the heat, when sweat production and evaporation are high during recovery in a cool, well-ventilated area, it is possible for a normal, healthy human to have, at the same time, a high core temperature (e.g., 39°C) and a low skin temperature (e.g., 32°C). This is, of course, an extreme situation. But daily living entails multiple adjustments which can result in a discrepancy between core temperature and peripheral temperature. Skin blood flow may be highly limited, for instance, when a person is standing in a cold wind or seated in an air-conditioned room, as the body attempts to conserve heat and maintain core temperature from dropping below 37°C; skin temperature will be considerably lower. Conversely, someone standing in the sun will be exposed to radiant heat, which will elevate skin temperature before core temperature begins to change.

It follows that a peripheral temperature reading may not be representative of core temperature. This bears directly on the theoretical validity of the temporal (forehead) temperature to detect a fever.
Who has a fever? Defining a fever cutoff temperature

Even when a valid core temperature reading is obtained, there is some discussion about what a normal body temperature is. Humans regulate core temperature between about 35°C and 39°C, depending on environmental thermal stress and some physiological changes such as fever. Circadian rhythms can make body temperature oscillate by about 0.5°C to 1.0°C (Castellani, 2003).

When humans are at rest in a thermoneutral environment, an elevated core temperature is considered fever. Fever is a normal response to viral or bacterial infection, consisting of an elevation of the set-point temperature mediated by the release of pyrogens. However, due to the mentioned variability of normal body temperature, the question is: what is the correct fever cutoff temperature?

Ivayla Geneva and her colleagues published a systematic review in 2019, where they calculated ranges for body temperature at different sites from 9227 measurements performed on 7636 subjects, published in the 36 studies reviewed. They found no clinically significant difference in normal body temperature when comparing males to females, but they found their adults aged 60 years or older to have a slightly lower average temperature (Geneva, Cuzzo, Fazili & Javaid, 2019). Considering what has already been discussed in this editorial about peripheral temperature readings, together with the fact that they showed up to a 1°C difference depending on measurement site, I suggest that rectal temperature should be used as the criterion. The authors obtained 37.04 ± 0.36°C (mean ± S.D.), or a range (mean ± 2 S.D.) of 36.32 to 37.76°C for normal rectal temperature (Geneva et al., 2019).

In other words, a resting individual in a thermoneutral environment could be said to have a fever if his/her rectal temperature is greater than 37.8°C. In the case of an individual who is 60 years old or older, a slightly lower cutoff point of 37.7°C could be used. This is a very important physiological criterion, but not the only one to be considered, since the definition of a cutoff point for body temperature as a screening test for fever will have a large impact on its sensitivity and specificity, to be discussed later. Furthermore, because rectal temperature is not practical as a screening method, other alternatives must be considered.

Prevalence of fever in COVID-19 confirmed cases

One of the big challenges with this pandemic is the fact that many individuals can be infected and contagious while being asymptomatic. According to the World Health Organization, the main symptoms are fever, dry cough and fatigue (World Health Organization, n.d.). How many of those infected present fever?

Guan et al. (2020) reported data on 1099 patients with laboratory-confirmed Covid-19 from 522 different hospitals. Only 43.8% presented fever on admission, but 88.7% had it during hospitalization. In a retrospective study, Bi et al. (2020) reported 84% of the 391 cases in Shenzhen had fever; 30% of their contact-based group did not have a fever at the time of first clinical assessment. Liang et al. (2020) reported a retrospective study in a teaching hospital in Beijing, where 21 individuals were confirmed with Covid-19; 85.7% of them presented to the clinic with a fever. Fu et al. (2020) published a systematic review and meta-analysis, and calculated that fever was present in 83.3% of confirmed cases. In the United States, in New
York, Richardson et al. (2020) reported on 5700 hospitalized patients for Covid-19 and found that only 30.7% of them had a fever upon hospital arrival and triage.

From this extremely limited information it can be gathered that while more than 80% of Covid-19 confirmed cases will have a fever, as little as 30% of them may present with fever initially. This has serious implications for the sensitivity of a screening test based on body temperature.

**What is a valid body temperature measurement?**

As explained earlier in this editorial, peripheral temperatures in humans may vary considerably from core temperature in different situations. Published validation studies have been carried out under strictly controlled conditions. As an example, Henker & Coyne (1995) performed a study in the critical care unit of a hospital. They compared the pulmonary artery temperature (the standard of reference) with a variety of peripheral temperatures in critically ill patients and concluded that most of them were not accurate enough to make diagnostic or treatment decisions; they deemed the electronic thermometer used as acceptable for those purposes if used to measure oral or axillary temperatures, but recommended against using them when patient infection is a concern. They also recommended against measuring rectal temperature because of patient discomfort and potential for cross-contamination.

The method under scrutiny in this editorial is anterior temporal artery (forehead) temperature (Ttemp). Figure 1 shows a thermograph taken with a thermographic camera (FLIR® model T-650sc, Wilsonville, OR); the thermograph shows actual skin temperature of the forehead, from which core temperature is commonly predicted. In normal life situations, different infrared devices are used to read the skin temperature of the forehead and predict core temperature using proprietary (black box) algorithms. In other words, the values displayed by every-day use devices are not the actual forehead temperature, but an estimation of core temperature made by each manufacturer. The correct site of measurement is an additional concern.

![Figure 1. Thermograph of a healthy adult male sitting in a comfortable 24°C room temperature. Maximum, minimum, and average values shown in the upper left corner correspond to the rectangular area on the forehead, where measurements are typically taken. Source: the author. Special thanks to Rodrigo Cordero-Tencio for his technical assistance.](image-url)
Also in a hospital setting, 164 intensive care patients were measured by both rectal temperature and infrared temporal thermometry in Norway (Dybwik & Nielsen, 2003). 70 of those patients had a fever, defined as rectal temperature equal or greater than 38°C; only 33 of them were detected by the temporal scanner. The authors concluded that “sensitivity of the infrared temporal thermometer for detecting rectally measured fever is too low to recommend its use in adult intensive care patients.” (Dybwik & Nielsen, 2003, p. 3025).

A group from the Netherlands compared the esophageal and rectal temperatures (as the standard of reference) with two infrared forehead skin thermometers from the same manufacturer (SensorTouch by Exergen) in two different experiments, reported together in the same article (Kistemaker, Den Hartog & Daanen, 2006). They manipulated core temperature with exercise and different environments, but again, their tests were performed under strictly controlled conditions. The results obtained from measuring forehead (superficial temporal artery) temperature were considered not reliable during periods of increasing body temperature.

Therefore, if under these highly-controlled conditions, the validity of various peripheral temperature measurements is questioned and, specifically, temporal (forehead) temperatures ($T_{temp}$) are not considered valid and reliable, what can be expected when using temporal temperature readings as a measure of core temperature in normal, day-to-day activities and settings? Exercise physiology studies shed some light on this issue.

Low et al. (2007) compared temporal thermometry with intestinal (core) temperature during passive heating of 16 healthy adults, using a water perfusion suit in a temperature-controlled laboratory (26±1 °C). They used ingestible pills (HQ Inc., Palmetto, FL) to measure core temperature, and performed readings of temporal temperature in strict compliance with the manufacturer’s instructions (TemporalScanner TAT-5000, Exergen, Watertown, MA), which involved moving the device from the forehead midline to the lateral hairline; in the case of diaphoretic (sweating) individuals, they also moved the device behind the ear, as instructed. As expected, intestinal temperature increased steadily after 20 minutes. Temporal temperature, however, decreased, and was significantly different from intestinal temperature at all time points after 20 minutes: while intestinal temperature detected an increase of about 0.7°C at the end of testing, temporal temperature measured a 0.2°C decrease. The authors concluded that both measurement methods provide similar values only while subjects are thermoneutral. However, during the passive-heating induced increase in core temperature, the temporal temperature readings did not track internal temperature. Because of the latter, they expressed serious concern about the validity of temporal scanning devices to measure internal temperature (Low et al., 2007).

In a study of 25 subjects during outdoor exercise in the heat, Doug Casa and his colleagues compared rectal temperature readings (the core temperature reference standard) with measurements at a variety of sites (inner ear, GI tract, forehead, and temporal artery) using different devices (Casa et al., 2007). They were concerned that, since many of the latter may be affected by wind, fluid ingestion, skin temperature and sweat evaporation, they may not be valid measures of core temperature during outdoor exercise. Their subjects were tested before exercise, every hour during a three-hour exercise session, and every 20 minutes during their one-hour recovery. Temporal measurement showed a -1.46°C bias; its correlation coefficient was -0.56 and the Bland-Altman limits of agreement ± 2.16°C. They concluded that only the
gastrointestinal temperature was accurate, compared with the rectal criterion. In addition, regarding the temporal temperature testing device, the authors stated that “using a correction factor would not validate this device because changes over time in [their values] were opposite those in RCT” (Casa et al., 2007, p. 340).

Matthew Ganio and colleagues published a similar study in 2009, but their 25 subjects exercised in a controlled laboratory environment and recovered indoors. The design was very similar to the one used by Casa et al. (2007), but the exercise was performed at 36.4 ± 1.2°C and 52% relative humidity in the laboratory, and the recovery took place in a normal air-conditioned room at approximately 23.3°C and 40% RH. Temporal measurement showed a -0.87°C bias, with an intraclass correlation coefficient of 0.44. The Bland-Altman limits of agreement were ±1.77°C. While these results are better than those obtained outdoors by Casa et al. (2007), the authors concluded that temporal temperature (Ttemp) is an invalid estimate of core (rectal) temperature in exercising athletes even indoors. They suggested that this may be due to the inconsistent blood flow in the superficial temporal artery or to the presence and evaporation of sweat (Ganio et al., 2009).

The scientific literature has a large number of studies comparing rectal and aural core body temperatures. Aural (tympanic) temperature should also be of interest, as it shares some of the advantages of temporal temperature testing. But it also shares its limitations. Huggins, Glaviano, Negishi, Casa & Hertel (2012) published a meta-analysis on this topic in exercising individuals, concluding that as core temperature increases during exercise, aural temperature appears to underestimate it, registering lower values (about 1°C) than those recorded by rectal thermometry. Because the discrepancy was shown to increase as subjects became more hyperthermic, the authors highlighted that solving it is not simply a matter of adding a constant value to the aural temperature reading (Huggins et al., 2012).

The ability of measuring body temperature in a practical, quick, and relatively inexpensive way has a very wide appeal. There are enough concerns, however, from the scientific literature to suggest that forehead (temporal) temperature readings (Ttemp) with an infrared device may not be valid.

Screening test criteria: specificity and sensitivity

Screening tests have an important role in public health. A screening test, as opposed to a more expensive, invasive, and often complicated diagnostic test, is meant to be applied to a large number of people in order to classify them regarding the probability that they present a particular condition of interest (an illness). These people are typically asymptomatic, and the screening test must be practical and inexpensive (Aragón-Vargas, 1995; Trevethan, 2017).

Screening tests are typically evaluated in terms of sensitivity and specificity. Sensitivity is often described as a test’s ability to correctly identify all people who have a condition or, in other words, the extent to which the screen will “catch” all people who present the condition of interest. Specificity, on the other hand, can be described as the test’s ability to correctly identify people who do not have a condition or, in other words, the extent to which the screen will “pass” all people who do not present the condition of interest (Trevethan, 2017). According to Trevethan, these are the properties or the “credentials” of the test, given that a person does or does not actually present a condition of interest. However, he warns, we are often more interested in what
Editorial

to do with a particular individual who tests positive or negative: how confident are we when a decision is made regarding whether they present the condition of interest? For that purpose, his recommendation is that we must consider the Positive Predictive Value (PPV) and the Negative Predictive Value (NPV) of the test (Trevethan, 2017).

Back to the limitations of temporal temperature ($T_{\text{temp}}$) as a screening test, because of the virulence of SARS-CoV-2, and how easily it has been transmitted by non-symptomatic individuals, it is highly desirable to have a screening test with good sensitivity, that is, a test which will detect most individuals who indeed have a fever and therefore will provide adequate protection to those potentially interacting with them. Accordingly, experts will define a cutoff point for fever that is low enough, while at the same time considering the literature about normal body temperature already discussed. Unfortunately, the fact that not all Covid-19 infected individuals present with fever undermines the sensitivity of the test. While it would also be desirable for the test to have a reasonable specificity, that is, it will not err too often on the side of flagging healthy people as being sick, fever is actually not specific to Covid-19, and therefore the test will potentially select many individuals who have other types of infections. This is particularly true if the screening is applied to groups where the prevalence of Covid-19 is low. After all that has been said, if a test is not valid to begin with, its sensitivity, specificity, PPV and NPV will all be impaired.

A sound evaluation of temporal temperature ($T_{\text{temp}}$) as a screening test should acknowledge the need to address three perspectives: the quality of the devices used for testing, test validity, and test credentials. The first one, device quality, is beyond the scope of this editorial; suffice to say that some of the devices being used will inevitably be more accurate and reliable when providing a reading of the actual forehead skin temperature. But even if the best temporal thermometer is used, if $T_{\text{temp}}$ is not a valid measure of body temperature, its utility as a screening test will be extremely limited. Furthermore, the fact that each manufacturer uses its own proprietary algorithm to predict core temperature from the infrared-detected forehead skin temperature undermines any attempt at standardization.

Final recommendations

In summary, temporal (forehead) temperature readings ($T_{\text{temp}}$) have serious limitations as a screening method for COVID-19: there is enough evidence to question the validity of $T_{\text{temp}}$ as a measure of core body temperature; the cutoff point used to define fever has been established for core temperature, but manufacturers of $T_{\text{temp}}$ devices use proprietary algorithms to estimate it; infrared devices are being used with minimum attention paid to established protocols for a reasonable reading; the prevalence of fever may not be high enough in Covid-19 infected individuals, undermining the sensitivity of the test; fever is not exclusive of other infections, affecting its specificity. Finally, decisions being made on the basis of $T_{\text{temp}}$ results can be very inconvenient to individuals and potentially serious for the general population.

It is obvious that we currently do not have a perfect screening test for Covid-19, and never will. But as we develop better ways to deal with this pandemic, we can improve the validity of a practical test currently in use, namely, temporal temperature reading with infrared devices ($T_{\text{temp}}$). We can follow several steps in an attempt to make it a more sensitive and specific test. To begin with:
• Infrared temporal thermometers should be selected carefully, and their manufacturers must be held to high standards.
• Personnel in charge of field testing at public venues, worksites, and public transportation should be properly trained, according to standards established by experts or respectable organizations.
• Tests should only be applied to individuals who have been at rest and in a thermoneutral environment, with limited air movement, for 10 minutes or longer.
• Experts should establish a well-documented temporal temperature cutoff point for fever that is independent of manufacturer-defined proprietary algorithms.

Finally, it is crucial to stay alert: experts have warned us regarding the danger of a false sense of security experienced by many people when using face masks or gloves; these experts insist on the right way to use them and emphasize basic hygiene rules. Likewise, $T_{\text{temp}}$ is not only an imperfect screening test but it can also give us a false sense of security during a very delicate phase of this pandemic. The test must be improved and must be always accompanied with all the preventive measures we should already know by heart, such as physical distancing, frequent hand washing, surface sanitizing, and avoiding touching our face.

References


