



OPEN

Proper use of noncontact infrared thermometry for temperature screening during COVID-19

Amber S. Hussain¹✉, Heather S. Hussain², Nathan Betcher², Robert Behm¹ & Burt Cagir¹

Among the myriad of challenges healthcare institutions face in dealing with coronavirus disease 2019 (COVID-19), screening for the detection of febrile persons entering facilities remains problematic, particularly when paired with CDC and WHO spatial distancing guidance. Aggressive source control measures during the outbreak of COVID-19 has led to re-purposed use of noncontact infrared thermometry (NCIT) for temperature screening. This study was commissioned to establish the efficacy of this technology for temperature screening by healthcare facilities. We conducted a prospective, observational, single-center study in a level II trauma center at the onset of the COVID-19 outbreak to assess (i) method agreement between NCIT and temporal artery reference temperature, (ii) diagnostic accuracy of NCIT in detecting referent temperature $\geq 100.0^\circ\text{F}$ and ensuing test sensitivity and specificity and (iii) technical limitations of this technology. Of 51 healthy, non-febrile, healthcare workers surveyed, the mean temporal artery temperature was 98.4°F (95% confidence interval (CI) = $[98.2, 98.6]^\circ\text{F}$). Mean NCIT temperatures measured from 1 ft, 3 ft, and 6 ft distances were 92.2°F (95% CI = $[91.8, 92.67]^\circ\text{F}$), 91.3°F (95% CI = $[90.8, 91.8]^\circ\text{F}$), and 89.6°F (95% CI = $[89.2, 90.1]^\circ\text{F}$), respectively. From statistical analysis, the only method in sufficient agreement with the reference standard was NCIT at 1 ft. This demonstrated that the device offset (mean temperature difference) between these methods was -6.15°F (95% CI = $[-6.56, -5.74]^\circ\text{F}$) with 95% of measurement differences within -8.99°F (95% CI = $[-9.69, -8.29]^\circ\text{F}$) and -3.31°F (95% CI = $[-4.00, -2.61]^\circ\text{F}$). By setting the NCIT screening threshold to 93.5°F at 1 ft, we achieve diagnostic accuracy with 70.9% test sensitivity and specificity for temperature detection $\geq 100.0^\circ\text{F}$ by reference standard. In comparison, reducing this screening criterion to the lower limit of the device-specific offset, such as 91.1°F , produces a highly sensitive screening test at 98.2%, which may be favorable in high-risk pandemic disease. For future consideration, an infrared device with a higher distance-to-spot size ratio approaching 50:1 would theoretically produce similar results at 6 ft, in accordance with CDC and WHO spatial distancing guidelines.

Since the emergence of COVID-19, over 100 guidance documents have been produced by the World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC)^{1,2}. Interim guidance for United States healthcare facilities recommend aggressive universal source control measures and well-equipped triage procedures at the entrances of facilities to actively screen individuals for fever. Fever is defined as either measured temperature greater than or equal to 100.0°F or subjective fever¹ and symptoms. Respiratory symptoms consistent with COVID-19 are cough, shortness of breath, and sore throat¹. Maintaining spatial separation at 6 ft is also advised by the CDC and keeping at least 1 m apart is recommended by the WHO¹⁻³.

Under this guidance, our center aimed to investigate the diagnostic accuracy of temperature measurement using the Fluke 561 Noncontact Infrared Thermometer (NCIT) from a distance of 6 ft. A cursory investigation into this device's use revealed variable results in temperature measurement, which degraded as a function of distance, and a notable offset compared to the expected body temperature.

Furthermore, there is a paucity of high-quality research comparing thermal measurements obtained by noncontact infrared thermometry (IRT) versus conventional methods in public health applications, such as pandemic disease, like COVID-19.

Several studies support this dissimilitude in diagnostic accuracy across a myriad of available devices⁴⁻⁶. In a review of six studies by Bitar et al.⁷ assessing measurement of forehead temperature using an NCIT, test sensitivity ranged widely from 4 to 89.6%, the specificity from 75.4 to 99.6%, and the Positive Predictive Value (PPV)

¹Department of General Surgery, Guthrie Robert Packer Hospital, 1 Guthrie Square, Sayre, PA 18840, USA. ²Seattle, WA, USA. ✉email: amber.hussain@guthrie.org

between 3.5 and 65.4%. Four studies failed to provide technical information about the NCIT device used. Three studies failed to report environmental conditions or stabilization factors, and a majority of studies did not detail the procedural methods employed for measurement testing, such as distance away from the target. Studies using handheld devices at less than 7.9 in (20 cm) from the forehead showed improved accuracy^{4,8}.

This variability called into question the efficacy of screening practices that similarly equipped healthcare facilities are implementing to reduce the spread of COVID-19 and comply with spatial distancing guidance. This study will analyze temperature data collected using the Fluke 561 infrared thermometer at 1 ft, 3 ft and 6 ft distances to assess the accuracy of this device compared to conventional temporal artery thermometry and survey the technological constraints inherent to NCITs. Our intent is to inform a concerned audience of explicit and implicit limitations and recommend an optimal screening process for infection prevention and control measures during pandemic disease.

Methods

Study design. This nonblinded prospective single-center study was designed to examine the validity of the Fluke 561 NCIT (Fluke Corp, Everett, WA) at variable distances compared to conventional temporal artery reference temperature of employees in a U.S. healthcare facility during the outbreak of COVID-19.

The study was conducted at Guthrie Robert Packer Hospital, a rural 267-bed tertiary care level II trauma center serving the southern tier of New York and the northern tier of Pennsylvania⁹.

Participants were recruited by study investigators at random from different hospital wards. An institutional review board-approved letter was used to inform participants about the study and informed consent was obtained prior to enrollment.

Ethical approval. The study protocol and informed consent documentation were reviewed and approved by the Institutional Review Board of The Guthrie Clinic. All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation and with the Helsinki Declaration of 1964 and its later amendments.

Participants. Eligible participants were healthy adults aged 18 or older who were employed by the Guthrie Clinic, working on-site during the COVID-19 pandemic. Participants were excluded if they were pregnant.

Procedures and data collection. Data was collected, by a single researcher, between April 6 and April 10, 2020 on different hospital wards. Age and gender were included in data collection.

A fever threshold ≥ 100.0 °F was adopted in accordance with CDC guidelines for U.S. healthcare facility screening¹. If participants were found to have a measured temperature ≥ 100.0 °F, they would be instructed to report to the Employee Health Office.

Mode of use, device calibration, temperature scanning and disinfection were followed according to manufacturer instructions for each apparatus and in compliance with CDC infection prevention standards. Similarly, Personal Protective Equipment (PPE) compliant with institution and CDC precautions for COVID-19, including the use of an N95 respirator, was worn.

Temporal artery thermometer (TAT). Each participants' baseline body temperature was measured using the Exergen TemporalScanner (Model TAT5000, Exergen Corp, Watertown, MA) Temporal Artery Thermometer (TAT). Technical specifications for the Exergen TemporalScanner are given in Table 1.

This accurate and noninvasive thermometer uses infrared technology to measure heat emitted from the skin's surface overlying the temporal artery¹⁰. Temporal artery temperature is a close approximate of rectal temperature and therefore accurately reflects a measure of core body temperature¹⁰⁻¹².

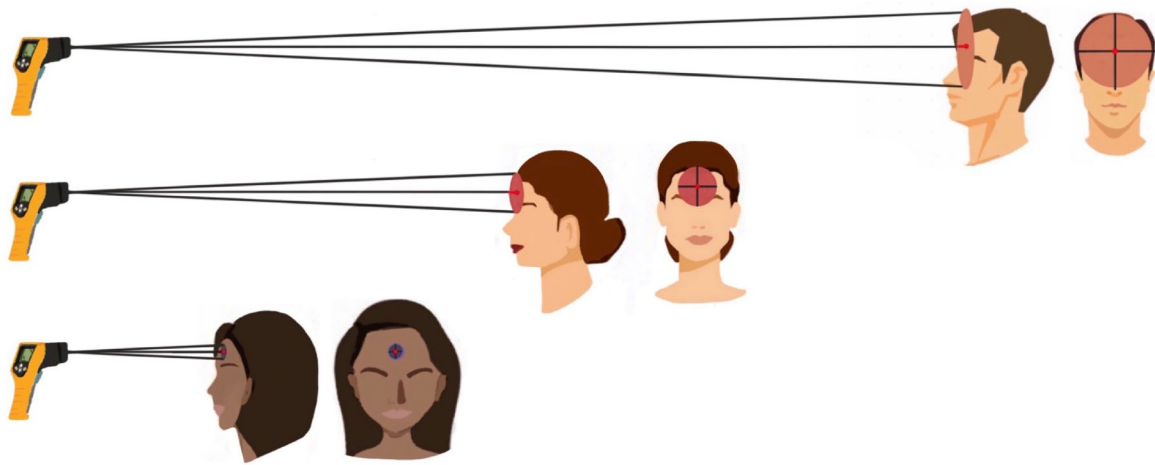
Non-contact infrared thermometer (NCIT). Next, a digital ruler was used to measure and mark 1 ft, 3 ft, and 6 ft distances on the floor of each testing location, relative to a fixed reference point.

Three sequential measurements were obtained using the Fluke 561 NCIT at device-to-target [Target, in this context, denotes the forehead of each study participant at the specified IR measurement distance] distances of 6 ft, 3 ft, and 1 ft (see Fig. 1a) with emissivity set to Hi (see Table 1). The focal point of the device, marked by a laser, was centered on each participants' forehead whom were asked to close their eyes during collection.

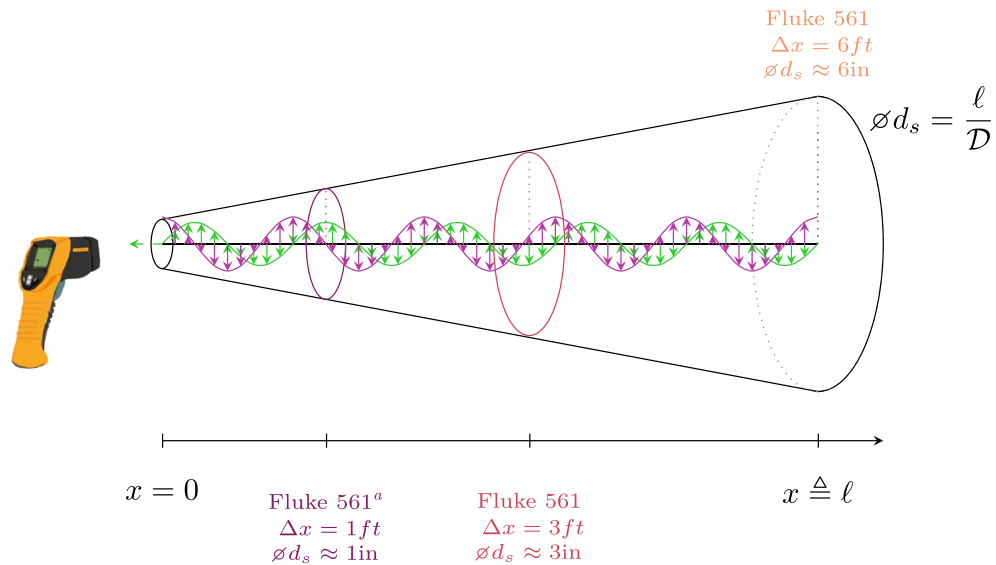
NCITs measure the amount of infrared energy emitted by an object's surface; it is not a direct measure of core body temperature and can be influenced by environmental conditions and physiological factors.

To ensure accurate temperature measurement, device-to-target distance is crucial since it governs the measurement spot size according to the NCIT's distance-to-spot size (D:S) specification. The spot size, i.e. diameter of the measurement area, indicates 90% encircled energy¹³ and increases proportionally with distance. For that reason, the spot size should not exceed the intended measurement area [Fluke recommends that the measurement area be twice as large as the measured spot size¹⁴]. The Fluke 561 D:S ratio is 12:1 which implies a 6 in spot diameter at 6 ft (see Fig. 1)¹³.

Outcome measures. Primary outcome measures were temperature values obtained with the Exergen TAT [For brevity, TAT and NCIT may be used interchangeably in reference to the Exergen TemporalScanner and Fluke 561 devices, respectively] and three temperature measurements with the Fluke 561 at 1 ft, 3 ft, and 6 ft distances. NCIT temperatures were compared to temporal artery reference temperatures and analyses of agreement were conducted. Secondary outcome measures of test sensitivity and specificity to assess diagnostic accuracy for



(a) Illustration of Distance-to-Spot Size (D:S) ratio for 1ft (bottom), 3ft (middle), and 6ft (top) using an NCIT with D:S equivalent to 12 : 1, such as the Fluke 561. Approximate spot sizes are depicted for each distance.



(b) Illustration of Distance-to-Spot Size (D:S) ratio and computation. The diameter of the measurement spot size is proportional to the distance between the measurement target and NCIT, i.e. $\varnothing d_s \propto \Delta x$. Specifically, $\varnothing d_s \triangleq \frac{\Delta x}{D}$.

^aThe D:S ratio gives an *approximate* spot size diameter. The Fluke 561 manual¹³ cites $\varnothing d_s = 1.5\text{in}$ at 1ft. We recommend that the reader verify this crucial measurement parameter in the corresponding instrument specific manual.

Figure 1. Illustration of distance-to-spot size (D:S) ratio.

Device	Exergen TemporalScanner	Fluke 561 (NCIT)	Fluke 568 (NCIT) ^a
Temperature range	[61, 110] °F ([16, 43] °C)	[−40, 1022] °F ([−40, 550] °C)	[−40, 1472] °F ([−40, 800] °C)
Accuracy	±0.2 °F or 0.1 °C ^b	> 32 °F : max(±1%, ±2 °F)	> 32 °F : max(±1%, ±2 °F)
Repeatability	Not available	max(±0.5%, ±2 °F (1 °C))	max(±0.5%, ±1 °F (0.5 °C))
D:S (90%)	Not applicable	12:1	50:1
Response time	40 ms	< 500 ms	< 500 ms
Emissivity	Not applicable	Lo (0.3)/Med (0.7)/Hi (0.95)	0.10 to 1.00 by 0.01
Resolution	0.1 °F or °C	0.1 °F or °C	0.1 °F or °C

Table 1. Technical specifications of temporal artery and non-contact infrared thermometers. ^aFor reference and discussion only, device was not tested in this study. ^bPer ASTM E1112.

Measurement method	Mean (°F)	SD ^a (°F)	Median (°F)	Range (°F)	IQR (°F) ^b	Min(°F)	Max (°F)
Temporal (TAT)	98.4	0.59	98.4	2.5	0.6	97.1	99.6
$\Delta x = 1$ ft (NCIT)	92.2	1.52	92.4	6.4	2.5	88.9	95.3
$\Delta x = 3$ ft (NCIT)	91.3	1.72	91.8	6.8	2.8	87.4	94.2
$\Delta x = 6$ ft (NCIT)	89.6	1.68	89.9	6.3	2.9	86.6	92.9

Table 2. Descriptive statistics for each measurement group. ^a SD: Standard Deviation. ^b IQR: Interquartile Range.

detecting temperature ≥ 100.0 °F by TAT were then analyzed using NCIT screening criteria on an open source data set. Third, this paper outlines the salient technical features of noncontact infrared thermometry.

Sample size. The sample size for this study ($N = 51$) exceeded minimum requirements for all statistical tests and the Bland–Altman agreement analyses. The degree of freedom for this study was 50 which corresponds to a t -value of 2.009 used in Confidence Interval (95% CI) computations.

Blinding. All participants and data collectors were unblinded to data collection. Fully deidentified objective outcome data was provided to an independent data analyst.

Data analysis. All data analyses were performed using MATLAB, version R2019a (The MathWorks Inc., Natick, MA). First, the validity of NCIT data at 1 ft, 3 ft, and 6 ft had to be established. Central tendency and dispersion characteristics for each dataset was computed and graphically visualized through histogram plots. Formal judgment of normality was performed using robust statistical testing. Those results, in combination with the limitations of the NCIT device, led to rejection of NCIT measurements at 3 ft and 6 ft. Then, methodological comparison between TAT and NCIT at 1 ft was conducted using Bland–Altman plots, the benchmark approach for assessing between-method differences^{15–17}. This provided an indication of bias and 95% Limits of Agreement (LOA) between the measurement methods. Diagnostic accuracy for detecting fever was then explored.

Prior to presenting the main results, we introduce the following nomenclature: $T_{\mathcal{X}}$ (°F) denotes the temperature measured by method \mathcal{X} . That is, T_{TAT} represents temperature measured by the TAT, while $T_{1\text{ft}}$, $T_{3\text{ft}}$, and $T_{6\text{ft}}$ correspond to NCIT measurements.

Results

A total of 51 healthy adults were sampled. Ages ranged 21 to 67 with a median age of 31 [Median excludes one participant who did not report their age.]. Of the participants, 37 (72.5%) were female and none exhibited fever or symptoms of illness ($N = 51$, $P = \emptyset$). The mean temporal artery temperature was 98.4 °F with all measurements falling between 97.1 and 99.6 °F. No data was excluded from the study.

Descriptive statistics. Measures of central tendency and dispersion characteristics for each measurement method were computed and summarized in Table 2. Histograms for each measurement method are provided in Fig. 2 where $T_{\mathcal{X}}$ (°F) was binned with width defined by $w_{\text{bin}} = \frac{3.49\sigma}{\sqrt{n}}$.

This figure illustrates the existence of a measurement bias, i.e. shift in the mean, between TAT and NCIT measurements. Mean temperature for the TAT was 98.4 °F (95% CI = [98.2 98.6] °F). The mean temperatures for NCIT at 1 ft, 3 ft, and 6 ft were 92.2 °F (95% CI = [91.8 92.7] °F), 91.3 °F (95% CI = [90.8 91.8] °F), and 89.6 °F (95% CI = [89.2 90.1] °F), respectively.

The dispersion of NCIT measurements is greater compared to the central TAT method. Specifically, the standard deviation (σ) for TAT is 0.6 (95% CI = [0.5, 0.7]) compared to the NCIT where $\sigma = 1.5$ (95% CI = [1.3, 1.9]) at 1 ft, and $\sigma = 1.7$ (95% CI = [1.4, 2.1]) for both the 3 ft and 6 ft measurements.

No distributions displayed readily apparent asymmetry or skew; however, formal statistical tests are performed in the next section to determine whether TAT and NCIT measurements are drawn from a normally distributed population.

Distribution tests for normality. Statistical test decisions, along with corresponding p values, are summarized in Table 3 for the Lilliefors' Composite and Anderson–Darling Goodness of Fit (AD GOF) hypothesis tests.

Lilliefors and AD GOF test outcomes for both TAT ($p = 0.19$ and $p = 0.28$) and NCIT at 1 ft ($p = 0.25$ and $p = 0.11$) were false, indicating that the null hypothesis, data comes from a normal distribution, cannot be rejected at the 5% significance level. Conversely, the Lilliefors and AD GOF results for NCIT measurements at 3 ft and 6 ft ($p < 0.1$, $p < 0.05$) reject the same null hypothesis at the 5% significance level.

Thus, 1 ft NCIT data is the only measurement set whose distribution coincides with the referent method's distribution.

Non-normal distributions of 3 ft and 6 ft NCIT data combined with insufficient distance-to-spot size ratings [The corresponding spot sizes were approximately 3 in and 6 in, respectively, which exceed reliable forehead area thresholds] for the intended measurement target using the Fluke 561 (see Fig. 1 and “Basics of noncontact

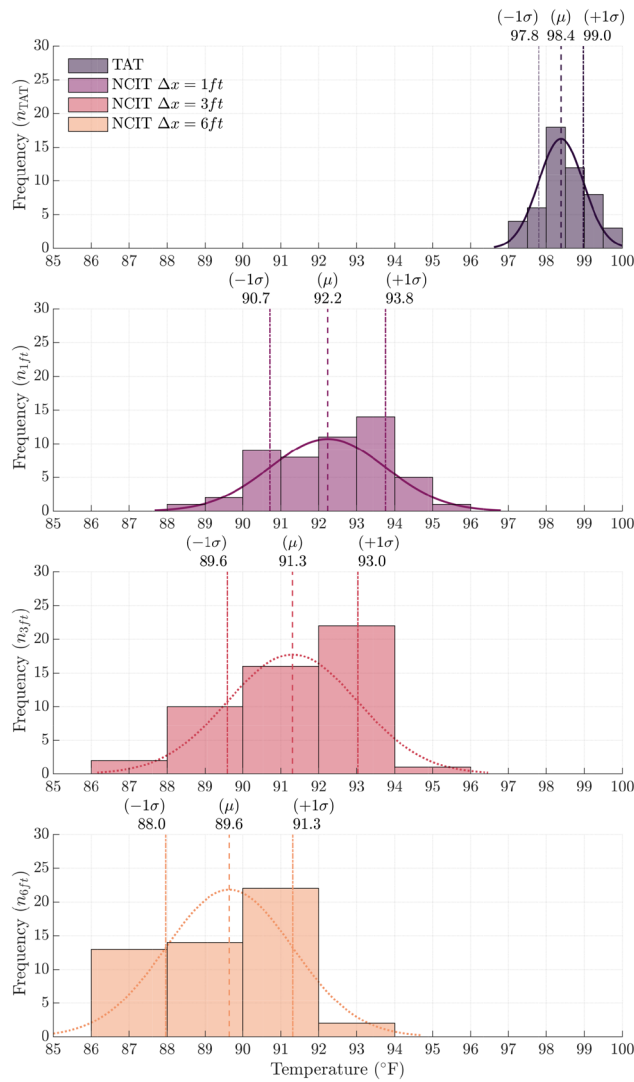


Figure 2. Overlay of histograms and normal density function fits for each measurement method. Vertical dashed lines annotate the mean of each measurement method and vertical dotted lines represent one standard deviation.

Measurement method	Lilliefors test	Anderson–Darling GOF
Temporal (TAT)	False ($p = 0.19$)	False ($p = 0.28$)
$\Delta x = 1$ ft (NCIT)	False ($p = 0.25$)	False ($p = 0.11$)
$\Delta x = 3$ ft (NCIT)	True ($p = 0.03$)	True ($p = 0.008$)
$\Delta x = 6$ ft (NCIT)	True ($p = 0.04$)	True ($p = 0.005$)

Table 3. Distribution tests of each measurement group.

infrared thermal measurements” section), invalidate the 3 ft and 6 ft NCIT procedures for temperature screening or further method comparison.

Bland–Altman method comparison. A Bland–Altman plot is provided in Fig. 3 for the 1 ft NCIT measurement in comparison with the reference (TAT) method. The difference between the IR measurement and temporal measurement is plotted as the dependent variable, with the average of the two measurements serving as the independent variable. The mean bias (dashed line) and 95% LOA (dotted lines) are illustrated with corresponding 95% CIs.

The mean difference between NCIT at 1 ft and TAT is -6.15 °F (95% CI = $[-6.56, -5.74]$ °F) with LOA equivalent to 2.84 °F (95% CI = $[2.14, 3.54]$ °F). In other words, 95% of the measurement differences at 1 ft fall

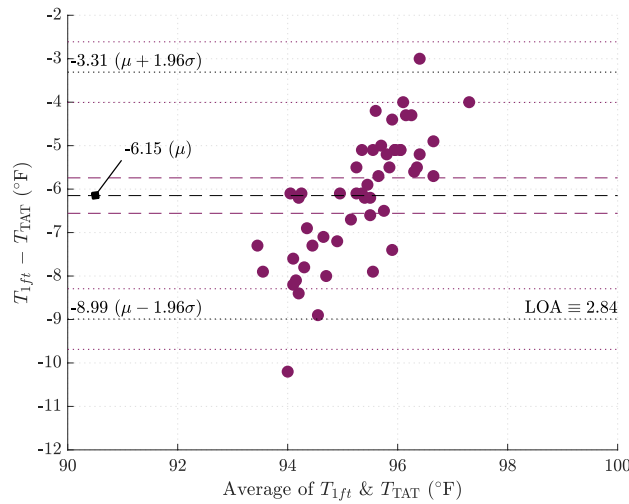


Figure 3. Bland–Altman method comparison plot for IR measurement at 1 ft versus temporal measurement.

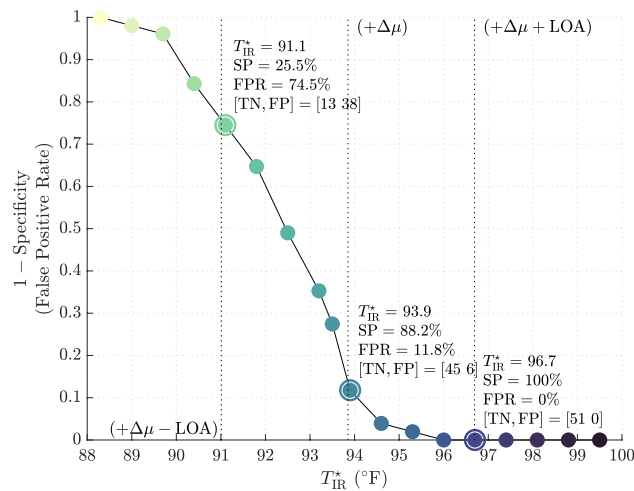


Figure 4. Specificity of fever screening protocol for varying NCIT temperature thresholds in determining afebrile observations.

	$T_{IR}^* = 91.1\text{ }^\circ\text{F}$	$T_{IR}^* = 93.9\text{ }^\circ\text{F}$	$T_{IR}^* = 96.7\text{ }^\circ\text{F}$
Specificity of NCIT at 1 ft^a			
SP ^b (%)	25.5	88.2	100
FPR (%)	74.5	11.8	0
TN (<i>n</i>)	13	45	51
FP (<i>n</i>)	38	6	0

Table 4. Diagnostic accuracy for detecting referent temperature above 100.0 °F using NCIT screening thresholds (T_{IR}^*) of 91.1 °F (left), 93.9 °F (middle), and 96.7 °F (right). These represent the mean, and upper and lower limits of agreement. ^a $T_{IR}^* \triangleq T^* + \Delta\mu + \delta T$, $\Delta\mu = -6.15\text{ }^\circ\text{F}$, $T^* = 100.0\text{ }^\circ\text{F}$. ^b Specificity (SP), False Positive Rate (FPR), True Negative (TN), False Positive (FP).

within $-8.99\text{ }^\circ\text{F}$ (95% CI = $[-9.69, -8.29]\text{ }^\circ\text{F}$) and $-3.31\text{ }^\circ\text{F}$ (95% CI = $[-4.00, -2.61]\text{ }^\circ\text{F}$) of the TAT measurement.

Assessing diagnostic accuracy. In this section, we quantify test sensitivity and specificity of NCIT at 1 ft for detecting temperature $\geq 100.0\text{ }^\circ\text{F}$ by the reference method based on NCIT screening thresholds.

With a mean measurement bias (-6.15°F) and LOA (2.84°F) established using Bland–Altman, diagnostic specificity can be computed for various NCIT thresholds, denoted T_{IR}^* . Figure 4 demonstrates the effect T_{IR}^* has on specificity (SP) with threshold values encircled and presented in Table 4. For example, $T_{\text{IR}}^* = 93.9^{\circ}\text{F}$ corresponds to a test specificity of 88.2% (FPR = 11.8%) with True Negative (TN) and False Positive (FP) group sizes equal to 45 and 6, respectively.

Open source data analysis—STRIDE cohort. Given the sampled population was constrained to a group of healthy adults, complete assessment of diagnostic accuracy, including test sensitivity, necessitates a population with fever occurrence. Thus, an open source cohort from the Stanford Translational Research Integrated Database Environment (STRIDE)¹⁸ was imported and synthesized to validate this study's findings on a larger population with fever prevalence. The STRIDE database is comprised of 578,522 adult outpatient temperature measurements at Stanford Health Care, gathered between 2007 and 2017. The mean of the STRIDE Cohort was 98.0°F with a standard deviation of 0.7°F .

The method of measurement used in the STRIDE cohort was oral thermometry. Therefore, the cohort obser-

	$T_{\text{IR}}^* = 91.1^{\circ}\text{F}$	$T_{\text{IR}}^* = 93.9^{\circ}\text{F}$	$T_{\text{IR}}^* = 96.7^{\circ}\text{F}$
Diagnostic accuracy of NCIT at 1 ft^a (Synthesized STRIDE cohort)			
SP ^b (%)	17.1	80	99.6
SE (%)	98.2	60.1	6.56
FPR (%)	82.9	20	0.39
FNR (%)	1.76	39.9	93.4
PPV (%)	3.99	9.54	37.1
NPV (%)	99.6	98.3	96.8
TN (<i>n</i>)	95, 620	446, 976	556, 458
FP (<i>n</i>)	463,017	111,661	2179
FN (<i>n</i>)	343	7809	18, 230
TP (<i>n</i>)	19, 241	11, 776	1285

Table 5. Diagnostic accuracy for detecting referent temperature above 100.0°F using different NCIT screening thresholds (T_{IR}^*) of 91.1°F (left), 93.9°F (middle), and 96.7°F (right).

^a $T_{\text{IR}}^* \triangleq T^* + \Delta\mu + \delta T$, $\Delta\mu = -6.15^{\circ}\text{F}$, $T^* = 100.0^{\circ}\text{F}$. ^b Specificity (SP), Sensitivity (SE), False Positive Rate (FPR), False Negative Rate (FNR), Positive Predictive Value (PPV), Negative Predictive Value (NPV), True Negative (TN), False Positive (FP), False Negative (FN), True Positive (TP).

variations were adjusted by $+0.8^{\circ}\text{F}$ to account for differences between oral and Exergen TAT thermometry^{10,12}. The adjusted mean is 98.8°F with $P = 19,584$ febrile observations.

A simulated 1 ft NCIT measurement set is then synthesized with the adjusted cohort and repeated random sampling from the distribution

$$\phi(\Delta T) = \frac{\rho}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\Delta T - \Delta\mu}{\rho} \right)^2} \quad (1)$$

where $\Delta\mu \triangleq -6.15^{\circ}\text{F}$ and $\rho \triangleq \frac{\text{LOA}}{1.96} = 1.45^{\circ}\text{F}$. Conceptually, this is a transformation producing NCIT measurements for the open source cohort.

Diagnostic accuracy for detecting referent temperature greater than 100.0°F using three different NCIT temperature thresholds is summarized in Table 5.

A receiver operating characteristic (ROC) curve is shown in Fig. 5 with the points in Table 5 encircled. This graph plots test sensitivity over (1-specificity). This provides a graphical depiction of the implicit trade-off between sensitivity and specificity at select screening criteria.

For the Fluke 561, this curve provides a visual guide for a priori selection of fever screening criteria adjusted to an acceptable sensitivity and specificity for a certain clinical context.

Discussion

Basics of noncontact infrared thermal measurements. *Thermal radiation (first principles)* An infrared thermometer measures the thermal radiation energy of an object and computes the temperature according to the fundamental Stefan–Boltzman law

$$M = \varepsilon\sigma T^4 \quad (2)$$

where M is the radiant exitance, a measure of radiation power emitted by an object into an imperfect vacuum, $\sigma = 5.670373 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$, ε denotes the emissivity of the emitting object, and T (K) represents its absolute Temperature¹⁹. The emissivity ε is a measure of how well an object can emit energy as thermal radiation. Almost perfect emitters, such as skin, have high emissivity ($\varepsilon_{\text{skin}} = 0.98$) while highly reflective surfaces such as polished metals are low.

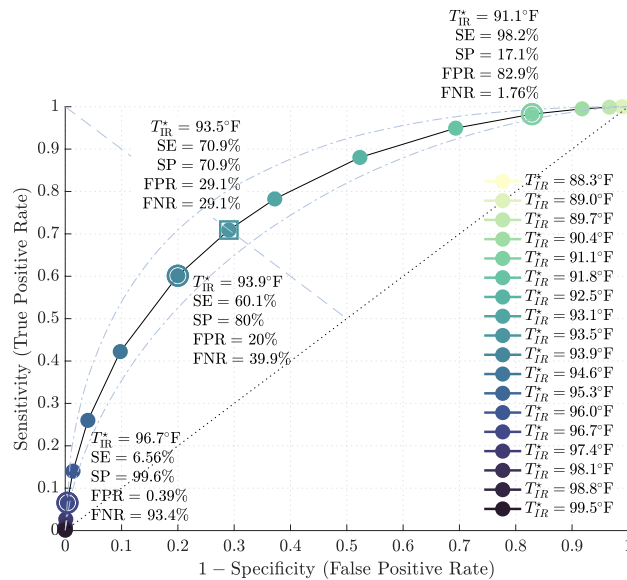


Figure 5. ROC curve for detecting temperature ≥ 100.0 °F for fever screening using open source cohort¹⁸ based on NCIT threshold value. Sensitivity and specificity for points 91.1 °F (right), 93.9 °F (middle), and 96.7 °F (left) are encircled. The dashed line represents a non-discriminatory test.

Sources of error in the computation shown in Eq. (2) stem from the focal resolution of the device, introduced in “Non-contact infrared thermometer (NCIT)” section and discussed further below, along with the prescribed ε value which can lead to large temperature errors due to the fourth power dependence.

Choosing a device with a sufficient D:S ratio such that the spot size is completely inscribed by the measurement area is crucial. If the spot size exceeds the target measurement area, such as the participant’s forehead, then the NCIT samples extraneous thermal radiation leading to incorrect forehead temperature measurements.

Lastly, thermal radiation is subject to atmospheric interference and distortion. End users should mitigate the effects of environmental factors affecting measurement error⁵, such as controlling for ambient temperature gradients, minimizing surface irregularity, and stabilizing device-to-target alignment in an effort to prevent scattering.

Discussion of results. Infected patients are the primary source of pathogen dissemination in healthcare settings³. If initial screening and containment efforts fail, the ramifications can be particularly severe. Healthcare personnel may fall ill and transmit disease to others, including high-risk patients². Infected staff may also require healthcare services themselves, placing additional strain on the medical system as healthcare providers become healthcare receivers.

In a case series by Wang et al.²⁰ of 138 hospitalized patients with COVID-19, 41% were presumed to be due to hospital-related transmission, affecting 40 (29%) health professionals and 17 (12.3%) hospitalized patients.

The lack of a highly sensitive screening test for COVID-19 undermines efforts to contain viral spread²¹, an issue compounded by imperfect COVID-19 diagnostic tests with low-moderate sensitivities estimated around 70%²². In a study analyzing 1099 COVID-19 patients in China this year, fever was present in 43.8% of cases upon admission and 88.7% during hospitalization²³.

A highly sensitive screening strategy allows timely detection of outbreaks and detects almost all cases of pandemic illness, which thereby limits disease spread and optimizes infection prevention and control measures. This approach, however, is much more resource and time consuming as a consequently higher rate of false positive tests require second or multi-stage screening surveys or confirmatory diagnostic testing to correctly classify diseased versus nondiseased individuals.

Alternatively, employing a screening strategy with lower sensitivity (i.e., lower level of detection) affords higher feasibility but at the cost of a higher false negative test rate²⁴. Using this approach, a negative test result cannot safely assume someone is uninfected²². This surveillance method is discouraged and more consequential since misdiagnosed infected persons are not identified or contained by screening strategies and can go on to infect others²².

Results in this study show that utilizing the Fluke 561 within manufacturer specifications yields an average temperature offset of 6.15 °F below temporal artery thermometer values when the device is held 1 ft away from the test subject’s forehead.

If the screening threshold is selected at 93.9 °F in healthcare facilities, the Fluke 561 NCIT will provide a 60.5% sensitive and 80% specific test for the detection of temperature ≥ 100.0 °F by temporal artery thermometry. Lowering the screening criterion to 91.1 °F greatly increases sensitivity to 98.3%, but this is at the expense of an unacceptably high FPR of 82.9%, producing an inefficient test. This phenomena is concisely illustrated in Fig. 6.

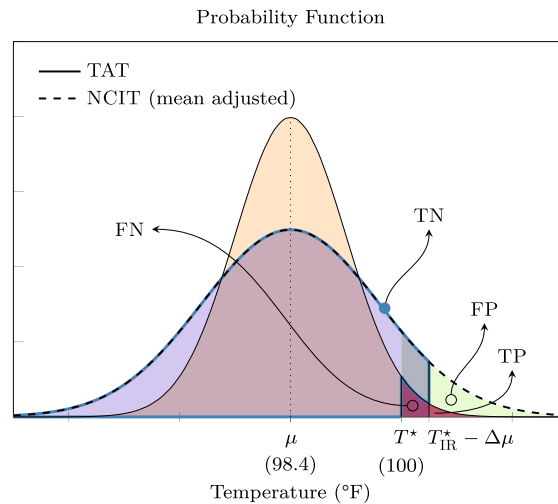


Figure 6. Overlay of probability distribution functions (PDF) for the reference measurement method, i.e. TAT, and the alternate or test measurement method, NCIT. Intersections of the PDFs define the FN, TN, FP, and TP groups given a referent fever screening temperature and a corresponding NCIT screening temperature. This plot illustrates the trade-off in sensitivity and specificity for temperature screening thresholds.

Designing a reference standard for the measurement of temperature elevation in individuals being screened at healthcare facilities during COVID-19 using noncontact infrared thermometry to maintain spatial separation remains a value judgement.

At our institution, it is reasonable to propose a reference standard set to 93.5 °F (Fig. 5). This would yield moderate test sensitivity and specificity of 70.9%, acknowledging a non-negligible FPR and FNR of 29.1%. Thus, the use of multi-step screening tools including issuance of symptom questionnaires and confirmatory TAT testing could further stratify individuals who screen positive by NCIT criteria.

That said, appropriate education and training of relevant staff, control of local conditions and environmental factors to reduce measurement error, and the acquisition of an NCIT with improved accuracy and a higher resolution D:S, such as the Fluke 568 with a distance to spot size ratio of 50:1 (see Table 1), may yield improved specificity while also maintaining a 6 ft distance between screening personnel and target individuals.

Individual facility resources and staff available for screening, population risk characteristics, epidemiological factors, and personnel throughput are factors that should be considered when selecting a screening criteria.

Limitations. The following limitations were identified in our study:

- (i) Because data collection was performed in the background of usual clinical activity, researchers did not adjust heating, ventilation, or air conditioning properties for each testing area per hospital ward.
- (ii) Incident angle of measurement varied due to height difference between the researcher and participants.
- (iii) Lack of blinding of researchers during temperature measurement and data collection.
- (iv) Sample population contained no febrile observations, the majority of participants were female, and the sample was drawn from healthcare providers and therefore not a random distribution of the general population.

Future work. A second-phase investigation is currently underway using the Fluke 568 infrared device (see Table 1), which has a D:S ratio of 50:1. Theoretically speaking, the main results of this paper extend to the Fluke 568 at 6 ft since the spot-size is comparable to the Fluke 561 at 1 ft, and both devices have similar instrument accuracy. Thus, the aim is to validate the main results of this paper for temperature screening at further distances that comply with WHO and CDC separation guidelines (1 m and 6 ft, respectively).

Conclusions

Measuring skin surface temperature in mass public screening applications is an imperfect method for detecting elevated body temperature in individuals potentially infected with coronavirus. Doing so with an NCIT, in an effort to maximize social distancing and minimize risks of exposure to screening staff and healthcare personnel, introduces the potential for additional measurement error. However, a moderately sensitive screening test in the setting of high-risk pandemic virus is possible with NCIT. This requires proper device selection to match the intended application, i.e. specifying a D:S requirement that ensures a sufficient target spot size for the measurement distance, and determination of the device-specific temperature offset compared to an institution's reference standard.

Under this method, the authors strongly advocate for appropriate training of staff, clear instructions-for-use, robust device calibration, and stabilization of environmental and procedural factors to increase success and maximize diagnostic accuracy.

The incorporation of an initial fever screening and subsequent quarantine, PPE, and sanitation regimens for febrile patients may prove beneficial beyond the course of this pandemic. A growing body of clinical research points to a decline in a number of infections normally endemic in hospital settings since the emergence of COVID-19. Studies such as the analysis of *Clostridium difficile* infection in health care settings by Bentivegna et al.²⁵ attribute a statistically significant reduction of this most common pathogen in health-care associated infections with a combination of such practices. A positive externality of NCIT screening for COVID-19 induced fever likely has been the identification and isolation of other 'sentinel' fever producing pathogens and reduction of secondary infections. Such screening and measures may prove invaluable post COVID-19.

With respect to further research, the authors recommend future studies be inclusive of febrile patients and welcome collaboration testing the conclusions of this work. Data for febrile patients was necessarily synthesized for this study, but the authors hold that sound statistical methodology and understanding of the given technology was applied in drawing the conclusions presented. The authors also recommend the inclusion and analysis of NCITs from additional manufacturers, infrared cameras, and nascent technologies.

With respect to NCITs, institutions must realize this is just one of several requisite components of an effective screening program. At their most effective, NCITs such as the Fluke 561, may simultaneously achieve sensitivity and specificity results of approximately 70%. Thus, when employed properly, such tools may correctly identify a majority of febrile individuals, but a non-trivial number of febrile and afebrile persons will be misclassified. It is therefore crucial to understand that temperature measurement through IR thermometry is simply a supplemental screening mechanism. It is not a substitute for accurate and reliable diagnostic tests. For this reason, confirmatory temperature measurement by reference thermometry should be used for those who are identified as positive by NCIT. Lastly, multi-step screening tools such as exposure risk and symptom questionnaires are essential and necessary components of an effective program.

Preliminaries

This study and the analyses herein were conducted and reported in the imperial system of units to ensure consistency of reporting in conjunction with U.S. federal agencies, healthcare agencies, and public and private healthcare systems. As such, all temperature measurements were taken, recorded, and analyzed in degrees Fahrenheit (°F) and all distances in feet (ft) or inches (in). Thus, we refer the reader to the following conversion factors.

Conversion from $T^{\circ}\text{F}$ to $T^{\circ}\text{C}$ is achieved through $T^{\circ}\text{C} = \frac{5}{9}(T^{\circ}\text{F} - 32)$. The conversion factor for ft to meter is approximately $0.3048 \frac{\text{m}}{\text{ft}}$.

Notable values used throughout the paper: 91.1°F (32.8°C), 93.9°F (34.4°C), 96.7°F (35.9°C), 100.0°F (37.8°C), and 1 ft (0.3048 m), 3 ft (0.9144 m), 6 ft (1.8288 m).

Received: 19 October 2020; Accepted: 4 May 2021

Published online: 04 June 2021

References

- Guidance for U.S. Healthcare Facilities about Coronavirus (COVID-19), available from <https://www.cdc.gov/coronavirus/2019-ncov/hcp/us-healthcare-facilities.html>.
- UNDCO and WHO. *Strategic Preparedness and Response Plan on COVID-19 by WHO*. Printed in Geneva, Switzerland. 2020 February; p. 20, available from <https://www.who.int/publications-detail/strategic-preparedness-and-response-plan-for-the-new-coronavirus>.
- World Health Organization (WHO). *Infection prevention and control of epidemic- and pandemic-prone acute respiratory infections in health care*. WHO Guidelines, 1–156, available from http://apps.who.int/iris/bitstream/10665/112656/1/9789241507134_eng.pdf?ua=1 (2014).
- Hayward, G. et al. Non-contact infrared versus axillary and tympanic thermometers in children attending primary care: a mixed-methods study of accuracy and acceptability. *Br. J. Gen. Pract.* **70**(693), E236–E244 (2020).
- Ng, D. K. K. et al. A brief report on the normal range of forehead temperature as determined by noncontact, handheld, infrared thermometer. *Am. J. Infect. Control* **33**(4), 227–229 (2005).
- Teran, C. G. et al. Clinical accuracy of a non-contact infrared skin thermometer in paediatric practice. *Child Care Health Dev.* **38**(4), 471–476 (2012).
- Bitar, D., Goubar, A., & Desenclos, J. C. International travels and fever screening during epidemics: a literature review on the effectiveness and potential use of non-contact infrared thermometers, available from <https://www.eurosurveillance.org/content/10.2807/ese.14.06.191115-en> (2009).
- Hausfater, P., Zhao, Y., Defrenne, S., Bonnet, P. & Riou, B. Cutaneous infrared thermometry for detecting febrile patients. *Emerg. Infect. Dis.* **14**(8), 1255–1258 (2008).
- Guthrie Robert Packer Hospital, available from <https://www.guthrie.org/hospitals-locations/robert-packer-hospital/about-guthrie-robert-packer-hospital>.
- Reference Manual Model TAT-5000, available from www.exergen.com/ww.
- Allegaert, K., Casteels, K., van Gorp, I. & Bogaert, G. Tympanic, infrared skin, and temporal artery scan thermometers compared with rectal measurement in children: a real-life assessment. *Curr. Therap. Res. Clin. Exp.* **76**, 34–38. <https://doi.org/10.1016/j.curtheres.2013.11.005> (2014).
- Changing The Way The World Takes Temperature, available from <https://www.exergen.com/tathermometry/TAT-5000-Training-Trifold.pdf> (2012).
- Fluke. 56x Infrared Thermometers Users Manual. Updated March 31 2020, last accessed 31 March 2020, <https://www.fluke.com/en-us/product/temperature-measurement/ir-thermometers/fluke-561>.
- Fluke. *Accurate Noncontact Infrared Temperature Measurement*, last accessed 31 March 2020, <https://www.flukeprocessinstruments.com/en-us/service-and-support/knowledge-center/infrared-technology/accurate-noncontact-infrared-temperature>.
- Bland, J. M. & Altman, D. G. Measuring agreement in method comparison studies. *Stat. Methods Med. Res.* **8**(2), 135–160 (1999).

16. Bland, J. M. & Altman, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **327**(8476), 307–310 (1986).
17. Altman, D. G. & Bland, J. M. Measurement in medicine: the analysis of method comparison studies. *J. R. Stat. Soc. Ser. D Stat.* **32**(3), 307–317 (1983).
18. Protsiv, M., Ley, C., Lankester, J., Hastie, T. & Parsonnet, J. Decreasing human body temperature in the United States since the industrial revolution. *eLife* **01**(9), e49555 (2020).
19. Kaplan, H. *Practical Applications of Infrared Thermal Sensing and Imaging Equipment*, Third Edition. SPIE Publications, available from <http://gen.lib.rus.ec/book/index.php?md5=bbdd9c98ec3667d1768ec321addb312f>.
20. Wang, D. *et al.* Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA J. Am. Med. Assoc.* **323**, 1061–1069 (2020).
21. Bentivegna, E. *et al.* New IgM seroconversion and positive RT-PCR test after exposure to the virus in recovered COVID-19 patient. *J. Med. Virol.* **93**, 97–98 (2020).
22. Woloshin, S., Patel, N. & Kesselheim, A. S. False negative tests for SARS-CoV-2 infection—challenges and implications. *New Engl. J. Med.* **383**, e38 (2020).
23. Guan, W. J. *et al.* Clinical characteristics of coronavirus disease 2019 in China. *New Engl. J. Med.* **382**, 1708–1720 (2020).
24. World Health Organization (WHO). Guidelines on core components of infection prevention and control programmes at the national and acute health care facility level, available from <https://apps.who.int/iris/bitstream/handle/10665/251730/9789241549929-eng.pdf?sequence=1> (2016).
25. Bentivegna, E. *et al.* Impact of COVID-19 prevention measures on risk of health care-associated *Clostridium difficile* infection. *Am. J. Infect. Control* **49**, 640–642. <https://doi.org/10.1016/j.ajic.2020.09.010> (2021).

Acknowledgements

Dr. Amir Bashiri, Dr. Robert Jameson and Dr. David Levine assisted with participant recruitment and data collection. Lastly, we want to thank all healthcare personnel and ancillary staff in the United States and worldwide, especially those closest to home at Guthrie Robert Packer Hospital. Through your efforts and selfless commission, humanity will rise again, recover and conquer this coronavirus pandemic. We are indebted.

Author contributions

Dr. R.B. proposed and initially designed the study. Dr. A.H., the corresponding author and guarantor, led the study and manuscript preparation, acquired data, and interpreted the results. Dr. H.H. analyzed the data, summarized results, and generated the tables and figures. N.B. assisted with data analysis, interpretation, and contributed to manuscript revision. Dr. B.C. oversaw the study and consulted throughout, providing valuable insight.

Funding

No specific funding received for this work.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.S.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021