Title: Diagnostic accuracy of non-contact infrared thermometers and thermal scanners: A systematic review and meta-analysis

Running title: Diagnostic accuracy of non-contact infrared thermometry

Order	Name of author	Highest qualification	Affiliation
1	Nishant Aggarwal	MBBS	Department of Medicine, All India Institute of Medical Sciences New Delhi, India 110029
2	Mohil Garg	MBBS	Department of Medicine, All India Institute of Medical Sciences New Delhi, India 110029
3	Vignesh Dwarakanathan	MD	Department of Community Medicine, All India Institute of Medical Sciences New Delhi, India 110029
4	Nitesh Gautam	MBBS	Department of Internal Medicine, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA
5	Swasthi S Kumar	MBBS	Department of Medicine, All India Institute of Medical Sciences New Delhi, India 110029
6	Ranveer Singh Jadon	MD	Department of Medicine, All India Institute of Medical Sciences New Delhi, India 110029
7	Mohak Gupta	MBBS	Department of Medicine, All India Institute of Medical Sciences New Delhi, India 110029

8	Animesh Ray* *(*corresponding author)*	DM	Department of Medicine, All India Institute of Medical Sciences New Delhi, India 110029

Financial support: None

***Corresponding Author:**

Dr. Animesh Ray

Room no 3070A, Department of Medicine, Teaching block,

All India Institute of Medical Sciences, New Delhi, India 110029

Email address: doctoranimeshray@gmail.com

Phone:- +91-956-009-3190

ABSTRACT

Infrared thermal screening, via the use of handheld non-contact infrared thermometers (NCITs) and thermal scanners, has been widely implemented all over the world. We performed a systematic review and meta-analysis to investigate its diagnostic accuracy for the detection of fever. We searched PubMed, Embase, the Cochrane Library, medRxiv, bioRxiv, ClinicalTrials.gov, COVID-19 Open Research Dataset, COVID-19 research database, Epistemonikos, EPPI-Centre, World Health Organization International Clinical Trials Registry Platform, Scopus and Web of Science databases for studies where a non-contact infrared device was used to detect fever against a reference standard of conventional thermometers. Forest plots and Hierarchical Summary Receiver Operating Characteristics curves were used to describe the pooled summary estimates of sensitivity, specificity and diagnostic odds ratio. From a total of 1063 results, 30 studies were included in the qualitative synthesis, of which 19 were included in the meta-analysis. The pooled sensitivity and specificity were 0.808 (95%CI 0.656-0.903) and 0.920 (95%CI 0.769-0.975), respectively, for the NCITs (using forehead as the site of measurement), and 0.818 (95%CI 0.758-0.866) and 0.923 (95%CI 0.823-0.969), respectively, for thermal scanners. The sensitivity of NCITs increased on use of rectal temperature as the reference. The sensitivity of thermal scanners decreased in a disease outbreak/pandemic setting. Changes approaching statistical significance were also observed on the exclusion of neonates from the analysis. Thermal screening had a low positive predictive value, especially at the initial stage of an outbreak, while the negative predictive value (NPV) continued to be high even at

later stages. Thermal screening has reasonable diagnostic accuracy in the detection of fever, although it may vary with changes in subject characteristics, setting, index test, and the reference standard used. Thermal screening has a good NPV even during a pandemic. The policymakers must take into consideration the factors surrounding the screening strategy while forming ad-hoc guidelines.

Keywords: COVID-19, Fever, Infection control, Infrared rays, Mass screening, Pandemics, Influenza

INTRODUCTION

The emergence of the SARS virus in 2003 pushed several nations to adopt border control measures. Thermal screening - via the use of thermal scanners (infrared thermal imaging systems) as well as handheld non-contact infrared thermometers (NCITs) - is deemed as the safest tool for screening of temperature during infectious disease outbreaks such as SARS[1], H1N1[2,3] and presently, COVID-19[4,5]. It works on the principle that the human body emits infrared radiation which, like other electromagnetic radiations, can be focused onto a detector that converts heat into electrical signals and displays the temperature of the area as a graphic profile (thermal scanners) or a numerical reading (NCITs)[6]. In the wake of COVID-19, thermal screening has been widely implemented all over the world. These sites include entry and/or exit screening at domestic and international airports[7], defense establishments[8], offices/workplaces, grocery stores, shopping malls, and hotels[9].

Screening for fever with non-contact infrared devices is operationally more favorable, especially in the setting of contagious diseases, over conventional methods of measuring temperature in which the instrument comes in contact with the human body. Potential advantages of using handheld NCITs include reduced discomfort to the subject as well as faster readings[4,10]. Infrared tympanic thermometers, a popular method of contact thermometry, require ear pinna to be pulled manually which may increase the risk of cross-infection, and the use of disposable plastic covers which may increase the financial burden during a disease outbreak[11]. Thermal scanners do not require close proximity to the subject (in contrast to NCITs and contact thermometers) and hence, the operator may be in a remote area to minimize the risk of transmission[5].

The efficacy of thermal screening during a pandemic would depend on several factors including, but not limited to: (a) the diagnostic accuracy of the devices for the detection of fever, and (b) the prevalence of fever in the disease infected individuals. We aimed to conduct a systematic review and meta-analysis to estimate the diagnostic accuracy of NCITs and thermal scanners for the detection of fever.

METHODS

This systematic review was based on the methodological approaches recommended by the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy[12]. This review complies with the Preferred Reporting Items for Systematic Reviews and Meta-analyses of Diagnostic Test Accuracy Studies, the PRISMA-DTA statement[13], and PRISMA-DTA checklist (eMethods 1 in the Supplement).

Database search

We searched the relevant databases for eligible articles without time restriction until May 29, 2020. Our search strategy is provided in eMethods 2 in the supplement. The databases searched for published or ongoing studies included PubMed, Embase, the Cochrane Library, medRxiv, bioRxiv, ClinicalTrials.gov, COVID-19 Open Research Dataset (CORD-19), COVID-19 research database, Epistemonikos, EPPI-Centre, and World Health Organization International Clinical Trials Registry Platform. The reference lists of the included articles and the relevant review articles were manually screened to search for additional studies. To include conference proceedings in our search, we also searched the Scopus and Web of Science databases. We transferred our results to Zotero 5.0 and removed duplicates manually.

Study eligibility

The detailed inclusion and exclusion criteria used in the study are mentioned in eMethods 3 in the Supplement.

Study selection

Two reviewers (NA and MGa) independently screened the articles on the basis of title and abstract to assess for potential inclusion in our study. Following this, full-text versions of articles were accessed and further screened for inclusion. If a clear consensus for a particular study was not reached, the differences were resolved by a collective discussion that included a third reviewer (AR).

Data extraction and qualitative synthesis

From the included studies, data extraction was carried out by two independent reviewers (NA and MGa). Extracted fields included study characteristics (first author name, year of publication, study setting), subject characteristics, index test characteristics (manufacturer, anatomical site), reference test characteristics (method, temperature threshold), and the indices of diagnostic test accuracy.

Methodological quality of included studies

The quality of included studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool (eMethods 4 in the Supplement). This was done independently by two reviewers (NA and MGa). All disagreements were resolved by consensus in consultation with a third reviewer (AR).

Statistical analysis

Data for 2x2 table (true positives, false negatives, false positives, true negatives) was extracted wherever reported or calculated from that provided in individual studies (eTable 1 in the Supplement). In the case of eligible studies where the 2x2 data was ambiguous, an attempt was made to contact the corresponding author and/or coauthors via email or on ResearchGate (https://www.researchgate.net/). If no satisfactory responses were received, the study was excluded from our quantitative synthesis. If one study reported different sets of values for different sites of use or different thresholds of fever, the set of values with the highest Youden's J index was used[14]. The other sets of values were used as appropriate for the subgroup analysis.

Heterogeneity assessment for studies was carried out by visual inspection of the Hierarchical Summary Receiver Operating Characteristics (HSROC) curves (95% prediction region and 95% confidence region). For diagnostic test accuracy reviews, the Cochrane handbook does not recommend the use of the I^2 statistics as it fails to account for the variation due to threshold effect and may overestimate the degree of heterogeneity[12]. To look for threshold effect, Spearman correlation coefficient was derived to look for strong negative correlations between sensitivity and specificity[15]. For detection of publication bias, we used Deeks' funnel plot asymmetry test using the STATA module "midas, pubbias"[16].

The statistical analysis was performed in STATA version 13 (StataCorp, College Station, Texas, USA) using the MIDAS module. The 2x2 table values of true positives, false negatives, false positives, and true negatives were used as input to fit the HSROC curves in order to obtain the pooled estimates of sensitivity, specificity, diagnostic odds ratio (DOR), positive and negative likelihood ratio along with 95% CIs.

A sensitivity analysis was also conducted to investigate the possible influence of neonates (excluding the studies which involved neonates or did not mention age distribution of the sample), the threshold of fever (analysis of studies with fever threshold of <38°C vs ≥38°C by the reference device), type of reference standard (comparison of studies according to different methods for core temperature measurement), disease outbreak (limiting the analysis to studies conducted during a disease outbreak or pandemic), and study setting (comparison of studies conducted in an 'inpatient' vs 'outpatient or airport' setting).

To calculate the statistical significance of the difference between two pooled sensitivities or specificities, we calculated the combined standard error of pooled estimate, followed by the Z statistic[17]. Using this value, the p-value for the difference was calculated. A p-value of <0.05 was considered to denote statistical significance.

We also recorded the positive and negative predictive values (PPV and NPV) in individual studies which were however found to be variable due, in part, to the varying prevalence of fever in different studies. Therefore, an analysis was performed to determine the PPV (the probability of test positives being true positives) and NPV (the probability of test negatives being true negatives) values from the pooled sensitivity and specificity data obtained from our quantitative synthesis. These values were calculated across a wide range of expected fever prevalence during a pandemic (from 0.00001% to 10%) and plotted in a graph using the GraphPad Prism 8 software.

RESULTS

Results of the search

Using our search criteria, we identified a total of 1063 studies, of which 700 were found to be from PubMed, 321 from Embase, 29 from the Cochrane library, 1 from medRxiv, and 12 from screening the reference lists of included articles and relevant review articles. Our literature search flow diagram is summarized in the PRISMA format (Figure 1). A total of 30 studies were included in the qualitative synthesis, of which 19 were included in the quantitative synthesis.

Characteristics of included studies

The 30 studies included in the qualitative synthesis were published between 2004 and 2020 across 15 countries, with most studies conducted in the USA[2,18–20], Singapore[6,21–23], Turkey[24–26], Taiwan[1,27,28], China (including Hong Kong)[29–31], Japan[3,32,33], and others being from Australia[34], Belgium[35], Bolivia[36], England[11], France[37,38], Italy[39], Netherlands[40], New Zealand[41], and Thailand[42]. Characteristics of included studies are summarized in table 1 (studies with the use of NCITs as index test), table 2 (studies with the use of thermal scanners as index test) and table 3 (studies where both NCITs and thermal scanners were used as index test).

Out of the 30 studies included in the qualitative synthesis, 19 were included in the meta-analysis. Most of these studies reported on one index test device (per study) except in the study by Selent et al,[20] where three devices were compared. Hence, we had a total of 21 individual devices, including 10 NCITs[18,20,21,26,29,35,36,38–40] and 11 thermal scanners[1–3,6,20,30–32,37,41]. These 21 devices obtained a total of 13,874 readings from 12,759 patients, with a number of readings ranging from 100 to 2026 per study. Four[1,6,37,41] of the 21 devices did not report the age distribution in the study population, while five[2,3,18,35,40] involved measurements of neonates along with adults or children. These 21 devices were used in different settings, which were classified as inpatient or outpatient/airport. The setting was designated as 'inpatient' setting (n=7)[21,32,35,36,38–40]

when the patients admitted to the hospital wards or the emergency department (ED), were included; while the setting was considered 'outpatient/airport' (n=12)[1–3,18,20,26,29–31,41]when the subjects presented to outpatient centers, clinics, emergency triage (but not admitted to the ED) or were healthy volunteers from a clinic or airport setting. The study by Hamilton et al., where >70% of subjects were clinic attendees and healthy volunteers, was also considered as an outpatient/airport setting[18]. Two of the studies did not mention sufficient information for study setting and were considered under the 'unclassified' setting (n=2)[6,37]. Seven of the 21 devices were used during a pandemic/disease outbreak- SARS[1,21], H1N1[2,3], seasonal influenza[32,41] or COVID-19[29].

The method of reference temperature measurement was variable across our studies. In the studies with the use of NCITs as the index test, the reference device used was tympanic[21,29,38], axillary[26,39], or rectal[35,36,40] thermometer. The studies using thermography as the index test reported their reference test being tympanic[1,6,30,37,41] or axillary[3,32] thermometer. In the study by Hewlett et al., the majority (93%) of participants had oral temperature used as reference[2]. In addition, there were some studies where the reference test was not uniform amongst all included subjects[18,20,31].

Fever thresholds, as per the reference thermometers, varied from 37.3°C to 38.5°C. In the studies where the NCITs were the index test, the prevalence of fever varied from 0.5% to 57.7%. In the studies reporting on thermography, the prevalence of fever ranged from 0.5% to 51.9%. The sensitivity and specificity of the included NCITs ranged from 0.182 to 0.970 and 0.599 to 1, respectively. In the case of thermal scanners, the sensitivity and specificity ranged from 0.148 to 0.929 and 0.310 to 0.997, respectively.

Of the 30 studies from qualitative synthesis, 11 studies were not included in the meta-analysis, due to various reasons (Figure 1): 2x2 data was unavailable[28,33,34] or inconsistent[11,19,23–25,27] or the study characteristics for risk of bias were unavailable[22,42].

Methodological quality of included studies

Results of quality assessment of the included studies (n=19) are summarized as eTable 2 in the Supplement. Overall, 12 of the 19 included studies had a high risk of bias in at least one of the four domains of the QUADAS-2 tool, while 3 studies[18,20,30] had a high risk of bias in two domains.

Quantitative data synthesis

Diagnostic accuracy of handheld non-contact infrared thermometers (NCITs)

Overall, 10 NCIT devices were included in our analysis, which involved a total of 5562 readings. The pooled sensitivity and specificity for NCITs, regardless of the site of temperature measurement, were 0.781 (95%CI 0.628-0.882) and 0.926 (95%CI 0.799-0.975), respectively (eFigure1 in Supplement). When the site of measurement was restricted to the forehead, the pooled sensitivity and specificity were 0.808 (95%CI 0.656-0.903) and 0.920 (95%CI 0.769-0.975), respectively (Figure 2A), which were not significantly different from the pooled measures obtained when the site of measurement was not restricted. In view of maintaining uniformity amongst the included studies and the fact that NCITs are almost exclusively used on the forehead (as seen in our qualitative analysis in Table 1 and stated by the U.S. Food and Drug Administration [FDA][4]), our further analysis was restricted to forehead site only. The DOR was 48.4 (95%CI 19.0-123.7). The pooled positive and negative likelihood ratios were 10.1 (95%CI 3.5-28.7) and 0.2 (95%CI 0.1-0.4). The area under the HSROC curve (Figure 2B) showed an

overall accuracy of 0.92 (95%CI 0.90-0.94). No publication bias was seen on Deeks' funnel plot asymmetry test (p=0.67, Figure 2C).

Diagnostic accuracy of thermal scanners

Amongst thermal scanners, 11 devices were included, which involved a total of 8312 readings. The pooled sensitivity and specificity of the devices was obtained to be 0.818 (95%CI 0.758-0.866) and 0.923 (95%CI 0.823-0.969) (Figure 3A). The DOR was 54.0 (95%CI 16.5-176.4). The pooled positive and negative likelihood ratios were 10.6 (95%CI 4.3-26.4) and 0.2 (95%CI 0.1-0.3). The area under the HSROC curve (Figure 3B) showed an overall accuracy of 0.88 (95%CI 0.85-0.91). No publication bias was seen on Deeks' funnel plot asymmetry test (p=0.07, Figure 3C).

Positive and negative predictive values (PPV and NPV)

As disease spreads in a community, the proportion of infected individuals, and with it, the prevalence of symptoms (fever in the present study) is expected to rise. PPV and NPV for the detection of fever will depend on the prevalence of fever in the community. In our analysis, we observed that PPV rises with an increase in the prevalence of fever for both NCITs and thermal scanners as shown in Figure 4. At an arbitrary prevalence of 1%, the PPV for detection of fever was 9.2% for NCITs and 9.7% for thermal scanners. This means that out of every 10 patients detected febrile by thermal screening, ~one actually turned out to be febrile. Interestingly, in contrast to PPV, there was only a comparatively smaller fall in the values of NPV- 2.3% (from ~100% to 97.7%) for NCITs and 2.1% (from ~100% to 97.9%)- even as the prevalence of fever increased 10^5 fold (Figure 4). This would mean that, even at a fever prevalence of 10% during a

pandemic, a patient who is detected to be afebrile by thermal screening has over a 97% probability of being truly afebrile by the reference method.

Heterogeneity

Wide heterogeneity was observed as demonstrated by visual inspection of the 95% prediction region of the HSROC curves (Figure 2B and 3B). The Spearman correlation coefficient was -0.56 (p=0.09) for NCITs and 0.25 (p=0.45) for thermal scanners, indicating the absence of a threshold effect. Further subgroup analysis was conducted to look for the likely sources of heterogeneity.

Sensitivity analysis

The results of the sensitivity analysis are depicted in Figures 5A and 5B. The forest plots and HSROC curves for these summary estimates are included as eFigures 1-19 in the Supplement.

In the case of handheld NCITs, on the exclusion of the studies on neonates (and where the age distribution of the sample was not mentioned), a difference in the pooled sensitivity (0.89 vs 0.81, p=0.11) and specificity (0.81 vs 0.92, p=0.07) was observed, which approached statistical significance (Figure 5A). Due to the non-availability of enough studies for each individual reference test, the analysis was performed comparing groups of the reference test used: (a) tympanic or axillary, (b) tympanic or rectal, and (c) axillary or rectal temperature. Pooled specificity was observed to be significantly higher, with no difference in sensitivity, when the rectal temperature was used as reference [pooled specificity in group (b) > group (a), p=0.006; pooled specificity in group (c) > group (a), p=0.0003]. There were no differences in the pooled sensitivity or specificity on comparison of studies with a fever threshold of <38°C vs

≥38°C. There were no changes in sensitivity or specificity with the exclusion of studies with a high risk of bias in ≥2 domains. The specificity of NCITs was not found to change in an outpatient/airport setting as compared to an inpatient setting (0.81 vs 0.95, p=0.10). There were only two studies where NCITs were used during a pandemic[21,29], due to which a subgroup analysis could not be performed.

On the exclusion of studies with neonates (and where the age distribution was not mentioned), there was a change in the pooled specificity of thermal scanners (0.86 vs 0.92, p=0.05), which approached statistical significance (Figures 5B). Due to the non-availability of enough studies for oral and axillary reference temperature, the analysis was performed in groups: (a) tympanic temperature only, (b) tympanic or oral, and (c) tympanic or axillary temperature. No differences were observed in the pooled summary estimates between these three groups. There were no differences in the pooled sensitivity or specificity on comparison of devices with a fever threshold of <38°C vs ≥38°C or on the exclusion of studies with a higher risk of bias in ≥2 domains. The sensitivity of thermal scanners was found to fall with their use in a pandemic setting (0.74 vs 0.82; p=0.04). On limiting the analysis to studies from an outpatient or airport setting (i.e. exclusion of studies from the inpatient setting[32] and where the study setting was not reported), there were no changes observed in the pooled summary measures.

DISCUSSION

The results of this review suggest that non-contact infrared thermometers (NCITs) and thermal scanners generally have reasonable sensitivity and specificity for the diagnosis of fever. An increase in the specificity of NCITs was noted when rectal temperature was used as the reference test. The sensitivity of thermal scanners decreased with the use of the devices during a

disease outbreak/ pandemic setting. On the exclusion of neonates from the analysis, differences approaching statistical significance were observed in the sensitivity of NCITs and the specificity of both NCITs and thermal scanners. In the case of both thermal screening devices, there were no changes in the pooled sensitivity or specificity with the exclusion of studies at a high risk of bias or with the comparison of studies with different thresholds for fever. Thermal screening was found to have a low PPV, especially in the initial phase of a disease outbreak in a given community. In contrast, the NPV was seen to be reasonably high even in case of a relatively large proportion of the population being febrile.

Wide heterogeneity was observed in the studies included in our review, in terms of the participant characteristics, the study design and setting, the index tests, and the reference standards used. The demographic details regarding the study participants were not available in some of our included studies. There was non-uniformity in the reference standard used for the confirmation of fever. In addition, differences in the type of index test used (NCITs/thermal scanners), the manufacturer specifications, the environmental conditions for optimal operation and the experience of the operator can lead to inaccuracies in the measurement of temperature and a further increase in heterogeneity.

Several factors can influence the detection of fever by infrared thermal devices[6,43]. Environmental factors such as absolute temperature, variation in the temperature, relative humidity, etc. play an important role in the accuracy of measurement. NCITs should not be used in direct sunlight or near radiant heat sources[4]. Factors related to the screened subject that may result in false negative readings include application of make-up on the target area, use of antipyretics or significant perspiration. At the stage of fever initiation, the rise in the hypothalamic set point is accompanied by cutaneous vasoconstriction, which may lead to cooling

of skin and a false negative reading on the thermal scanner[44]. On the other end, false positive results may be seen in subjects who are menstruating, pregnant, on hormone replacement therapy, or have recently consumed alcohol, hot beverages or have recently done strenuous physical activity[6]. These factors may have played in role in the low PPV observed in our study.

The target body site for the measurement may be subject to differential vascularity leading to variation in heat distribution. Forehead is a more feasible site for scanning but is thought to be more prone to physiological and environmental variations. On the other hand, sites such as external auricular area and inner eye canthi[6,45] are less subject to variations but are not as accessible and the removal of eyewear, scarves, etc. may lengthen the preparation time for the subject[43]. Wrist temperature may be useful since rolling up the sleeves may not lengthen the preparation time significantly[29]. In our study, we found no significant changes in the pooled sensitivity or specificity when the analysis was restricted to the forehead as the site.

Disease outbreaks, such as the COVID-19, necessitate the use of a screening device wherein the sensitivity of the device plays a vital role, as false negatives should be minimized at all costs. In a pandemic setting, the sensitivity of thermography decreased significantly in our analysis. This may be linked to the use of thermal scanners for mass screening[1,3], contrary to the recommendations by the FDA, which state that only one person's temperature should be measured at a time[5]. Any face obstructions such as masks, glasses, headbands or scarves must also be removed prior to screening with a thermal scanner; this may be challenging to enforce in a pandemic situation. Incidentally, the FDA recommends confirmation of a positive result on thermal scanner with a secondary method of evaluation, such as an NCIT or a contact thermometer[46].

On the exclusion of neonates from the analysis, differences approaching statistical significance were observed in the sensitivity of NCITs and the specificity of both NCITs and thermal scanners. Several factors, unique to neonates, may hamper the detection of fever by infrared devices as well as reference tests. Neonates are more prone to temperature instability from ambient temperature changes due to a higher evaporative heat loss, higher metabolic rates and inability to make behavioral adaptations[47]. Discomfort to the baby during handling may affect the rectal, oral and axillary measurements, as well as make it challenging to achieve an optimal viewing angle for the use of infrared devices. Additionally, infants have brown fat located in their axillary pockets, which takes part in non-shivering thermogenesis, and hence, may affect the axillary temperature measurements[47,48].

In our analysis, we found that thermal screening had a high NPV for fever but there was considerable variation in PPV with change in fever prevalence. On assuming a fever prevalence of 1%, the NPV obtained in our study (99.8%, both for NCITs and thermal scanners) agrees well with the results obtained by Bitar et al. (>99%)[43]. But, it is generally in the early stages of a pandemic (prevalence of fever<1%) that thermal screening is used as a means of delaying the introduction of infection in the given community[43]. At these initial stages, we found thermal screening to have a poor PPV, meaning that most of the subjects deemed to be febrile on screening would turn out to be afebrile (false positives), which may also evoke undue anxiety and anguish amongst these individuals[43].

In addition to the concerns about the diagnostic accuracy, there are other factors that determine if thermal screening is relevant in the case of COVID-19. Being a symptom-based surveillance approach, thermal screening will be unable to identify asymptomatic (estimated 40-45% of infections[49]) or presymptomatic individuals (account for 30-60% of total transmission[50–

[52]). In addition, a Centers for Disease Control and Prevention (CDC) report (n=373,883) showed only 43.1% of the COVID-19 infected individuals to have fever[53]. Other studies have reported variable prevalence of fever[54–57], suggesting that fever is far from a universal finding at presentation. In our analysis, we observed that thermal screening will be able to detect ~81% of these febrile individuals (sensitivity of NCITs: 80.8%, thermal scanners: 81.8%). This implies that a high proportion of infected individuals (afebrile and/or false negatives) would be missed at thermal screening, which can drastically multiply the risk of spread in the community. A simulation study in an airport setting estimated that thermal screening at airports would miss 46% of travelers with COVID-19[58]. Similar results have been obtained earlier in SARS[59] and H1N1 influenza[3] epidemics. An international experts committee led by Bell et al. reported that thermal scanning of over 35 million travelers at borders did not detect any incoming SARS cases and hence, had little role in infection control[60]. Clifford et al. reported that syndromic screening of air travelers at entry or exit along with their sensitization at arrival only delayed the local spread of SARS-CoV-2 by a few days[61]. Therefore, temperature screening alone does not appear to be an effective way to detect cases and to help curb the international spread of COVID-19. Despite the psychological reassurance provided by thermal screening, public health officials and policymakers must take into consideration the quality of scientific evidence that drives such measures and the guidelines must reflect a wholesome approach to the prevention of community transmission. A recent study suggested that the best strategy to reopen travel restrictions is the administration of COVID-19 test to all incoming travelers followed by isolation of test positives[62]. While it is important to rule out more common infections like COVID-19, other imported infections must also be taken into consideration in the workup of febrile travelers[63].

This study had a few limitations. First, there was high heterogeneity across the studies, which persisted even on subgroup analysis. Second, in our overall analysis including all NCITs and thermal scanners, only the single best sets of values (with the highest Youden's index) for each of the 21 devices were considered. Hence, our estimates of pooled sensitivity and specificity may reflect the best-case parameters of diagnostic accuracy for the included devices, which may be higher than in the case where the other sets of 2x2 data values are considered. Third, there were several included studies where the index test temperature threshold for fever was not pre-specified but obtained retrospectively from the study data, making them less reliable.

CONCLUSIONS

Handheld non-contact infrared thermometers (NCITs) and thermal scanners have a reasonable sensitivity and specificity in detecting fever. However, variation in the diagnostic performance was observed in different study settings: notably, an increase in specificity of NCITs with the use of rectal temperature as reference, and differences in sensitivity of NCITs and specificity of both NCITs and thermal scanners with the exclusion of neonate subjects. Despite an observed fall in the sensitivity of thermal scanners in a pandemic setting, our study shows that the NPV continues to be high even when the disease affects a large proportion of the community. Thermal screening may be considered as a method of detection of fever in symptomatic individuals, but only as a part of a larger approach to pandemic response. The demographic, epidemiological, environmental, and psychosocial factors that surround the screening strategy must be taken into consideration, both by present public health policymakers as well as future researchers.

Author contributions

NA, VD, and AR completed study design. Literature search and data extraction were carried out by NA and MGa. Assessment of quality of included studies was carried out by NA and MGa under the supervision of VD and AR. Qualitative synthesis was carried out by all authors. Data analysis was performed by NA and VD. AR consulted on the analysis. All authors were involved in the interpretation of the results. NA and MGu created the figures. NA and MGa drafted the initial manuscript. The manuscript was revised by NA, VD, MGu and AR, and approved for final submission by all authors.

Funding

No financial support was obtained for this study.

Conflicts of interest/Disclosure

The authors have declared no conflict of interest.

REFERENCES

1. Chiu WT, Lin PW, Chiou HY, et al. Infrared thermography to mass-screen suspected SARS patients with fever. *Asia Pac J Public Health* 2005;17:26-28.

2. Hewlett AL, Kalil AC, Strum RA, et al. Evaluation of an infrared thermal detection system for fever recognition during the H1N1 influenza pandemic. *Infect Control Hosp Epidemiol* 2011;32:504-506.

3. Nishiura H, Kamiya K. Fever screening during the influenza (H1N1-2009) pandemic at Narita International Airport, Japan. *BMC Infect Dis* 2011;11:111.

4. Health C for D and R. Non-contact Infrared Thermometers. *FDA* 2020. Available at: https://www.fda.gov/medical-devices/general-hospital-devices-and-supplies/non-contact-infrared-thermometers. Accessed June 25, 2020.

5. Health C for D and R. Thermal Imaging Systems (Infrared Thermographic Systems / Thermal Imaging Cameras). *FDA* 2020. Available at: https://www.fda.gov/medical-devices/general-hospital-devices-and-supplies/thermal-imaging-systems-infrared-thermographic-systems-thermal-imaging-cameras. Accessed June 25, 2020.

6. Ng EYK, Kaw GJL, Chang WM. Analysis of IR thermal imager for mass blind fever screening. *Microvasc Res* 2004;68:104-109.

7. Anon. LAWA Official Site | News Release | June 22, 2020. Available at: https://www.lawa.org/news-releases/2020/news-release-038. Accessed July 17, 2020.

8. Anon. Department Uses Thermal Imaging to Detect COVID-19. *US DEPARTMENT OF DEFENSE*. Available at: https://www.defense.gov/Explore/News/Article/Article/2178320/department-uses-thermal-imaging-to-detect-covid-19/. Accessed July 14, 2020.

9. Anon. MoHFW | Home. Available at: https://www.mohfw.gov.in/?aFaqsOrder=Sorter_faqID&aFaqsDir=ASC&aWnDir=ASC&s_keyword=covid&aWnOrder=Sorter_wnid. Accessed June 20, 2020.

10. Wang K, Gill P, Wolstenholme J, et al. Non-contact infrared thermometers for measuring temperature in children: primary care diagnostic technology update. *Br J Gen Pract* 2014;64:e681-683.

11. Hayward G, Verbakel JY, Ismail FA, 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* 2020;70:e236-e244.

12. Deeks J, Bossuyt PM, Gatsonis C. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0: The Cochrane Collaboration. 2010. Available at: /resource/cochrane-handbook-systematic-reviews-diagnostic-test-accuracy. Accessed July 13, 2020.

13. McInnes MDF, Moher D, Thombs BD, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA* 2018;319:388-396.

14. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32-35.

15. Devillé WL, Buntinx F, Bouter LM, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol* 2002;2:9.

16. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005;58:882-893.

17. Anon. Practical Statistics For Medical Research. *ResearchGate*. Available at: https://www.researchgate.net/publication/273097098_Practical_Statistics_For_Medical_Research. Accessed July 14, 2020.

18. Hamilton PA, Kasbekar RS, Monro R. Clinical performance of infrared consumer-grade thermometers. *J Nurs Meas* 2013;21:166-177.

19. 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;16:1710-1717.

20. Selent MU, Molinari NM, Baxter A, et al. Mass screening for fever in children: a comparison of 3 infrared thermal detection systems. *Pediatr Emerg Care* 2013;29:305-313.

21. Ng DK, Chan C-H, Lee RS, et al. Non-contact infrared thermometry temperature measurement for screening fever in children. *Ann Trop Paediatr* 2005;25:267-275.

22. Tan YH, Teo CW, Ong E, et al. Development and deployment of infrared fever screening systems. In: *Thermosense XXVI*.Vol 5405. International Society for Optics and Photonics; 2004:68-78. Available at: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/5405/0000/Development-and-deployment-of-infrared-fever-screening-systems/10.1117/12.542993.short. Accessed May 22, 2020.

23. Tay MR, Low YL, Zhao X, et al. Comparison of Infrared Thermal Detection Systems for mass fever screening in a tropical healthcare setting. *Public Health* 2015;129:1471-1478.

24. Apa H, Gözmen S, Bayram N, et al. Clinical accuracy of tympanic thermometer and noncontact infrared skin thermometer in pediatric practice: an alternative for axillary digital thermometer. *Pediatr Emerg Care* 2013;29:992-997.

25. Apa H, Gözmen S, Keskin-Gözmen Ş, et al. Clinical accuracy of non-contact infrared thermometer from umbilical region in children: A new side. *Turk J Pediatr* 2016;58:180-186.

26. Ataş Berksoy E, Bağ Ö, Yazici S, et al. Use of noncontact infrared thermography to measure temperature in children in a triage room. *Medicine (Baltimore)* 2018;97. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5805428/. Accessed June 23, 2020.

27. Chiang M-F, Lin P-W, Lin L-F, et al. Mass screening of suspected febrile patients with remote-sensing infrared thermography: alarm temperature and optimal distance. *J Formos Med Assoc* 2008;107:937-944.

28. 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;25:1109-1111.

29. Chen G, Xie J, Dai G, et al. Validity of Wrist and Forehead Temperature in Temperature Screening in the General Population During the Outbreak of 2019 Novel Coronavirus: a prospective real-world study. *medRxiv* 2020:2020.03.02.20030148.

30. Chan L-S, Cheung GTY, Lauder IJ, et al. Screening for fever by remote-sensing infrared thermographic camera. *J Travel Med* 2004;11:273-279.

31. Chan LS, Lo JLF, Kumana CR, et al. Utility of infrared thermography for screening febrile subjects. *Hong Kong Med J* 2013;19:109-115.

32. Sun G, Saga T, Shimizu T, et al. Fever screening of seasonal influenza patients using a cost-effective thermopile array with small pixels for close-range thermometry. *Int J Infect Dis* 2014;25:56-58.

33. Suzuki T, Wada K, Wada Y, et al. The validity of mass body temperature screening with ear thermometers in a warm thermal environment. *Tohoku J Exp Med* 2010;222:89-95.

34. McBride WJH, Buikstra E, FitzGerald M. Investigation of febrile passengers detected by infrared thermal scanning at an international airport. *Aust N Z J Public Health* 2010;34:5-10.

35. Allegaert K, Casteels K, van Gorp I, et al. Tympanic, infrared skin, and temporal artery scan thermometers compared with rectal measurement in children: a real-life assessment. *Curr Ther Res Clin Exp* 2014;76:34-38.

36. Teran CG, Torrez-Llanos J, Teran-Miranda TE, et al. Clinical accuracy of a non-contact infrared skin thermometer in paediatric practice. *Child Care Health Dev* 2012;38:471-476.

37. Bardou M, Seng P, Meddeb L, et al. Modern approach to infectious disease management using infrared thermal camera scanning for fever in healthcare settings. *J Infect* 2017;74:95-97.

38. Hausfater P, Zhao Y, Defrenne S, et al. Cutaneous Infrared Thermometry for Detecting Febrile Patients. *Emerg Infect Dis* 2008;14:1255-1258.

39. Chiappini E, Sollai S, Longhi R, et al. Performance of non-contact infrared thermometer for detecting febrile children in hospital and ambulatory settings. *J Clin Nurs* 2011;20:1311-1318.

40. Paes BF, Vermeulen K, Brohet RM, et al. Accuracy of tympanic and infrared skin thermometers in children. *Arch Dis Child* 2010;95:974-978.

41. Priest PC, Duncan AR, Jennings LC, et al. Thermal image scanning for influenza border screening: results of an airport screening study. *PLoS One* 2011;6:e14490.

42. Sumriddetchkajorn S, Chaitavon K. Field test studies of our infrared-based human temperature screening system embedded with a parallel measurement approach. *Infrared Physics & Technology* 2009;52:119-123.

43. 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.

44. Ng EYK. Is thermal scanner losing its bite in mass screening of fever due to SARS? *Med Phys* 2005;32:93-97.

45. Ring EFJ, Jung A, Kalicki B, et al. New standards for fever screening with thermal imaging systems. *J Mech Med Biol* 2013;13:1350045.

46. Health C for D and R. Enforcement Policy for Telethermographic Systems During the Coronavirus Disease 2019 (COVID-19) Public Health Emergency. *US Food and Drug Administration* 2020. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/enforcement-policy-telethermographic-systems-during-coronavirus-disease-2019-covid-19-public-health. Accessed July 9, 2020.

47. Martin SA, Kline AM. Can there be a standard for temperature measurement in the pediatric intensive care unit? *AACN Clin Issues* 2004;15:254-266.

48. Duran R, Vatansever U, Acunaş B, et al. Comparison of temporal artery, mid-forehead skin and axillary temperature recordings in preterm infants <1500 g of birthweight. *J Paediatr Child Health* 2009;45:444-447.

49. Oran DP, Topol EJ. Prevalence of Asymptomatic SARS-CoV-2 Infection: A Narrative Review. *Ann Intern Med* 2020.

50. He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine* 2020;26:672-675.

51. Chun JY, Baek G, Kim Y. Transmission onset distribution of COVID-19. *Int J Infect Dis* 2020.

52. Hu S, Wang W, Wang Y, et al. Infectivity, susceptibility, and risk factors associated with SARS-CoV-2 transmission under intensive contact tracing in Hunan, China. *medRxiv* 2020.

53. Stokes EK. Coronavirus Disease 2019 Case Surveillance — United States, January 22–May 30, 2020. *MMWR Morb Mortal Wkly Rep* 2020;69. Available at: https://www.cdc.gov/mmwr/volumes/69/wr/mm6924e2.htm. Accessed June 20, 2020.

54. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* 2020;323:2052-2059.

55. Guan W-J, Ni Z-Y, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* 2020;382:1708-1720.

56. Jutzeler CR, Bourguignon L, Weis CV, et al. Comorbidities, clinical signs and symptoms, laboratory findings, imaging features, treatment strategies, and outcomes in adult and pediatric patients with COVID-19: A systematic review and meta-analysis. *Travel Med Infect Dis* 2020;37:101825.

57. Grant MC, Geoghegan L, Arbyn M, et al. The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies from 9 countries. *PLoS ONE* 2020;15:e0234765.

58. Quilty BJ, Clifford S, Flasche S, et al. Effectiveness of airport screening at detecting travellers infected with novel coronavirus (2019-nCoV). *Euro Surveill* 2020;25. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7014668/. Accessed June 20, 2020.

59. Gumel AB, Ruan S, Day T, et al. Modelling strategies for controlling SARS outbreaks. *Proc Biol Sci* 2004;271:2223-2232.

60. Bell DM, World Health Organization Working Group on International and Community Transmission of SARS. Public health interventions and SARS spread, 2003. *Emerging Infect Dis* 2004;10:1900-1906.

61. Clifford S, Pearson CAB, Klepac P, et al. Effectiveness of interventions targeting air travellers for delaying local outbreaks of SARS-CoV-2. *J Travel Med* 2020;27.

62. Dickens BL, Koo JR, Lim JT, et al. Strategies at points of entry to reduce importation risk of COVID-19 cases and re-open travel. *J Travel Med* 2020.

63. Norman FF, Chamorro-Tojeiro S, Crespillo-Andújar C, et al. Travel-related fever in the time of COVID-19 travel restrictions. *J Travel Med* 2020. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7454772/. Accessed September 18, 2020.

Address for correspondence:

Dr. Animesh Ray,

Room no 3070A, Teaching block,

Department of Medicine,

All India Institute of Medical Sciences,

New Delhi, India

Email id: doctoranimeshray@gmail.com

Phone:- +91-95600-93190

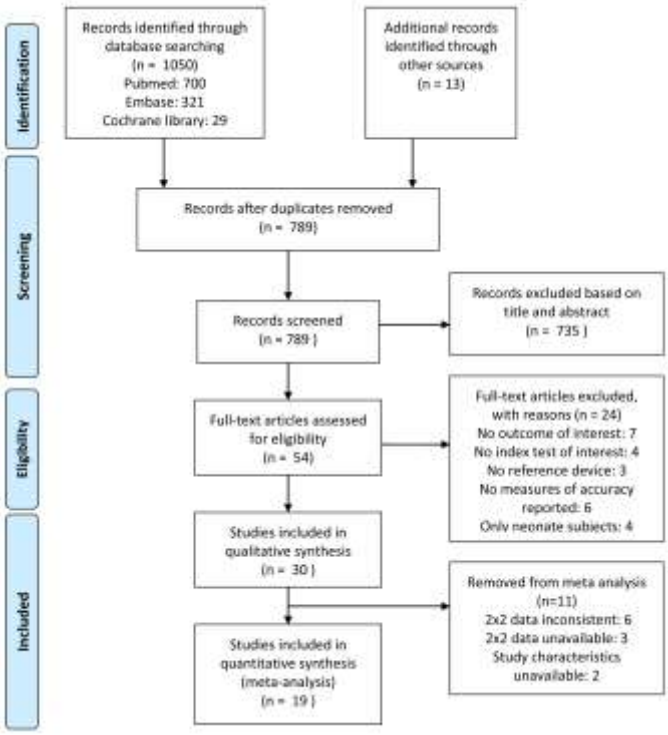
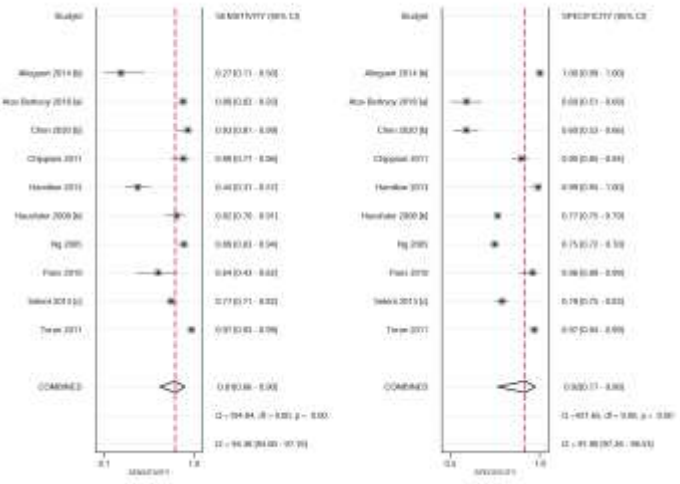
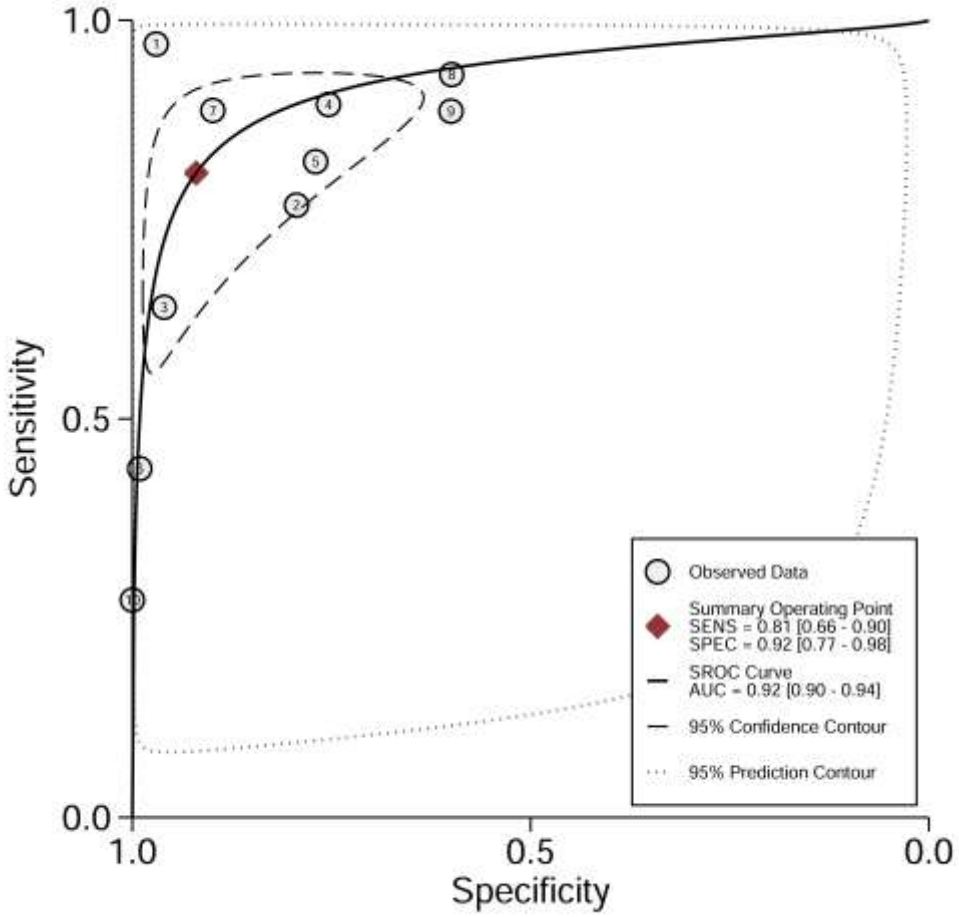


Figure 1: Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram of the study selection process.




1.0
0.5
0.0
Sensitivity
1.0
0.5
0.0
Specificity
Observed Data
Summary Operating Point
SENS = 0.81 [0.66 - 0.90]
SPEC = 0.92 [0.77 - 0.98]
SROC Curve
AUC = 0.92 [0.90 - 0.94]
95% Confidence Contour
95% Prediction Contour

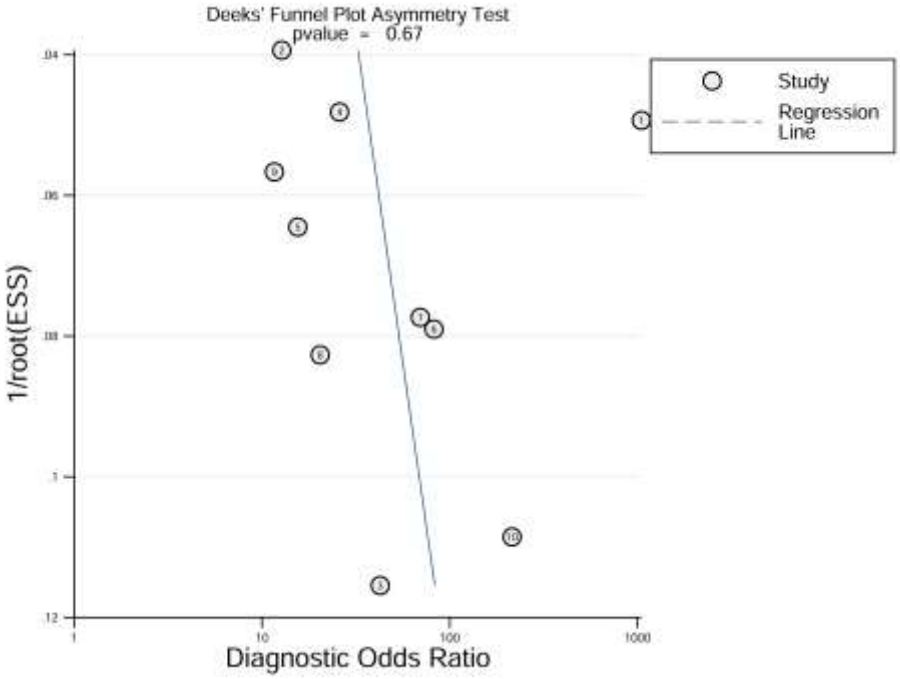
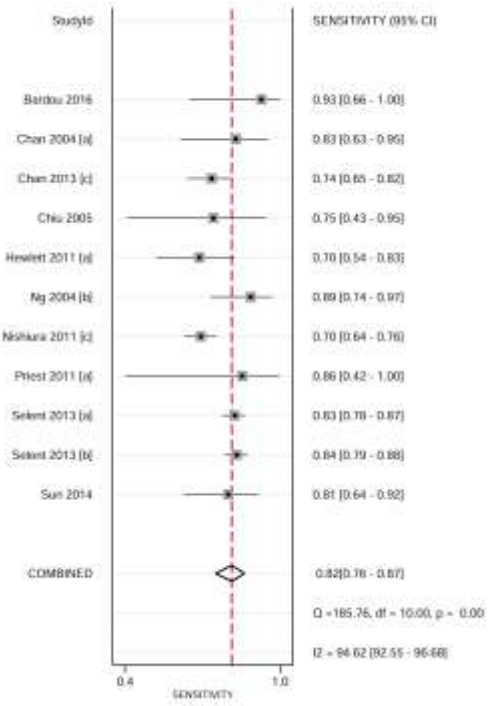
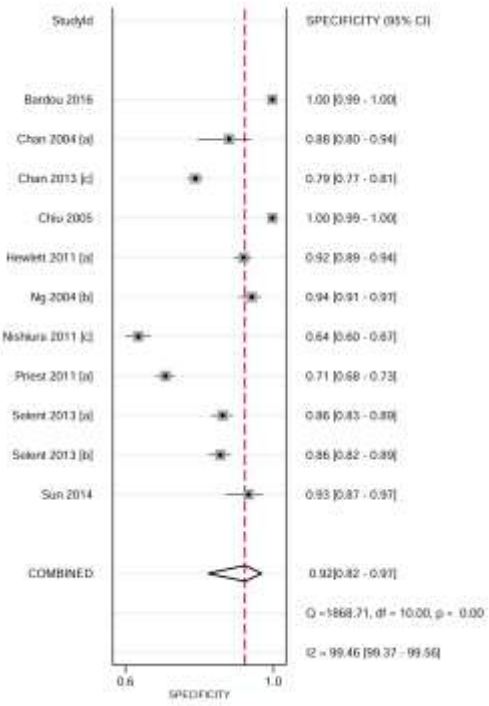
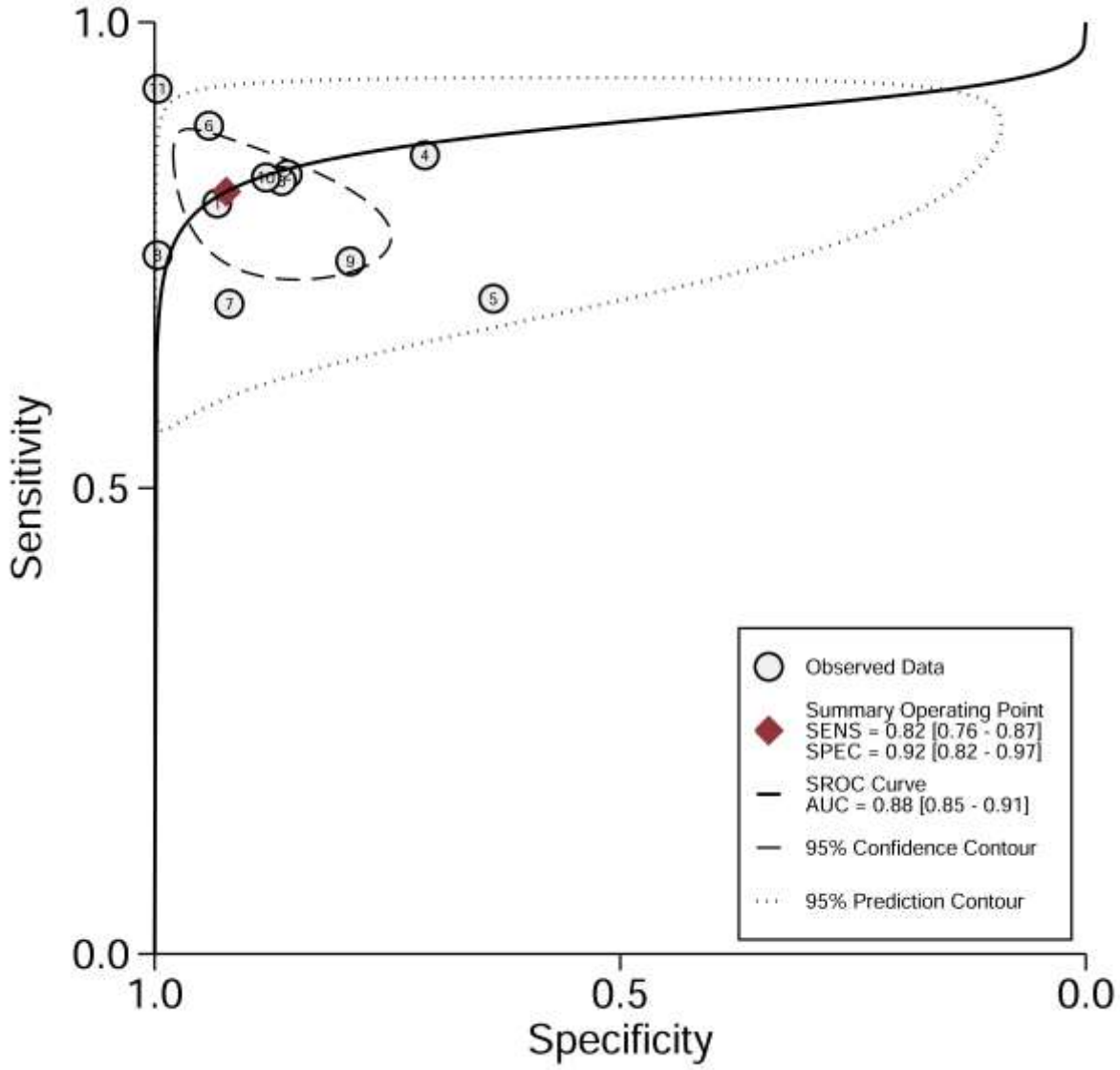


Figure 2: Quantitative analysis for the overall diagnostic accuracy of handheld non-contact infrared thermometers (NCITs) for the detection of fever with forehead as the site of measurement. (A) Forest plot showing pooled sensitivity and specificity; (B) Hierarchical Summary Receiver Operating Characteristic (HSROC) curves;(C) Funnel plot depicting publication bias.






1.0
0.5
0.0
Sensitivity
1.0
0.5
0.0
Specificity
Observed Data
Summary Operating Point
SENS = 0.82 [0.76 - 0.87]
SPEC = 0.92 [0.82 - 0.97]
SROC Curve
AUC = 0.88 [0.85 - 0.91]
95% Confidence Contour
95% Prediction Contour

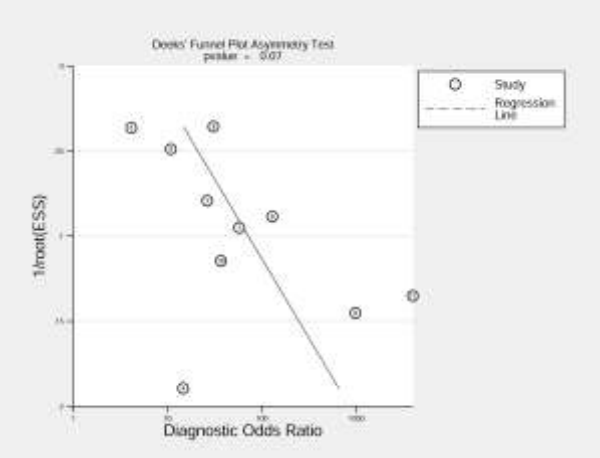


Figure 3: Quantitative analysis for the overall diagnostic accuracy of thermal scanners for the detection of fever. (A) Forest plot showing pooled sensitivity and specificity; (B) Hierarchical Summary Receiver Operating Characteristic (HSROC) curves; (C) Funnel plot depicting publication bias.

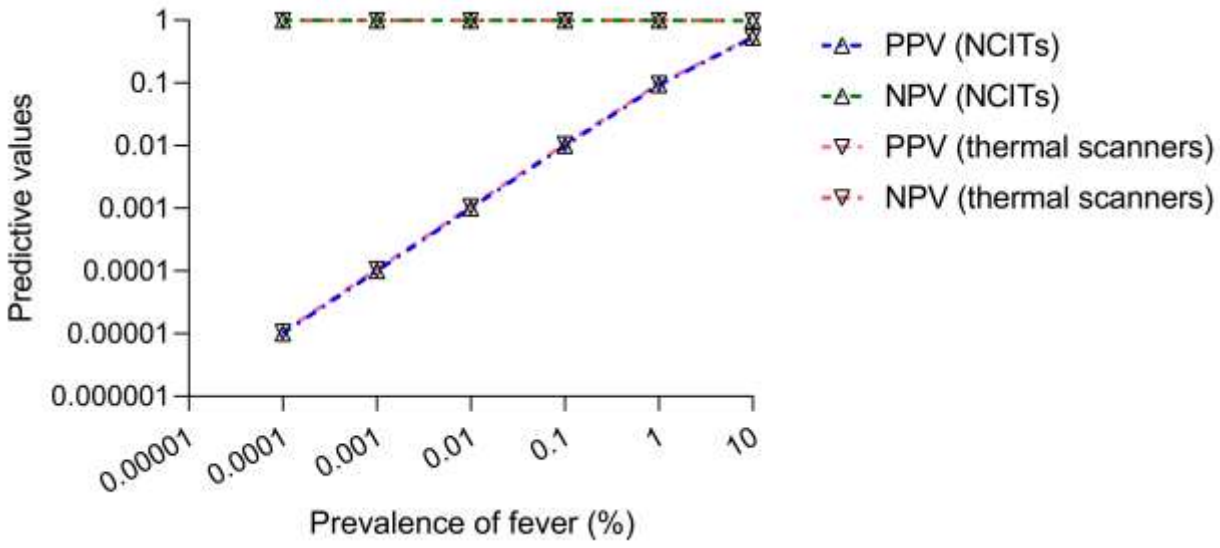


Figure 4: Positive and negative predictive values (PPV and NPV) for the thermal screening devices (handheld non-contact infrared thermometers and thermal scanners) with change in prevalence of fever in a given community.

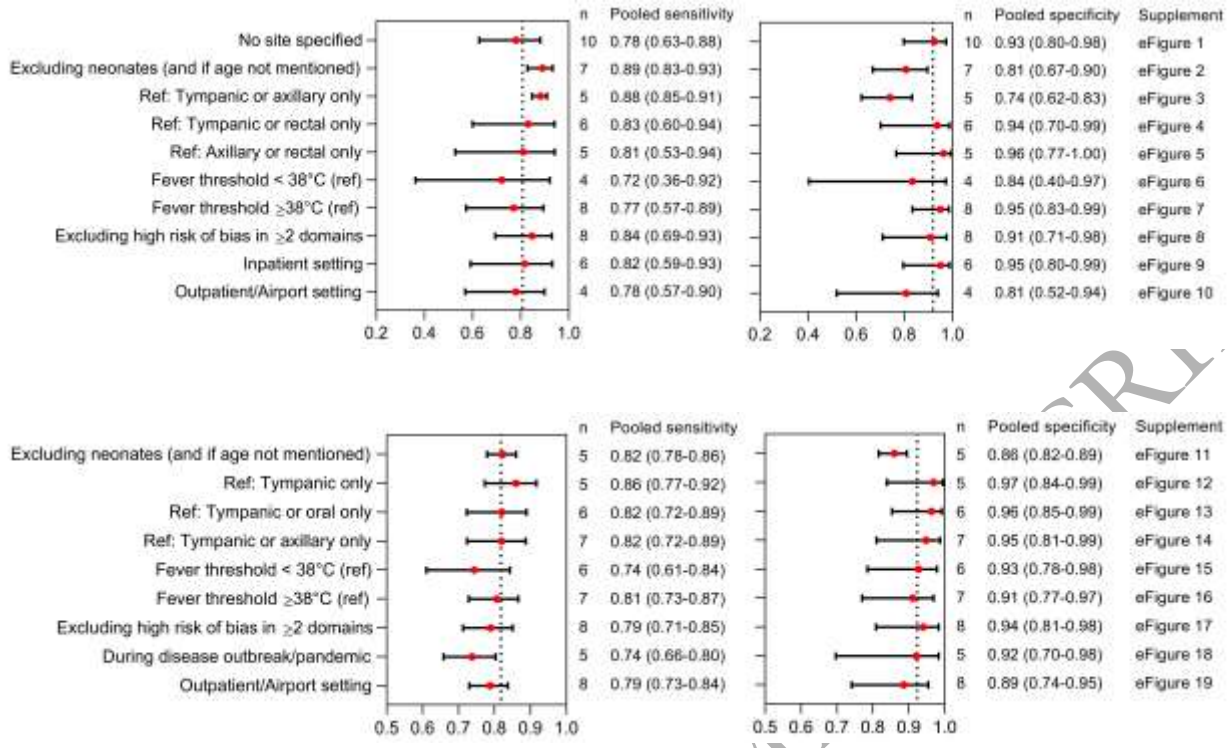


Figure 5: Depiction of sensitivity analysis with pooled sensitivity and specificity for each of our subgroups. (A) Pooled sensitivity and specificity of handheld non-contact infrared thermometers (NCITs) in different subgroups. (Dotted lines: Overall pooled sensitivity and specificity of NCITs with forehead as the site of measurement) (B) Pooled sensitivity and specificity of thermal scanners in different subgroups. (Dotted line: Overall pooled sensitivity and specificity of thermal scanners). Refer to eFigures 1-19 in the Supplement for individual forest plots and HSROC curves. (n: Number of index test devices included in the subgroup; Ref: Reference test)

Table 1:Characteristics of the included studies with handheld non-contact infrared thermometers (NCITs) as the index test.

Author, Year of publication	**Setting**	**Sample characteristics**	**Index test device**	**Reference test cut off value (°C)**	**Site for index test**	**AUC (95 %CI)**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**PPV (95% CI)**	**NPV (95% CI)**
Allegaertet al., 2014[35]	Pediatric patients admitted to different pediatric wards of University Hospitals, Belgium	294 children Age 0.02-17 yrs (median 3.2 years)	Infrared skin scan thermometer (ThermoFlash Contactless Medical Electronic Thermometer, Visiomedlab, Paris, France)	Rectal temperature 37.8			0.18	0.99	0.86	0.91
				38			0.27	1	1	0.92
Apaet al., 2013[24]	Pediatric population hospitalized in Infectious Disease Unit of DrBehcetUz Children's Training and Research	50 children (30 male), total 1639 readings for each device Mean age: 54 months	ThermoFlash LX-26, Visiomed SAS France, Paris, France	Axillary temperature: 38	Forehead	0.96	0.94	0.90	0.66 (0.58 – 0.73)	0.95 (0.94 – 0.96)

	Hospital, Turkey									
Apaet al., 2016[25]	Pediatric population hospitalized in Infectious Disease Unit of Dr Behcet Uz Children's Training and Research Hospital, Turkey	100 children (53 male), total 2048 readings for each device Age range: between 1 and 168 months Mean age: 56.3 ± 50.2 months	Thermo Flash LX-26, Visiomed SAS France, Paris/France	Axillary temperature : 38	1.5 cm below umbilicus	0.93 (0.92-0.94)	0.71 (0.67-0.75)	0.95 (0.94 - 0.96)	0.86 (0.82 - 0.89)	0.90 (0.89 – 0.92)
Ataş Berksoy et al., 2018[26]	Pediatric population at emergency	319 children (176 male)	IFR thermoscope (model DT-8806)	Axillary temperature: 37.5	Forehead	0.791 (0.740-0.841)	0.88	0.60		

	triage room of DrBehçetuz Children Teaching Hospital, Turkey	Age range: 1 month – 18 years. Median age: 30 months			Neck	0.815 (0.767-0.864)	0.75	0.77		
					Nape	0.810 (0.762-0.859)	0.78	0.71		
Chenet al., 2020*[29]	Ningbo First Hospital, China	261 participants ("Indoor participants") from fever clinic and emergency department		Tympanic temperature: 37.3 C	10 cm above the palmar side of wrist	0.790 (0.725-0.854)	0.86	0.67		
					Mid forehead	0.816 (0.757-0.876)	0.93	0.60		
Chiappiniet al., 2011[39]	5 centre study (3paediatric clinics, 1 paediatric emergency department, 1 primary health centre) in 5 cities, Italy	251 childrden (127 male) (age> 1mo) Median age 4.5years (3-8.6 IQR)	Thermofocus, model 0800; Tecnimed, Varese, Italy	Axillary temperature: 38	mid-forehead	0.968 (0.949 – 0.986)	0.89 (0.80-0.97)	0.90 (0.86-0.94)	0.70 (0.59-0.81)	0.97 (0.94-0.99)

Hamilton et al., 2013[18]	Adult subjects recruited from Scripps Hospital, La Jolla, California and pediatric subjects, (including subjects 0-12 mo of age)recruited from Advanced Pediatric Medical Group, San Diego, California.	171 patients (90 male) Age range: 4 days to 87 years (52 patients in 0-12 mo of age)	Visiomed SAS Thermoflash LX-26	< 5 yrs: Axillary temperature: 38 >5yrs: Oral temperature: 38	Forehead		0.44	0.99		
Hausfater et al., 2008[38]	Emergency department of a large academic hospital, Paris, France	2026 patients (1146 men) Mean age 46 ± 19 years (range 6–103	Infrared thermometer (Raynger MX; Raytek, Berlin, Germany)	Tympanic temperature 37.5	Forehead	0.935 (0.876-0.966)	0.76 (0.69-0.82)	0.65 (0.63-0.67)	0.16 (0.14-0.19)	0.97 (0.96-0.98)
				38.0 (best)		0.873 (0.807-0.917)	0.82 (0.71-0.90)	0.77 (0.76-0.79)	0.10 (0.08-0.13)	0.99 (0.99-1.00)

		years)		38.5		0.792 (0.749-0.829)	0.82 (0.67-0.91)	0.90 (0.88-0.91)	0.13 (0.09-0.18)	1.00 (0.99-1.00)
Haywardet al., 2020[11]	GP practice (nine sites) and out-of-hours (OOH) service (one site), Oxfordshire, UK	401 children [203 male; median age 1.6 years (IQR 0.79 to 3.38 years)] Five children were <4 weeks old.	Thermofocus 0800	Axillary temperature: 38			0.293 (0.161 – 0.455)	0.969 (0.946 – 0.985)	0.522 (0.306 – 0.732)	0.923 (0.892 – 0.948)
			Firhealth forehead thermometer				0.049 (0.006 – 0.165)	0.989 (0.972 – 0.997)	0.333 (0.043 – 0.777)	0.901 (0.868 – 0.929)
Liuet al., 2004[28]	Outpatient department of a medical center, Taiwan	500 patients	Thermofocus Thermometer, Tecnimed, Italy	Tympanic temperature :37.5	Forehead		0.17	0.98		
					Auditory meatus		0.82	0.98		

Nget al., 2005[21]	Children admitted to general paediatric department, Kwong Wah Hospital, Hong Kong	1000 readings from 567 children (335 male) Median age= 2 years	Standard ST 8812 (Standard Instruments Co., Hong Kong SAR, China)	Tympanic temperature >38 FirstTempH Genius, Intelligent Medical Systems Inc., Carlsbad, California	Forehead	0.868	0.89 (0.83 – 0.93)	0.75 (0.74 – 0.76)	0.33 (0.31 – 0.35)	0.98 (0.96 – 0.98)
Paeset al., 2010[40]	Children admitted to general pediatric ward of Spaarne Hospital, Netherlands	100 children (50 male) Age range 2wks to 18 yrs (Mean = 3.24 years)	Thermofocus 700 A2, Technimed, Varese, Italy	Rectal temperature: 38			0.64	0.96	0.84	0.89
Teranet al., 2011[36] _1264471	Triage of emergency room and inpatient unit of Pediatric Hospital Albina	434 (208male) Age range 1-48 months (Mean: 14.6 mo)	Thermofocus, model 01500, TECNIMED, Varese, Italy	Rectal temperature: 38	Forehead center		0.97 (0.93 -0.99)	0.97 (0.94-0.99)	0.95 (0.91-0.98)	0.98(0.95 - 0.99)

	R. de Patiño, Bolivia									
* Preprint publication										

Table 2: Characteristics of the included studies withthermal scanners as the index test.

Author, Year of publication	**Setting**	**Sample characteristics**	**Index test device**	**Reference test cut off value (°C)**	**Site for index test**	**AUC (95 % CI)**	**Sensitivity (95 % CI)**	**Specificity (95 % CI)**	**PPV (95 % CI)**	**NPV(95% CI)**
Bardouet al., 2016[37]	University hospital centre, Southern France (Both Patients and healthcare workers)	625 subjects	MOBOTIX M15D infrared thermal camera (MOBOTIX, Germany)	Tympanic temperature: 38.5			0.928 (0.661-0.998)	0.996 (0.988-0.999)	0.866 (0.595-0.983)	0.998 (0.990 -1)
Chan et al., 2004[30]	Queen Mary Hospital, two health clinics and the University of Hong Kong Sports Center (USC)	198 readings from 176 patients (98 male)	FLIR thermovision-Three different infrared cameras (models PM595, SC320C and S60)	Aural infrared temperature: 38	Ear pinna (n=116)		0.67	0.96		
					Forehead (n=188)		0.04	0.99		
Chan et al., 2013[31]	Accident and Emergency Department, Queen Mary	1517 patients (747 male) Mean age-	FLIR Systems ThermaCAM S40 infrared camera with 24° lens	core temperature: 38 Oral or aural temperature (If	Max frontal temperature ("AREAMAX")	0.812 (0.761 - 0.863)	0.64	0.86	0.27	0.97

	Hospital, Hong Kong	45.8 yrs		both available, higher reading was used.)	Forehead	0.780 (0.723 - 0.837)	0.50	0.93	0.36	0.96
					Maximum lateral temperature ("LATMAX")	0.815 (0.763 - 0.867)	0.74	0.79	0.22	0.97
Chiang et al., 2008[27]	People visiting Municipal Wang Fang Hospital, Taipei, Taiwan	1032 subjects	Digital infrared thermal imaging (DITI) (Spectrum 9000MB Medical Thermal Imaging System; Telesis Technologies Inc., Kaohsiung, Taiwan)	Eardrum infrared thermography: 37.5	Frontal view DITI (max temp)	0.812	1	0.524	2.1	0
					Lateral view DITI (max temp)	0.801	1	0.483	1.93	0
			Thermoguard		Thermoguard	0.716	1	0.42	1.72	0
Chiue t al., 2005[1]	Taipei Medical University-Wan Fang Hospital (TMU-WFH), Taiwan	993 subjects	Telesis Spectrum 9000MB digital infrared thermal imaging [DITI] system	Eardrum temperature: 37.5	Face especially the frontal area		0.75	0.99	0.69	0.99

Hewlett et al., 2011[2]	Triage area of the emergency department at the University of Nebraska Medical Centre, Omaha, Nebraska, USA.	566 subjects (246 male) Age (Range -15 daysto 89 years; Mean-32 years)	ThermoScreen Infrared Fever Screening System (OptoTherm)	Oral/Perirectal/axillary temperature 37.8 C		0.862 (0.80-0.93)	0.70 (0.54-0.83)	0.92 (0.90-0.94)	0.42 (0.31-0.55)	0.97 (0.96-0.99)
				38.0 C		0.896 (0.83-0.96)	0.58	0.96	0.40	0.98
				38.3 C		0.945 (0.89-0.99)	0.60	0.97	0.43	0.98
McBride et al., 2010[3][4]	Cairns airport, north Queensland, Australia	181,759 passengers *	FLIR Thermoscan infrared camera	Ear temperature: 37.8					0.12	
		285 passengers †		Aural Temperature: 37.8						1 (0.98 - 1)
Ng et al., 2004[6]	Tan Tock Seng Hospital, Singapore Civil Defense Forces and Civil Aviation Authority,	502 subjects Out of 502, 310 included in regression	Handheld radiometric IR ThermaCAM S60 FLIR system (FLIR Systems, 2004)	Aural temperature: 37.7	Maximum temperature in the eye region ("Eye range max") [n=310]	0.972 (0.947-0.987)	0.85 (0.72-0.93)	0.95 (0.91–0.97)		

	Singapore	and ROC analysis.			Forehead ("Forehead range max") [n=310]	0.960 (0.932-0.979)	0.89 (0.77–0.96)	0.94 (0.90–0.96)		
Nguyen et al., 2010[19]	Emergency department of 3 urban tertiary-care hospital in the United States - Albuquerque, New Mexico; Atlanta, Georgia; and Chicago, Illinois	2873 subjects (≥18 yrs of age). [1514 men] Age Range 18-92 years (mean = 42 yrs)	FLIR ThermoVision A20M [FLIR Systems Inc., Boston, MA, USA] [n=2515]	Oral temperature : 37.8	Face/ Neck	0.92 (0.88-0.96)	0.90 (0.84-0.97)	0.80 (0.76-0.84)	0.18 (0.13-0.23)	0.995 (0.991-0.997)
			OptoThermThermoscreen [OptoTherm Thermal Imaging Systems and Infrared Cameras Inc., Sewickley, PA, USA] [n=2507]			0.96 (0.94-0.98)	0.91 (0.85-0.97)	0.86 (0.81-0.90)	0.18 (0.14-0.22)	0.996 (0.993-0.998)
			Wahl Fever Alert Imager HSI2000S [Wahl Instruments Inc., Asheville, NC,			0.78 (0.72-0.84)	0.80 (0.76-0.85)	0.65 (0.61-0.69)	0.057 (0.041-0.073)	0.991 (0.986-0.995)

			USA] [n=2061]							
Nishiura et al., 2011[3]	Passengers arriving at Narita International Airport, Japan	1049 passengers ‡ [Mean age: 30.3 yrs; 653 male]	TVS-500 infrared thermoscanners (NEC/AVIO Infrared Technologies Co. Ltd., Tokyo, Japan)	Axillary temperature 37.5		70.5 (67.7-73.2)	0.58 (0.54 – 0.62)	0.70 (0.66-.074)	0.68 (0.64-0.71)	0.61 (0.58-0.63)
				38.0		72.4 (69.6-75.0)	0.51 (0.45-0.55)	0.81 (0.78-0.84)	0.62 (0.57-0.66)	0.74 (0.71-0.76)
				38.5		73.1 (70.4-75.7)	0.70 (0.65 – 0.76)	0.63 (0.60 – 0.67)	0.37 (0.34-0.40)	0.87 (0.85-0.89)
Priest et al., 2011[4][1]	Airline travellers from Australian airports arriving at Christchurch airport, New Zealand	1275 subjects	Infrared thermal image scanner THERMACAM™ E45, FLIR Systems, Sweden	Tympanic temperature 37.5	Front of face	0.86 (0.75–0.97)	0.86	0.71	0.015	
				37.5	Side of face	0.76 (0.54–0.97)	0.86	0.51	0.01	
				37.8	Front of face	0.71 (0.62–0.81)	0.84	0.39	0.04	
				37.8	Side of face	0.67 (0.58–0.77)	0.84	0.31	0.036	

Sumriddetchkajornet al., 2009[42]	Triage section in Rajvithi hospital, Bangkok, Thailand	221 subjects	Thermo Vision A40-M	Aural temperature: 37.4	Max facial temperature		1	0.368		
Sunet al., 2014[32]	Patients with seasonal influenza at Self-Defense Forces Central Hospital, Japan	155 patients Mean age: 25 years.	Thermopile array (Chino Corp., Tokyo, Japan)	Axillary Temperature: 37.5	Face		0.80	0.93		
Suzuki et al., 2010[33]	Healthy volunteers in Tokyo, Japan	50 subjects (26 male)	NEC Avio Infrared Technologies Co., Ltd., TH5108 ME, Tokyo, Japan	Axillary temperature: 36.7	Face/ Head	0.65 (0.54-0.76)				
						0.57 (0.45-0.68)				
Tan et al., 2004[22]	Tan Tock Seng Hospital (TTSH), Singapore	46 patients	Infrared Fever Screening System (IFSS), ST Electronics	Core temperature: 38.0 C	Maximum facial temperature		1	0.83		

* Reference temperature checked only if index test (FLIR Thermoscan) was positive
† All 285 passengers were afebrile on the index test. Reference temperature measured to ensure that febrile patients were not being missed.
‡ Out of 9,140,435 passengers arriving at the airport, 1049 were grouped into the "selected and suspected fraction" consisting of 930 individuals detected by thermal scanners and rest by symptoms or history of exposure.

Table 3: Characteristics of included studies with both handheld non-contact infrared thermometers (NCITs) and thermal scanners were used as index test.

Author, Year of publication	**Setting**	**Sample characteristics**	**Index test device**	**Reference test cut off value (°C)**	**Site for index test**	**AUC**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**PPV(95 % CI)**	**NPV(95 % CI)**
Selent et al., 2013[20]	Pediatric emergency department, Georgia, USA	855 children , (469 male) Age: 6 mo- 17 years (with 27.8% between 3-5 years)	OptoTherm; OptoTherm Thermal Imaging Systems and Infrared Cameras Inc, Sewickley, Pa [n=854]	Rectal : 38 [n=218] Oral :38 [n=422] Axillary : 37 [n=215] (1°C added to axillary temperatures to approximate rectal/oral temperature)	Face & neck	0.922	0.83 (0.78-0.87)	0.86 (0.83-0.89)		
			FLIR; FLIR Systems Inc, Boston, Mass [n=852]		Face & neck	0.923	0.84 (0.79-0.88)	0.86 (0.82-0.88)		
			Thermofocus 0800H3; TECNIMED Srl, P. le Cocchi, Italy – hand held device [n=706]		Forehead	0.852	0.77 (0.71-0.82)	0.79 (0.75-0.83)		

Tayet al., 2015[2][3]	Singapore military personnel seeking medical care at a high volume primary healthcare centre, Singapore	430 military personnel (99.1% male) Mean age 19.1 yrs)	STE Infrared Fever Screening System (IFSS) (Singapore Technologies Electronics, Singapore)	Oral Temperature: 37.5	Face/ Neck		0.44 (0.39 -0.48)	0.99 (0.98 - 1)		
			Omnisense Sentry MKIII (Omnisense Systems Ptd Ltd, Singapore)				0.90 (0.87 - 0.93)	0.92 (0.89 - 0.94)		
			The handheld Quick Shot Infrared Thermoscope HT-F03B (Shenzhen WTYD Technology Limited, Guangdong, China)				0.29 (0.25 - 0.33)	0.96 (0.95 - 0.98)		