COVID-19 Evidence Update

COVID-19 Update from SAHMRI, Health Translation SA
and the Commission on Excellence and Innovation in Health

12 May 2020

Predictive value of temperature screening for COVID-19

Executive Summary

Key issues influence the value of widespread temperature screening in general settings (i.e. for staff and visitors at hospital entrances; airports; schools; other workplaces and community settings):

- How well does elevated temperature predict active cases of COVID-19?
- How good is temperature screening using available technology in detecting elevated temperature in population screening settings (vs in clinical diagnostic settings)?
- What are the practicalities of implementing temperature screening in setting; i.e. the feasibility and the costs of mass temperature screening in settings?
- The benefits of screening, and significantly, the likely prevalence of undetected cases of virus in the community[1].

Prevalence of fever as a symptom of COVID-19 in adults: COVID-19 has many known symptoms. The latest data from Australia indicated that the prevalence of **fever in confirmed cases was 47%**. The prevalence of **cough was higher (69%)**. [1]

Limitations of point prevalence assessment: Elevated temperature is not present in incubating (pre-symptomatic) cases. People can be infectious 1-3 days prior to developing symptoms. For symptomatic cases, the relationship between fever and infection is complex. Fever is not necessarily constant and infection may be present with a slightly elevated body temperature that does not meet the threshold for fever (38.0°C).

The prevalence of fever as a symptom of COVID-19 in children was 36% in a small pediatric study in a hospital setting in China. Data are not published on prevalence of fever as a symptom in all confirmed cases in Australia by age.

Available technology: Non-contact thermometers measure skin temperature rather than core body temperature and can be influenced by individual differences and environmental conditions. A range of non-contact thermometers have been evaluated for fever detection (pre-COVID-19) typically using contact thermometers as a reference, however, results vary greatly. In general, tympanic (ear cavity) thermometers have better results than forehead thermometers when compared to contact thermometers. **Low sensitivity and low PPV is commonly reported**. This suggests that there is considerable risk in missing febrile individuals and that, because the prevalence of fever is so low, there is increased cost with limited benefit.

[1] *Likely prevalence of undetected cases in SA and Australia is not covered here but is likely to be low at the time of writing*

Detecting fever in adults (any cause): A systematic review found the ability of non-contact infrared thermometers (NCIT) targeting the forehead area to detect elevated temperature varies enormously. The **sensitivity** varied from **4.0% to 89.6%,** the **specificity** from **75.4% to 99.6%**, the **PPV** from **0.9% to 76.0%** and the **NPV** from **86.1% to 99.7%0**. [2]

Detecting fever in children (any cause): A systematic review and meta-analysis looked at studies assessing the accuracy of infrared tympanic thermometry in the diagnosis of fever in children (<18 years). The pooled **sensitivity was 70%** and **pooled specificity was 86%**. The **pooled positive likelihood ratio was 9.14**, indicating that the febrile children had more than 9-fold higher chance of being diagnosed as fever by infrared tympanic thermometry, compared with afebrile children. [3]

Studies evaluating the effectiveness of screening in other diseases such as **SARS** and **influenza**, using non-contact mass temperature screening for fever, have shown that **very few cases are identified** and there is evidence of cases that were later identified but were missed during the screening process. Compulsory temperature checking in schools in Singapore during the 2003 SARS outbreak did not detect any cases.

Other considerations:

- Other limitations of temperature screening: fever can fluctuate throughout the course of a day; people may take medications to treat a fever; it is difficult to standardise or eradicate factors that influence the reliability of body temperature recordings.
- Other potential risks and benefits of screening (mostly commentary and limited empirical evidence):
    - may be **better than self-report** [4]
    - could serve as a **deterrent** if unwell people public stay at home to avoid screening [5]
    - may provide a **false sense of security** [5]
- Implications of screening: Decisions must be made about **what to do** with people with elevated temperatures in different settings i.e. turned away (e.g. from visiting, school or work); or secondary temperature screening; disease testing (border control); and conditions for return. All of these have associated costs for individuals, health systems and organisations.
- With any population screening tool, costs (and implications) of screening must be weighed against benefits gained from screening. For COVID-19, these may vary by settings (e.g. aged care, visitors to health care, schools, airports).

Summary: The benefits of widespread temperature screening are limited. Fever is a common symptom of COVD-19, but temperature screening has many limitations (in non-clinical settings) and may not be a better screener than other symptoms (e.g. cough). Technology available for rapid body temperature measurement also has many limitations. Widespread screening would be costly in terms of technology and staff resourcing, and the outcomes of 'positive' cases of high temperature must be managed. In most settings, the limitations of general screening for temperature are unlikely to be outweighed by the benefits.

Context and Definitions

- The value of temperature screening largely depends on the ability to detect true positive (TP) and true negative (TN) results given that false positive (FP) and false negative (FN) results may occur. A range of metrics are used to evaluate the diagnostic performance of tests*; the choice of metrics used to evaluate tests varies considerably across studies.
- Body temperature can be measured using a range of devices. Core body temperature is most accurately measured via contact thermometers (axillary (armpit), rectal, oral), however, these methods are considered invasive and poorly tolerated. Technological advances have resulted in more widespread availability of non-contact thermometers.
- **Non-contact infrared thermometers (NCIT):** detect infrared emission coming from the body, most commonly the forehead (blood is supplied by the temporal artery). Note: these devices measure skin temperature NOT core body temperature.
    - Range of products available, not all approved for use as thermometers, and performance in detecting fever ranges across devices.
    - Advantages: very quick, reduces discomfort and reduces close contacts with infected individuals (although handheld devices may need to be within 5cm of the patient).
    - Limitations: results can be influenced by individual factors (consumption of hot beverages or alcohol, pregnancy, menstrual period or hormonal treatments, intense perspiration or heavy face make-up), the body area that is targeted due to physiological differences in vascularisation and heat distribution (e.g. forehead compared to inner eye corner has more variable results but is more feasible for screening programs), and environmental factors (subject-sensor distance, ambient temperature or humidity and surrounding ventilation systems). [6]
- **Tympanic thermometers**: infrared ray to measure the temperature inside the ear canal. The tympanic membrane shares the same blood flow with the hypothalamus and are therefore considered to be the most accurate values. [7]
    - Advantages: quick, reduces discomfort
    - Limitations: not suitable for patients who have had ear surgery, requires the removal of hearing aid if present, earwax and otitis media can lead to incorrect values
- **Thermal scanner cameras**: produce a heat map image using infrared radiation. Note: these devices measure skin temperature NOT core body temperature.
    - Range of products available that range in cost and sophistication
    - Advantages: can measure temperature at a greater distance, can be used to screen a large number of people
    - Limitations: Not necessarily approved for fever detection, require controlled environment and are difficult to use effectively (compared to other devices).
- ***Diagnostic performance of tests**
  Sensitivity: ability of a test to correctly classify an individual as having a fever (=TP/(TP+FN). High sensitivity indicates low false negatives.
  Specificity: ability of a test to correctly classify an individual as not having a fever (=TN/FP+TN). High specificity indicates low false positives.
  Positive predictive value (PPV): percentage of individuals with a positive test who actually have a fever (=TP/TP+FP).
  Negative predictive value (NVP): percentage of individuals with a negative test who do not have a fever (=TN/FN+TN)].
  ROC curve: identifies the most appropriate cut-off given a test's sensitivity and specificity.

Summary of key evidence

Reviews

- A systematic review [2] was conducted on the sensitivity, specificity, and predictive values of **non-contact infrared thermometers (NCIT)** used with the objective of **fever screening**. **Compared NCIT with tympanic thermometer** results.
    - Six studies were included in the review; most took place in a health care setting but varied in terms of sample size, temperature threshold, target area for taking temperature, device used and environmental conditions.
    - The sensitivity, specificity and predictive values of NCIT targeting the forehead area largely differed between studies. The sensitivity varied from 4.0% to 89.6%, the specificity from 75.4% to 99.6%, the PPV from 0.9% to 76.0% and the NPV from 86.1% to 99.7%.

TABLE 2

Fever screening by non-contact infrared thermometers, 2004-2008: sensitivity, specificity and predictive values according to the body area targeted

First author, country, publication year	Sample size	Target area(s)	Temperature threshold	Fever prevalence %	Sensitivity %	Specificity %	PPV %	NPV %
Ng E Singapore 2004	310	Forehead	37.7°C	16.9	89.6	94.3	***76.0****	***97.8****
	310	Inner eye corner	37.7°C	16.9	85.4	95	***77.7****	***97.0****
Liu CC Taiwan 2004	500	Forehead	37.5°C	-	17.3	98.2		
	500	Auricular meatus	37.5°C	-	82.7	98.7		
Chan LS Hong Kong 2004	188	Forehead	38°C	14.3	4	99	***40.1****	***86.1****
	-	Forehead	37.5°C	Na	15	98		
	116	Auricular meatus	38°C	20.7	67	96	***81.4****	***91.8****
Ng DK Hong Kong 2005	500	Forehead	37.5°C †	12.3 †	89.4	75.4	33.7	98.1
Chiu W Taiwan 2005	993	Forehead	37.5°C	1.2	75	99.6	***69.9****	***99.7****
	72.327	Forehead	37.5°C	-	-	-	***0.9****	
Hausfater P France 2008	2.026	Forehead	38.0°C	3.0	82	77	10	99

* Values derived from the available information are in ***bold italic***
† The 37.5°C cut-off corresponds to the optimal sensitivity and specificity values reported by the authors whereas the prevalence (12.3%) is based on a 38°C threshold.
PPV: Positive predictive values; NPV: Negative predictive values

    - Fixing fever prevalence at 1% in all studies and using the reported sensitivity and specificity values, the derived PPV for the forehead area varied from 3.5% to 65.4% and the derived NVP was =>99%.

- A systematic review and meta-analysis [3] of studies assessing the **accuracy of infrared tympanic thermometry in the diagnosis of fever in children** (<18 years). Included studies were restricted to those using rectal measurement as the reference value as rectal measurement is considered gold standard in pediatric settings. The temperature cut-off was 38.0°C (or closest to this value when multiple values were used).
    - A total of 25 articles, encompassing 31 studies were included in the meta-analysis. The pooled sensitivity was 0.70 (95% CI =0.68-0.72) and pooled specificity was 0.86 (95% CI = 0.85-0.88). The pooled positive LR was 9.14 (95% CI =6.37-13.11), indicating that the febrile children had more than 9-fold higher chance of being diagnosed as fever by infrared tympanic thermometry, compared with afebrile children.

Table 2. Main Results of Studies Included in the Meta-Analysis.[a]

Study	Size	True Positive	False Positive	False Negative	True Negative	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	QUADAS Score
Allegaert et al[25]	294	53	0	188	54	0.22	1.00	1.00	0.94	12
Batra and Goyal[26]	100	40	0	10	50	0.80	1.00	1.00	0.83	11
Teller et al[1]	254	72	4	28	150	0.72	0.97	0.95	0.84	12
Abdulkadir and Johnson[2]	400	366	0	4	0	0.92	—	—	—	12
Hamilton et al[27]	205	69	4	26	106	0.73	0.96	0.95	0.80	10
Duru et al[28]	300	34	3	11	252	0.76	0.99	0.93	0.97	11
Edelu et al[29*]	45	17	2	6	20	0.73	0.91	0.89	0.77	12
Edelu et al[29*]	114	52	0	5	57	0.91	1.00	1.00	0.92	
Edelu et al[29*]	84	35	1	7	41	0.83	0.98	0.97	0.85	
Edelu et al[29*]	60	20	1	4	35	0.83	0.98	0.96	0.85	
Edelu et al[29*]	55	25	2	2	26	0.93	0.93	0.93	0.93	
Edelu et al[29*]	42	21	1	1	21	0.98	0.95	0.95	0.98	
Paes et al[30]	100	22	2	6	70	0.80	0.97	0.91	0.94	13
Vertedor-Hurtado et al[31]	119	50	17	0	52	1.00	0.75	0.74	1.00	10
Pavlović et al[32]	94	14	16	5	59	0.74	0.79	0.47	0.92	13
El-Radhi and Patel[33]	106	36	16	12	42	0.75	0.72	0.69	0.78	10
van Staaij et al[34]	40	14	2	1	23	0.93	0.92	0.88	0.96	10
Jean-Mary et al[35]	198	43	5	20	130	0.66	0.95	0.90	0.87	12
Banitalebi and Bangstad[36]	199	68	5	28	98	0.71	0.95	0.93	0.78	10
Loveys et al[37]	135	13	30	58	87	0.72	0.74	0.30	0.78	13
Wilshaw et al[38]	120	16	44	0	60	1.00	0.58	0.27	1.00	13
Lanham et al[39]	178	60	1	58	60	0.51	0.99	0.99	0.61	13
Montoya-Cabrera et al[40]	120	54	20	0	46	1.00	0.70	0.73	1.00	12
Hooker et al[41]	180	74	3	75	78	0.75	0.96	0.98	0.79	13
Brennan et al[42]	370	156	13	48	153	0.76	0.92	0.92	0.76	13
Selfridge and Shea[43]	102	16	9	2	75	0.88	0.89	0.74	0.97	13
Terndrup et al[44#]	303	59	10	35	199	0.63	0.95	0.85	0.83	12
Terndrup et al[44#]	303	79	46	24	155	0.77	0.77	0.63	0.86	
Chamberlain et al[45]	100	35	4	9	52	0.80	0.93	0.90	0.85	13
Kenney et al[46]	964	231	175	61	497	0.79	0.74	0.56	0.89	10
Rhoads and Grandner[47]	65	7	3	20	35	0.26	0.92	0.70	0.64	12

Abbreviation: QUADAS, Quality Assessment of Diagnostic Accuracy Studies.
[*#] and # indicate different studies in same article.

  - Contrary to the authors expectations, the results show that the accuracy of infrared tympanic thermometry in the diagnosis of pediatric fever is moderate, if not poor. It may be necessary to reduce the cut-off for fever diagnosis by 0.6°C to 0.2°C, which is the commonly reported difference between tympanic and rectal values.

- A **narrative review** [8] was conducted on the usefulness and applicability of infectious disease control measures in air travel. The authors indicated that **infrared thermal image scanners (ITIS)** used to identify febrile travellers are advantageous for being quick and contact free but have many disadvantages that make them less suitable than other strategies for disease control arising from air travel. They cited studies that showed that different scanners have different levels of accuracy, controlled environments are needed for accurate measurements, individual differences can influence the results, body location can influence results with forehead generally less accurate than ear temperature recordings, case detection depends on fever cut-off point which varies by study and device used, and the equipment and staff resourcing can be costly.

Primary studies

Temperature measurement to detect fever

Adults - Clinical settings

- A study [9] prospectively assessed accuracy of **cutaneous infrared thermometry** (non-contact, measures temperature on the **forehead**) for **detecting patients with fever in an emergency department**. Reference was tympanic temperature ≥38.0°C (measured in both ears) using infrared tympanic thermometer. On a sample of 2026 patients, sensitivity was lower than expected, PPV was low, correlation was poor, and body temperature was underestimated at low values and overestimated at high values. However, NPV was high. Variables correlated with the magnitude of difference between cutaneous and tympanic temperature measurements included: tympanic temperature, age, and outdoor temperature.
- A study [10] prospectively assessed the utility of temperature measurement using **infrared thermal detection systems (ITDS)** in an **emergency department** using routine oral temperature measurements as a reference. In a sample of 566 patients, the sensitivity, specificity, positive predictive value, and negative predictive value of the ITDS to detect temperatures of ≥38.0°C in all enrolled patients were 0.58, 0.96, 0.40, and 0.98, respectively, and for temperatures of ≥38.3°C were 0.60, 0.97, 0.43, and 0.98, respectively. The positive likelihood ratio for all subjects was 8.9. The prevalence of febrile patients was 7.6%. The authors concluded that low PPV results have high potential for false-positive results which may need to be followed up using alternative methods. The NPV values were very high, indicating that the ITDS has the ability to screen out individuals who are afebrile.
- A prospective study [11] assessing an inexpensive single channel **Cutaneous Infrared Thermometry (CIT)** in **detecting a fever in patients presenting to an emergency department**, relative to standard oral thermometry. The cut-off was defined as 38°C and the CIT was measured over the **forehead** at a distance of approximately 0.5m. Of the 548 cases, 67 (12.2%) had a fever when temperature was measured by an oral thermometer. Based on ROC analysis, detection of fever at a CIT reading of 35.3°C provided the best utility with a sensitivity of 0.236, specificity of 0.977, PPV of 0.589, NVP of 0.904 and accuracy of 0.888. Conclusion: at the optimum CIT reading, more than 50% of febrile patients could be incorrectly identified as afebrile.
- **3 ITDS were compared in 3 emergency departments** using oral temperature taken by clinical staff as the reference [4]. Patients also self-reported fever symptoms and medication taken for pain or fever. Confirmed fever was defined as ≥37.8°C. Of the sample of 2873, 64 (2.2%) had confirmed fever. Antipyretic or analgesic drug use within 8 hours was reported by 39% of patients. Compared with oral thermometry, sensitivity for self reported fever was 75%, specificity was 84.7%, and PPV was 10.1%. Sensitivities of the 3 ITDS at their respective optimal thresholds did not differ significantly from that of self-reported fever. However, specificities and PPVs of OptoTherm and FLIR at optimal thresholds were significantly greater than those of self-reported fever ($p<0.001$ for both comparisons), and specificity and PPV of Wahl were significantly lower than those of self-reported fever ($p<0.001$). NVP was above 99% for all four comparators.
    - Note author conclusions: “In our study, in which patients had no disincentive to report, we found that one fourth of febrile patients did not report having fever, which suggests true unawareness of fever among some persons. Only one tenth of those who reported having a fever were actually found to be febrile.” and
    - “In settings where secondary evaluation is available or during a pandemic with high illness severity, ITDS temperature can be set at a lower cutoff to ensure fewer false negatives, each of which represents a potential public health threat. However, setting the cutoff to achieve very high sensitivity can result in many false positives, which could have adverse consequences to the population being screened (e.g., unnecessary travel delays, missed work) and increase the workload of staff who are conducting the screening.”

Adults - Airport screening

- A Retrospective study [12] assessing the feasibility of **detecting influenza cases based on fever screening of airline travellers**. The diagnostic performance of **infrared thermoscanners** in detecting fever at cut-off levels of 37.5°C, 38.0°C and 38.5°C were also assessed (relative to axillary temperature. The sensitivity and specificity of the infrared thermoscanners in detecting hyperthermia ranged from 50.8-70.4% and 63.6-81.7%, respectively. The PPV value ranged from 37.3-68.0% and NPV ranged from 61.1-87.5%. The proportion of cases for each cut-off level was: 37.5°C =51.9%, 38.0°C =37.0% and 38.5°C =23.5%. The authors conclude that the sensitivity of fever for detecting influenza upon arrival was low and many of the confirmed cases were under antipyretic medications. The low PPV implies more false-positive passengers during mass screening when relying on infrared thermoscanners. Also, PPV of infrared thermoscanners was insufficient for actively detecting febrile passengers.
- A study [13] compared front of face **infrared thermal image scanner (ITIS)** to tympanic thermometer in measuring body temperature in 1275 **airline travellers** as an option for mass screening for influenza. Using 37.8°C as the cut-off and setting sensitivity at 86% gave specificity of 71%. The PPV in this population (of whom 0.5% were febrile) was 1.5%. Conclusion: while ITIS can have moderately high sensitivity and specificity for a high body core temperature ≥37.8°C, the low prevalence of fever in arriving travellers means that the PPV is very low. The low PPV of ITIS measures for fever suggests that false positives would be high - resulting in extra resources needed to conduct follow-ups. Moreover, the low PPV of ITIS detecting influenza infection means that the use of ITIS would have failed to identify all the influenza-infected travellers in this study.
- A study [14] measured **fever prevalence** detected using **infrared thermal scanning** at an international airport in **Cairns, Australia**, over a six-month period in 2006 (not outbreak specific). Passengers whose surface temperature was 1.3°C higher than the average surface temperature of other passengers were approached for measurement of body temperature using an ear thermometer. Of the 181,759 screened, 1052 (0.6%) had an elevated surface temperature. Of these, 963 agreed to have an ear temperature measurement and 118 were identified as febrile, equating to 0.06% of screened passengers. Using a subset of passengers (n=285; not identified through primary screening), 4 passengers had temperatures 37.5 to 37.6°C, giving an NVP of 98.6%. If 37.8°C was used as the cut-off, sensitivity would be 100%, PPV 12.3%, and NVP 100%.
- A study [15] assessed the performance of **airport screening procedures for detecting Dengue infection using fever as a screening tool** in Taiwan. International travellers were screened upon arrival using **thermal non-contact infrared thermometers (NCITs)**. Travellers with an NCIT detected temperature of higher than 37.5°C were detained and rechecked with a symptoms survey and ear thermometer. Travellers with a temperature above 38°C were defined as confirmed fever cases. A rapid dengue diagnostic system was used to confirm dengue infection. Fever prevalence ranged from 0.08-0.10% for inbound travellers. Overall, 44.9% (95%CI: 35.73-54.13%) of the confirmed imported dengue cases with apparent symptoms were detected by the thermal screening program with a PPV of 2.36% (95%CI: 0.96-3.75%), and NVP of >99.99% and a specificity of 99.97%. Sensitivity was 44.93%.
- A study [16] measured **fever prevalence** and the effectiveness of a fever screening procedure (**infrared thermal scanning camera**) in detecting febrile arrivals at an international airport in Korea during 2012 (not outbreak specific). Symptomatic arrivals were classified as passengers with fever above 36°C and those above 37.8°C were identified as febrile arrivals. Fever prevalence was 0.002%. Among febrile arrivals (n=6), the re-checked tympanic temperature (38.14°C) was similar to the first one (38.20°C), with the thermal temperature (36.83°C) slightly lower than the ear temperature; this difference was not statistically significant. Of 31 self-reported fever cases, 2 (6.5%) were confirmed as febrile; of 577 self-reported non-fever cases, 4 (0.7%) were identified as febrile (statistically significant association).
- A study [17] examined the effectiveness of **detecting influenza [not fever] airport screening** in NSW during the 2009 influenza pandemic. **Thermal imaging scanners** with a set point of 38°C ±2°C as well as self report health declaration cards were used to detect febrile passengers. Of 625,147 passenger arrivals, 5845 (0.93%) were identified as being symptomatic or febrile, and of these, 3 were subsequently confirmed as having influenza (case detection rate of 0.05 per 10,000). There were 45 people with

overseas acquired influenza who would have likely passed through the airport during this time, giving an airport screening sensitivity of 6.67% (95%CI, 1.4%-18.27%). PPV was 0.05% (95%CI, 0.02%-0.15%) and specificity was 99.10% (95%CI 99%-100%). Of 1296 passengers requiring further assessment, most (88.27%) were detected through health declaration cards and only 11 (0.85%) passengers were detected by the thermal scanners. It was more common to detect cases at emergency departments (52.1%) and general practices (24.2%) and at the airport (0.5%).

- A study [18] of 471,733 passengers who were screened on arrival in Japan, 9 individuals with **influenza [not fever]** were identified through entry screening and 1 was detected during quarantine (note: passengers were only tested for influenza if they had 2 or more of 4 symptoms). Of the 10 passengers, 2 had body temperatures above 38°C at entry screening. There were 24 individuals who were missed during entry screening despite being symptomatic but were identified as having influenza during the enhanced surveillance period. Of the 24 cases, 12 had a body temperature above 38°C at entry screening.

Adults - other settings

- A study [7] compared body temperature measurement values obtained with a **non-contact infrared thermometer, a tympanic thermometer and a chemical dot thermometer** in healthy young adults. The chemical dot and tympanic values were highly correlated whereas a weak correlation was found between the non-contact infrared and other thermometers. Furthermore, the difference between the measurements increases as the temperature increases

Children - clinical settings

- A study [19] of **36 children with confirmed COVID-19** admitted to hospital in China showed that **fever** (>37°C) **was a symptom in 13 cases (36%)**. Of the 13 cases, body temperature was 38.5°C or higher in four cases and between 37.5°C and 38.5°C in nine cases. The authors concluded, based on comparisons with other studies, that the prevalence of fever in Children with COVID-19 was much lower than adults with COVID-19 (86%) [20], children with SARS (100%) [21] and children with H1N1 influenza (92%) [22].
- A prospective study [23] **comparing non-contact infrared forehead temperature (NIFT) to tympanic thermometry** in children (n=567; 1 month to 18 years) admitted to **hospital**. The **optimal NIFT cut-off point** derived from this ROC curve for fever definition was **35.1°C**. The sensitivity, specificity, PPV and NPV of this cut-off point for fever screening were 89.4% (95% CI 83.1–93.6), 75.4% (95% CI 74.5–76.0), 33.7% (95% CI 31.4–35.3) and 98.1% (95% CI 96.9– 98.8), respectively. For this cut-off point, the positive likelihood ratio was 3.63 (95% CI 3.26–3.90) and 0.14 (95% CI 0.08– 0.23) for the negative test. Conclusion: screening fever by NIFT gave reasonable accuracy to rule out fever as the NVP was high but not to rule in fever as PPV was low. The authors concluded that:
    - "During outbreaks of SARS or similar communicable diseases, the main requirement of a fever-screening programme is a very low false-negative rate. NIFT for screening for fever is acceptable as the false-negative rate using this method was low, i.e. NPV=98%."
    - "We conclude that NIFT measurement has reasonable accuracy in detecting fever and could be recommended for fever screening in children during outbreaks of infectious diseases as the false-negative rate was very low. However, the high false positive rate of NIFT should be borne in mind."
- A study [24] compared body **temperature measurements of infrared tympanic and forehead non-contact thermometers with an axillary digital thermometer** in a sample of 50 pediatric patients in hospital. The **cut-off for fever was 38.0°C** measured using the axillary digital thermometer; 176 out of 1639 (10.7%) measurements had an axillary temperature reading of 38℃ or above. The comparison with tympanic measurements showed sensitivity of 89.8%, specificity of 92.2%, PPV of 68.2% and NVP of 97.8%. The comparison with forehead measurements showed sensitivity of 94.3%, specificity of 90.5%, PPV of 66.5% and NVP of 95.6%. Positive likelihood ratios were 17.9 and 16.5 and negative likelihood

ratios were 0.2 and 0.4 for the tympanic and forehead measurements, respectively. ROC curve analysis for determining the best threshold for axillary temperature above 38.0°C was **37.7°C for tympanic temperature and 37.5°C for forehead temperature**. The forehead temperature readings were found to have a higher level of bias than tympanic temperature readings.

- A study [25] compared accuracy and utility of **3 infrared (IFR) thermographs to axillary digital thermometer readings** in a group of 184 febrile and 135 afebrile children presenting to an emergency triage room. IFR skin scans were performed on the forehead, neck (over carotid artery) and nape. **Fever was defined as an axillary temperature of ≥37.5°C**. Correlations between axillary and IFR measurements were weak. The axillary recordings were consistently significantly higher than the forehead and nape but were similar to the neck readings. The predictive ability of forehead IFR data was reduced in children <6 years of age. **Overall, the use of forehead IFR data would mean that about 11.4% of febrile children would be missed.** When neck IFR temperatures were evaluated in children aged 2 to 6 years, the proportion of febrile children missed was 4.5%. In children aged >6 years, the febrile proportion missed using any IFR measure was 9.7%. Thus, neck IFR measures may be more accurate in children aged 2 to 6 years.
- A study [26] compared temperatures of 855 children (under 18 years old) presenting at a pediatric emergency department using **parent report, 3 infrared thermal detection systems (ITDSs) and age-appropriate contact temperature measurements**. **Confirmed fever was defined as ≥38.0°C (oral or rectal) and ≥37.0°C (axillary)**. Parents reported fever in 400 children (31% tactile, 39.5% measured temperature at home, remaining unknown). The PPV of parent-reported fever that was confirmed by traditional thermometry was 62.9% (tactile) and 64.6% (measured at home). Use of antipyretic medication was reported in 46% of children. Prevalence of confirmed fever was 35.8% when measured using traditional thermometry using a cut-off of 38.0°C. **The three ITDSs varied in optimal fever threshold (range 34.7-37.0**°C**)** with prevalence ranging from 38.4% to 40.1%. Sensitivity ranged from 0.7680 (Thermofocus ITDS [small, inexpensive, handheld]) to 0.8385 (parent report) and specificity ranged from 0.7084 (parent report) to 0.8634 (OptoTherm ITDS [large, expensive]). The false positive rate was highest for parents (0.3700) and the false negative rate was highest for the Thermofocus ITDS (0.1381). Conclusion: the two more expensive ITDS may be useful as an objective form of fever detection but should not be used as diagnostic tools.

Children - school settings

- On the basis that SARS patients were infectious only when febrile, twice daily monitoring of all schoolchildren aged 6 to 16 years was made mandatory in Singapore from 30 April to 25 July 2003 [27]. Students took their own temperatures and were not allowed to attend school if their temperature reading was ≥37.8°C (12 years and under) or ≥37.5°C (over 12 years). Some children with high temperatures were asymptomatic but were still turned away from school. **Asymptomatic children with persistent high readings and no contact/travel history with SARS cases were followed up (67 out of 499778 children (0.01%)**. Most (89.5%) had not identifiable pathological cause and were classified as **normal variants with high basal body temperatures**. None had SARS. The authors indicated that none of the children with SARS were detected through school temperature screening. However, temperature screening may have provided reassurance to parents that schools were safe during the SARS outbreak.

Other approaches to detect infectious disease beyond temperature measurement

- A portable system using multiple sensors was developed to detect infectious disease (i.e. influenza) based on vital signs (respiration rate, heart rate, and facial temperature). The system can classify patients into none, low-risk and high-risk influenza groups. Facial temperature was significantly different for the none and low-risk groups but there was no difference between the low and high risk groups [28]. A related study showed that conducting respiratory and heart rate screening along with facial temperature screening was more efficient and was possible using the non-contact machines [29]. Also, a test of efficacy showed that the system, when detecting influenza in a small sample (n=68) had a sensitivity of

91.7%. Specificity of 92.9%, NVP of 98.1% and PPV of 73.3% [30]. This corresponded to 1 false positive and 4 false negatives.

Commentary/correspondence regarding temperature measurement devices

- Letter ([31] J Hosp Infect) Suggests **tympanic infrared thermometer as a better measure of core body temperature than handheld infra-red thermometers** on the basis of the later having low sensitivity and its performance being operator dependent.

Effectiveness of screening for fever to detect infectious disease

Reviews

- A scoping and systematic review [5] examines evidence from 114 documents on entry and exit screening in the past 15 years. Screening measures have been used for a range of diseases (Ebola, SARS, influenza, Dengue fever) and in multiple countries. Very few cases were identified using exit and entry screening measures. The authors conclude that the primary objective of detecting imported cases of infectious disease at boards via entry screening has not been previously achieved. Screening protocols varied across studies but generally included questionnaires, temperature measurement (primary screening: non-contact thermometers; secondary screening: contact or minimal contact)
    - Limitations of screening - Cannot detect incubating or asymptomatic travellers, false declarations by passengers, antipyretic drugs conceal fever, non-specific questionnaires, language barriers, exit screening not dissuading ill travellers from attempting to return home.
    - Beneficial concomitant effects of screening - Obtaining passenger contact information for contract tracing purposes, informing travellers about risks and preventative measures, facilitating rapid and appropriate clinical care for ill travellers, maintaining air travel, may dissuade ill persons from travelling.
    - Adverse concomitant effects of screening - may give false sense of security, stigmatisation, high costs.
- A perspective piece [32] presents a narrative review of border-screening experiences during the 2003 SARS and 2009 H1N1 virus pandemics. The authors conclude that most previous examples of border screening during the influenza A (H1N1) and SARS infectious disease pandemics were ineffectual at preventing the spread of these infectious diseases. During the 2009 H1N1 pandemic, border screening protocols including the use of infrared thermal image scanners identified between 6.6-12.9% of infected persons with recent international travel in Japan, New South Wales and Singapore. Comparatively, Auckland, which did not use infrared thermal image scanners, identified 5.8% of infected persons. The authors suggest that diseases where fever is a more consistent symptom and where persons are not infectious when asymptomatic or during the incubation period are more amenable to border screening.

Modelling to detect COVID-19 using temperature screening

- A simulation [33] of 100 SARS-CoV-2 infected travellers estimated that, under conservative assumptions on sensitivity (86%) of infrared thermal image scanners, 46% of infected travellers would enter a country with the infection undetected by airport entry/exit screening. The authors conclude that the prevention of infected travellers entering countries via air is only achievable if there is negligible asymptomatic infection, sensitivity of screening is almost perfect and the incubation period is short.
- A modelling study [34] estimated that COVID-19 screening (symptom screening via infrared thermal scanners and exposure risk self-report questionnaire) will detect less than half of infected travellers during a growing epidemic. However, screening effectiveness will marginally increase as growth of the

epidemic decelerates. The greatest contributor to case detection is the departure fever screen, followed by the arrival fever screen. The two main factors influencing the effectiveness of screening were the difficulty in detecting infected individuals during incubation period (which is amplified with longer incubation periods) or during early onset of symptoms, and limitations with sensitivity of case detection by exposure risk questionnaire.

Selected primary studies (see [5] for comprehensive list)

- A study [35] used daily COVID-19 incidence data and global airport network connectivity from mainland China to estimate country level exportation risks of the outbreak and to estimate the impact of control measures. The modelled estimates suggest that 64% of exported COVID-19 cases were in the pre-symptomatic incubation period upon arrival at their destination airport. Travel restrictions reduced the rate of disease exportation by 81% from Wuhan and 71% from Hubei by 15 February 2020 compared to no border restrictions.
- A study [36] aimed to assess the effectiveness of the Australian border entry screening program to detect arriving travellers with symptoms of SARS between 5 April 2003 and 16 June 2003. Of the 241,491 travellers arriving from countries with local transmission of SARS, only 19 were deemed probable cases of SARS via the border screening protocol; none of whom were later confirmed cases. The border screening protocol missed 25 passengers who were symptomatic on arrival and deemed probable cases of SARS via the Australian SARS Case Register, although none of whom were later confirmed cases.
- During 2007–2012, 1.9% of tested symptomatic passengers (n=30,000) were positive for dengue fever; 46% of cases were recognised at the airport border [37].

Commentary/correspondence

- Correspondence ([38] NEJM) In an evacuation of 126 people from Wuhan to Frankfurt, a symptom-based screening process (including temperature screen) was ineffective in detecting **SARS-CoV-2** infection in 2 positive passengers. The results suggest that asymptomatic passengers may be infectious while travelling.
- Correspondence ([39] MJA) In Singapore, to reduce transmission of **SARS-CoV-2**, temperature screening was implemented at the border, encouraged in the workplace, implemented in most schools and businesses (which at the time had remained open), and mandated for all front-line health care workers in all public hospitals (measured twice daily). As of 15 March, none of the 569 health care workers who had been evaluated had tested positive.
- Letter ([40] J Hosp Infect) Almost all Taiwanese hospitals have established temperature monitoring at outdoor quarantine stations to detect fever associated with **COVID-19**. A community hospital in Taiwan reported 5 out of 40,887 patients attending the hospital in March presented with fever (>38°C) prior to hospital entry. However, a further 37 patients were identified with fever after a second temperature recording was made after acclimatizing to being indoors within the hospital. Authors conclude that repeated temperature measurements are needed.
- A narrative opinion article [41] discusses the similarities and differences of **SARS and COVID-19**. COVID-19 has a higher transmissibility than SARS, and many more patients with COVID-19 rather than SARS have mild symptoms that contribute to spread because these patients are often missed and not isolated. Pre-symptomatic and asymptomatic transmission of COVID-19 make temperature screening less effective for COVID-19 than SARS.
- Letter to the editor ([42] Tropical medicine and health). Authors argued that **screening for fever is likely to miss cases** and other strategies such as PCR testing and quarantine are better for controlling transmission. However, the authors did point out that screening for fever at airports could help with partial blocking of transmission if some cases were identified and quarantined.

- News article ([43] Science) Suggests that thermal scanners and handheld thermometers which measure skin temperature rather than core body temperature, the key metric for fever, often produce false positives and false negatives. The report suggests screening protocols in the US and China have been ineffective at picking up cases of COVID-19 among inbound travellers.

Authors: Prof Caroline Miller, Dr Jacqueline Bowden, Jo Dono (SAHMRI)

Searchers: Nikki May & Dr Ingrid Lensink (Commission on Excellence and Innovation in Health)

Expert input: Prof Steve Wesselingh

Acknowledgement: Aimee Brownbill

References

1. Covid-19 National Incident Room Surveillance Team, *COVID-19, Australia: Epidemiology Report 13 (Reporting week to 23:59 AEST 26 April 2020).* Commun Dis Intell, 2020. **44**. 10.33321/cdi.2020.44.35
2. Bitar, D., A. Goubar, and J.C. Desenclos, *International travels and fever screening during epidemics: a literature review on the effectiveness and potential use of non-contact infrared thermometers.* Eurosurveillance, 2009. **14**(6): p. 19115. doi:https://doi.org/10.2807/ese.14.06.19115-en
3. Zhen, C., et al., *Accuracy of infrared tympanic thermometry used in the diagnosis of Fever in children: a systematic review and meta-analysis.* Clin Pediatr (Phila), 2015. **54**(2): p. 114-26. 10.1177/0009922814545492
4. Nguyen, A.V., et al., *Comparison of 3 infrared thermal detection systems and self-report for mass fever screening.* Emerg Infect Dis, 2010. **16**(11): p. 1710-7. 10.3201/eid1611.100703
5. Mouchtouri, V.A., et al., *Exit and Entry Screening Practices for Infectious Diseases among Travelers at Points of Entry: Looking for Evidence on Public Health Impact.* Int J Environ Res Public Health, 2019. **16**(23). 10.3390/ijerph16234638
6. Ng, E.Y., G.J. Kaw, and W.M. Chang, *Analysis of IR thermal imager for mass blind fever screening.* Microvasc Res, 2004. **68**(2): p. 104-9. 10.1016/j.mvr.2004.05.003
7. Basak, T., et al., *Comparison of three different thermometers in evaluating the body temperature of healthy young adult individuals.* Int J Nurs Pract, 2013. **19**(5): p. 471-8. 10.1111/ijn.12097
8. Huizer, Y.L., et al., *Usefulness and applicability of infectious disease control measures in air travel: a review.* Travel Med Infect Dis, 2015. **13**(1): p. 19-30. 10.1016/j.tmaid.2014.11.008
9. Hausfater, P., et al., *Cutaneous infrared thermometry for detecting febrile patients.* Emerg Infect Dis, 2008. **14**(8): p. 1255-8. 10.3201/eid1408.080059
10. Hewlett, A.L., et al., *Evaluation of an infrared thermal detection system for fever recognition during the H1N1 influenza pandemic.* Infect Control Hosp Epidemiol, 2011. **32**(5): p. 504-6. 10.1086/659404
11. Hogan, D.E., S. Shipman, and K. Smith, *Simple infrared thermometry in fever detection: consideration in mass fever screening.* Am J Disaster Med, 2015. **10**(1): p. 69-74. 10.5055/ajdm.2015.0190
12. Nishiura, H. and K. Kamiya, *Fever screening during the influenza (H1N1-2009) pandemic at Narita International Airport, Japan.* BMC infectious diseases, 2011. **11**(1): p. 111.
13. Priest, P.C., et al., *Thermal image scanning for influenza border screening: results of an airport screening study.* PLoS One, 2011. **6**(1): p. e14490. 10.1371/journal.pone.0014490
14. McBride, W.J., E. Buikstra, and M. FitzGerald, *Investigation of febrile passengers detected by infrared thermal scanning at an international airport.* Aust N Z J Public Health, 2010. **34**(1): p. 5-10. 10.1111/j.1753-6405.2010.00466.x
15. Kuan, M.M. and F.Y. Chang, *Airport sentinel surveillance and entry quarantine for dengue infections following a fever screening program in Taiwan.* BMC Infect Dis, 2012. **12**(1): p. 182. 10.1186/1471-2334-12-182
16. Cho, K.S. and J. Yoon, *Fever screening and detection of febrile arrivals at an international airport in Korea: association among self-reported fever, infrared thermal camera scanning, and tympanic temperature.* Epidemiology & Health, 2014. **36**.
17. Gunaratnam, P.J., et al., *Airport arrivals screening during pandemic (H1N1) 2009 influenza in New South Wales, Australia.* Med J Aust, 2014. **200**(5): p. 290-2. 10.5694/mja13.10832
18. Sakaguchi, H., et al., *Assessment of border control measures and community containment measures used in Japan during the early stages of pandemic (H1N1) 2009.* PloS one, 2012. **7**(2).
19. Qiu, H., et al., *Clinical and epidemiological features of 36 children with coronavirus disease 2019 (COVID-19) in Zhejiang, China: an observational cohort study.* Lancet Infect Dis, 2020. 10.1016/S1473-3099(20)30198-5
20. chen, d., et al., *Hypokalemia and Clinical Implications in Patients with Coronavirus Disease 2019 (COVID-19).* medRxiv, 2020: p. 2020.02.27.20028530. 10.1101/2020.02.27.20028530
21. Leung, C.W., et al., *Severe acute respiratory syndrome among children.* Pediatrics, 2004. **113**(6): p. e535-43. 10.1542/peds.113.6.e535
22. Wang, Z., et al., *Clinical features of 167 children with the novel influenza A (H1N1) virus infection in Xi'an, China.* Turk J Pediatr, 2012. **54**(2): p. 99-104.
23. Ng, D.K., et al., *Non-contact infrared thermometry temperature measurement for screening fever in children.* Ann Trop Paediatr, 2005. **25**(4): p. 267-75. 10.1179/146532805X72412
24. Apa, H., 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**(9): p. 992-7. 10.1097/PEC.0b013e3182a2d419
25. Berksoy, E.A., et al., *Use of noncontact infrared thermography to measure temperature in children in a triage room.* Medicine, 2018. **97**(5).
26. Selent, M.U., et al., *Mass screening for fever in children: a comparison of 3 infrared thermal detection systems.* Pediatr Emerg Care, 2013. **29**(3): p. 305-13. 10.1097/PEC.0b013e3182854465
27. Chng, S.Y., et al., *Mandatory temperature monitoring in schools during SARS.* Arch Dis Child, 2004. **89**(8): p. 738-9. 10.1136/adc.2003.047084
28. Sun, G., et al., *An infectious disease/fever screening radar system which stratifies higher-risk patients within ten seconds using a neural network and the fuzzy grouping method.* J Infect, 2015. **70**(3): p. 230-6. 10.1016/j.jinf.2014.12.007
29. Nakayama, Y., et al. *Non-contact measurement of respiratory and heart rates using a CMOS camera-equipped infrared camera for prompt infection screening at airport quarantine stations*. in *2015 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*. 2015. IEEE.

30. Sun, G., et al., *A portable infection screening system designed for onboard entry screening based on multi-parameter vital signs.* International Journal of E-Health & Medical Communications, 2013. **4**(3): p. 20-35.
31. Aw, J., *The non-contact handheld cutaneous infra-red thermometer for fever screening during the COVID-19 global emergency.* J Hosp Infect, 2020. **104**(4): p. 451. 10.1016/j.jhin.2020.02.010
32. Selvey, L.A., C. Antao, and R. Hall, *Entry screening for infectious diseases in humans.* Emerg Infect Dis, 2015. **21**(2): p. 197-201. 10.3201/eid2102.131610
33. Quilty, B.J., et al., *Effectiveness of airport screening at detecting travellers infected with novel coronavirus (2019-nCoV).* Euro Surveill, 2020. **25**(5). 10.2807/1560-7917.ES.2020.25.5.2000080
34. Gostic, K., et al., *Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19.* Elife, 2020. **9**: p. e55570. 10.7554/eLife.55570
35. Wells, C.R., et al., *Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak.* Proc Natl Acad Sci U S A, 2020. **117**(13): p. 7504-7509. 10.1073/pnas.2002616117
36. Samaan, G., et al., *Border screening for SARS in Australia: what has been learnt?* Medical Journal of Australia, 2004. **180**(5): p. 220-223. 10.5694/j.1326-5377.2004.tb05889.x
37. Su, C.P., et al., *Clinical and epidemiological characteristics of imported dengue fever among inbound passengers: Infrared thermometer-based active surveillance at an international airport.* PLoS One, 2019. **14**(12): p. e0225840. 10.1371/journal.pone.0225840
38. Hoehl, S., et al., *Evidence of SARS-CoV-2 Infection in Returning Travelers from Wuhan, China.* N Engl J Med, 2020. **382**(13): p. 1278-1280. 10.1056/NEJMc2001899
39. Lin, R.J., T.H. Lee, and D.C.B. Lye, *From SARS to COVID-19: the Singapore journey.* The Medical Journal of Australia, 2020.
40. Hsiao, S.H., et al., *Body Temperature Measurement to Prevent Pandemic COVID-19 in Hospitals in Taiwan: Repeated Measurement is Necessary.* J Hosp Infect, 2020. 10.1016/j.jhin.2020.04.004
41. Wilder-Smith, A., C.J. Chiew, and V.J. Lee, *Can we contain the COVID-19 outbreak with the same measures as for SARS?* The Lancet Infectious Diseases, 2020. **20**(5): p. e102-e107. 10.1016/s1473-3099(20)30129-8
42. Bwire, G.M. and L.S. Paulo, *Coronavirus disease-2019: is fever an adequate screening for the returning travelers?* Trop Med Health, 2020. **48**: p. 14. 10.1186/s41182-020-00201-2
43. Normile, D., *Airport screening is largely futile, research shows.* Science, 2020. **367**(6483): p. 1177-1178. 10.1126/science.367.6483.1177