

Accuracy and acceptability of non-contact infrared compared to axillary and tympanic thermometers in children under 6 attending primary care: a method comparison study and nested qualitative evaluation

Authors: Gail Hayward¹(0000-0003-0852-627X), Jan Y Verbakel¹², Fatene Abakar Ismail³, George Edwards¹(0000-0002-0048-4178), Kay Wang¹, Susannah Fleming¹, Gea A Holtman¹⁴, Margaret Glogowska¹, Elizabeth Morris¹, Kathryn Curtis⁵, Ann van den Bruel²

1: Nuffield Department of Primary Care Health Sciences, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG.

Dr Gail Hayward (Associate Professor), Dr Jan Y Verbakel (Honorary Clinical Researcher), Mr George Edwards (Research assistant), Dr Kay Wang (NIHR Postdoctoral Fellow), Dr Susannah Fleming (Quantitative Researcher), Dr Gea Holtman (Academic visitor), Dr Margaret Glogowska (Senior qualitative researcher), and Dr Elizabeth Morris (NIHR in-practice fellow).

2: Academic Centre for Primary Care, Department of Public Health and Primary Care, KULeuven, Kapucijnenvoer 33 J, 3000 Leuven, Belgium

Dr Jan Y Verbakel (Assistant professor), Dr Ann van den Bruel (Associate Professor)

3: Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow Royal Infirmary, New Lister Building, 10 Alexandra Parade , Glasgow ,G31 2ER.

Dr Fatene Abakar Ismail (Research Associate)

4: Department of General Practice and Elderly Care Medicine, University Medical Centre Groningen, University of Groningen, PO Box 196, 9700 AD Groningen, the Netherlands

Dr Gea Holtman (Postdoctoral researcher)

5: Office G16, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS

Mrs Kathryn Curtis (Trial Manager)

Corresponding author: Gail Hayward. gail.hayward@phc.ox.ac.uk

ABSTRACT

Background: Guidelines recommend the measurement of temperature in children presenting with fever symptoms using electronic axillary, or tympanic thermometers. Non-contact thermometry offers advantages yet have not been tested against the recommended methods in primary care.

Aim: To compare two different non-contact infrared thermometers (NCITs) to axillary and tympanic thermometers in children aged 5 or under attending their GP with an acute illness.

Design and setting: Prospective design in primary care

Method: Methods comparison study with nested qualitative component

Results: 401 children were recruited (median age 1.6 years, 50.62% boys). The mean difference between the Thermofocus NCIT and axillary thermometer was -0.14°C (95% CI -0.21 to -0.06); the lower limit of agreement was -1.57°C (95% CI -1.69 to -1.44) and the upper limit 1.29°C (95% CI 1.16 to 1.42). A second NCIT (Firhealth) had similar levels of agreement. However, the limits of agreement between tympanic and axillary thermometers were also wide. Parents expressed a preference for the practicality and comfort afforded by NCITs, and were predominantly negative about their child's experience of axillary thermometers. However, there was a willingness to adopt whichever device was medically recommended.

Conclusions In a primary care paediatric population, temperature measurements with NCITs varied by over a degree Celsius compared to axillary and tympanic approaches. However, there was also poor agreement between tympanic and axillary thermometers. Since clinical guidelines often rely on specific fever thresholds, clinicians should interpret peripheral thermometer readings with caution, and in the context of an holistic assessment of the child.

Trial registration ISRCTN15413321

How this fits in:

Non-contact infrared thermometers (NCITs), allow clinicians to measure temperature from the forehead without touching the child, reducing discomfort and distress and the need for disposable covers. However NCITs have not been tested against the currently recommended methods in primary care.

We compared two different NCITs to electronic axillary and tympanic thermometers in children aged 5 and under seeing their GP with an acute illness. Our analysis suggests that in 95 out of 100 cases we would expect the difference between the NCIT and electronic axillary thermometer or tympanic thermometer to be up to a degree Celsius in either direction.

Since we also found poor agreement between axillary and tympanic readings, it is hard to know which type of thermometer is giving results closest to core body temperature. Clinicians should be aware of the variability in peripheral thermometer readings when assessing febrile children according to clinical guidelines.

INTRODUCTION

Acute infections in children are one of the most common problems in general practice and are associated with considerable burden on NHS resources. Nearly 40% of parents with children aged 6 to 17 months consult a healthcare professional when their child has a high temperature.(1) In the UK, acute infections result in 4 consultations per person-year in children under 1 year old, and 1.3 consultations per person-year in children aged 1-15 years.(2) Febrile illness accounts for 20% of all visits to the paediatric emergency department.(3)

Guidelines recommend the measurement of temperature in children presenting with fever symptoms using electronic axillary thermometers, or tympanic thermometers in children aged 4 weeks and older.(4) However, axillary thermometers require healthcare professionals to undress the child and hold the thermometer in the axilla for at least 30 seconds.(5) Tympanic thermometers are easier to use, but may be inaccurate due to presence of ear wax or insufficient straightening of the ear canal,(6) and both types of device require disposable covers to avoid cross- infection. Non-contact infrared thermometers (NCITs) convert measurements of the intensity of infrared radiation emitted by the body into temperature readings. The non-contact approach offers potential advantages, including reduced child discomfort or distress, rapid readings, measurement without interrupting sleep, minimal risk of cross-infection, and no requirement for disposable covers.(7)

Reports of agreement between NCITs and conventional thermometers have been variable, with larger mean differences reported between NCITs and tympanic thermometers than with NCITs and mercury-in-glass axillary thermometers.(8)(9) Comparisons with electronic and mercury rectal thermometry have also yielded variable results,(10)(11) and performance varies between devices.(12) Finally, whilst NCITs are mostly reported to have high sensitivity and specificity in detecting a fever of 38°C or higher measured with conventional thermometry,(8–10,13) in two studies sensitivity was estimated as only 27%(11) and 12%.(12)

In addition to the lack of clear conclusions from existing studies, there is a lack of generalisability of this data to primary care settings. Most previous studies were conducted in paediatric inpatient populations(9,11,12) or mixed hospital ambulatory care and ward settings.(10)(14) Furthermore NCITs have been mainly compared with temperature measurement approaches which are not

currently recommended for use in children, including rectal measurements(10–12,15) and using mercury-in-glass axillary thermometers.(8)

Understanding the performance of NCITs compared to currently recommended tympanic and electronic axillary thermometers in a primary care paediatric population could support introduction of this potentially beneficial technology into routine practice. In this mixed methods study, we evaluated the agreement between two NCIT models with electronic axillary and tympanic thermometers in children who present with acute illness in primary care and explored their acceptability to parents and children.

METHODS

This was a cross-sectional method agreement study with a nested qualitative study.

Method comparison

Children between 0 and 5 years with an acute illness of a maximum of 14 days presenting to a GP practice (9 sites) or an out-of-hours (OOH) service (1 site) in the UK were eligible for inclusion.

Children for whom acute trauma was the main reason for presentation, who were clinically unstable, had already been enrolled in the study or where parents were unable to understand trial materials in English were excluded from the study.

Parents and children were approached consecutively in the OOH or surgery waiting room by a study researcher between April 2017 and August 2018. Parents were provided with an information sheet and gave verbal consent to their child's involvement, with the option to consent to further contact by telephone for involvement in the qualitative substudy, and the option to withdraw consent subsequently by email or telephone. Temperature measurements were conducted either prior to or after the child's GP appointment. Demographic information and the history of fever were recorded.

Four thermometers were compared: Electronic axillary (Welch Allyn SureTemp), tympanic (Braun Thermoscan), NCIT Thermofocus 0800 and NCIT Firhealth Forehead Thermometer. The Thermofocus NCIT was included as it had been most extensively evaluated in other settings.(8, 10, 12,15,16) The Firhealth device was included as an example of a cheaper NCIT.

Measurements were performed consecutively in the shortest time frame possible, and no medication or drinks were administered between measurements. The order in which the thermometers were used was randomised prior to study start for each participant using a Random Number Generator (www.random.org). Tympanic measurements were not taken in babies under 4 weeks in line with UK guidance.(4) Once the four primary measurements were complete, a second measurement was taken with the Thermofocus and Firhealth thermometersto evaluate their reproducibility. We recorded failed measurements due to lack of cooperation of the child after three attempts, mechanical issues (operational or technological failure) and clinically implausible readings (based on researcher's assessment).

Children's reaction to the different measurements was rated by parents using the Patient Discomfort Scale.(17) Children aged 4 or 5 additionally completed the Wong-Baker Faces® Pain Rating Scale.(18) Parents scored the acceptability of each thermometer on a 10cm visual analogue scale before they were informed of the temperature measurements.

Sample Size

The sample size calculation was based on the desired accuracy of the limits of agreement on the Bland-Altman plots for our primary outcome.(19) The limits of agreement are defined as the mean difference (bias) ± 1.96 standard deviations(20). If the differences are normally distributed, approximately 95% of them may be expected to lie within the limits of agreement, so they are a useful estimate of the typical range of differences between the two measurements. Assuming an accuracy of $\pm 0.075^{\circ}\text{C}$ would be desired for the 95% confidence intervals of the limits of agreement, and the standard deviation of the agreement between temperatures measured by NCIT and electronic axillary thermometers (based on previous studies) (8–10, 13,15) would be 0.5°C , a minimum sample size of 533 participants would have been required. We revised this sample size calculation on 14th May 2018. Based on the data already available, the standard deviation for our primary analysis (Thermofocus NCIT vs electronic axillary thermometer) was 0.65°C . Using this standard deviation, the original sample size of 533 children would give us a 0.10°C accuracy for each limit of agreement (rather than 0.075 as anticipated). A reduced sample size of 400 children would give us 0.11°C accuracy. Considering thermometers only measure temperatures up to 0.1°C , we felt this would be sufficient as the rounding would make the two estimates equivalent. Secondary outcomes of agreement between the other thermometer types have been estimated with the same precision.

Analyses

Statistical methods focus on the agreement between thermometers, the accuracy of detecting fever and failure rates. All children contributed data to each analysis, when available.

Our primary outcome is the agreement between the Thermofocus NCIT and the electronic axillary thermometer. Analyses of agreement were conducted based on Bland Altman plots(19) which provide an indication of bias and 95% limits of agreement between the measurements. 95% confidence intervals (CI) around these estimates have been calculated.

Diagnostic accuracy for detecting fever (temperature $\geq 38^{\circ}\text{C}$ measured by the electronic axillary thermometer) was analysed by calculating sensitivity, specificity, predictive values, and likelihood ratios, with 95% confidence intervals. Failure rates are reported as proportions.

The scores on the Visual Analogue and Patient Discomfort Scales have been analysed using non-parametric techniques resulting in median acceptability (and interquartile ranges) for each thermometer.

Qualitative data collection and analysis

Parents who consented to contact were purposively sampled to achieve maximum variation in gender of parent, age of parent, age of child, ethnicity, and sibling number. Recruitment continued until the research team agreed data saturation had been achieved and sufficient explanation for the categories generated was reached.

Interviews were semi-structured and conducted by telephone (20) or face-to-face (1) using a flexible topic guide developed by the research team and PPI panel, which evolved in response to emerging themes. The topic guide explored parent's experiences with the different thermometers, their thoughts about future use of thermometers, and wider exploratory questions about motivations for and experience of temperature measurement and fever in children. A detailed example of the topic guide is presented in Supplementary materials.

All participants gave written or recorded verbal informed consent prior to the interview. Interviews were conducted separately by two researchers trained in qualitative methodology (EM, a female clinical researcher and salaried GP, and FAI, a female research assistant), were audio-recorded and transcribed. Consistency was ensured by regular discussion and review of transcripts and topic guides by the team (EM, FAI, GH, MG).

Data analysis followed a thematic approach, with the assistance of NVivo (version 11). This included familiarisation with the data, open coding, and subsequent inductive reasoning to identify salient categories and relationships between emerging themes derived from the data. Data and codes were then checked by two researchers (EM, MG). The codes and themes were developed and interpreted in discussion with the wider research team.

Patient and Public Involvement

The PPI panel for the NIHR Community Healthcare MedTech and IVD Cooperative which includes two mothers of children under 5, supported this project from inception to dissemination. The PPI panel have provided feedback on the study design, suggesting that we include reproducibility as a secondary outcome and endorsing the inclusion of a cheaper NCIT as a more affordable option for

home use. They also provided feedback on study materials, particularly the patient information leaflets and acceptability ratings by children and parents, commented on the topic guide for the qualitative interviews with the parents and provided feedback on our sample size revision. We sought their advice on acceptable limits of agreement between thermometry devices, and discussed our emerging themes on the acceptability of the thermometer types from our qualitative interviews. We are also being advised by them regarding dissemination of our findings to a lay audience.

RESULTS

We recruited 401 children with a median age of 1.6 years (IQR 0.79-3.38); 203 (50.62%) were boys. Five children were under 4 weeks old. Most children were of White British ethnicity (69.83%). 29.7% of children were believed by their parents to be feverish at the time of inclusion. Participant characteristics are listed in Table 1.

The Bland Altman plot for our primary outcome, which is the agreement between the Thermofocus NCIT (1st measurement) and the electronic axillary thermometer, is presented in Figure 1. The mean difference between the two methods was -0.14°C (95% CI -0.21 to -0.06), with the lower limit of agreement being -1.57°C (95% CI -1.69 to -1.44) and the upper limit 1.29°C (95% CI 1.16 to 1.42). This means that in 95 out of 100 cases we would expect the difference between the NCIT and electronic axillary thermometer to be between 1.57°C lower and 1.29°C higher than the average of the NCIT and electronic axillary measurements of temperature.

The mean difference between the Firhealth NCIT and the electronic axillary thermometer was -0.16°C (95% CI -0.23 to -0.09); the lower limit of agreement was -1.54°C (95% CI -1.66 to -1.41) and the upper limit 1.22°C (95% CI 1.10 to 1.34).

Full method comparison results, including NCIT versus tympanic and repeated measurements with both NCITS are summarized in Table 2 below. Accompanying plots are available as supplementary figure 1. Of note, there were wide limits of agreement when the tympanic and electronic axillary methods were compared. The mean difference was 0.06 (95%CI -0.02 to 0.14), lower limit of agreement -1.49 (95% CI -1.63 to -1.34) and upper limit of agreement 1.61 (95% CI 1.47 to 1.75). The reproducibility of the NCITs was reasonable; for the Thermofocus the mean difference was -0.04 (-0.07 to -0.01), the lower limit of agreement was -0.56 (-0.60 to -0.51) and the upper limit of agreement was 0.47 (0.43 to 0.52), meaning that the two measurements varied by less than half a degree Celsius in most cases. Firhealth reproducibility was similarly acceptable. Although there is some indication in the Bland-Altman plots of temperature-related relationships in the difference between measurements made by NCITs and axillary thermometry (Figures 1 and 2), we have not explored this formally as they are dominated by a small number of measurements at extremes of the physiological temperature range, where it is expected that thermometers for physiological use will be less accurate.

We calculated accuracy of NCITs for diagnosing fever, defined as a temperature of at least 38°C by electronic axillary measurement. Based on the axillary thermometer, there were 41 children with a body temperature of 38°C or more, a prevalence of fever of 10.22%. The proportion of children detected as febrile by the Thermofocus NCIT out of all children found to be febrile using the

electronic axillary thermometer (i.e. the 'sensitivity' for this alternate imperfect reference standard) was 29.3% (95% CI 16.1 to 45.5), and the proportion found not to be febrile out of all the children classified as not febrile by electronic axillary measurement (specificity) was 96.9% (95% CI 94.6 to 98.5). For the Firhealth NCIT, sensitivity was 4.9% (95% CI 0.6 to 16.5) and specificity 98.9% (95% CI 97.2 to 99.7). See table 3 for further details.

The number of attempts for each thermometer that were required to obtain a valid reading and the technical failures are detailed in Table 4.

Thermometer acceptability

Children (n = 69) who were aged 4 or 5 completed the Wong-Baker Faces® Pain Rating Scale. The median score and IQR was 0 (0-0) for both NCITs, and 0 (IQR 0-2) for the electronic axillary and tympanic thermometers.

Overall, most children were rated by their parents as relaxed during temperature measurements with each thermometer. The median patient discomfort scale (completed by 306 parents) for each thermometer was 2, although the IQR was slightly larger for the electronic axillary thermometer than for the other thermometers: Thermofocus median 2 (IQR 2-2), Firhealth median 2 (IQR 2-2), electronic axillary median 2 (IQR 2-3) and tympanic median 2 (IQR 2-2).

Parental acceptability as assessed by a 10 cm visual analogue scale by 398 parents was highest for the Firhealth NCIT (median 9 cm; IQR 7.6-9.5), followed by the Thermofocus NCIT (median 8.5 cm; IQR 6.9-9.4), the tympanic thermometer (median 7.6 cm; IQR 5.5-9) and lastly the electronic axillary thermometer (median 5 cm; IQR 2.1-7.6).

Qualitative Interviews

The characteristics of 21 parents who participated in the interviews are described in supplementary Table 1. Themes relating to device attributes –what parents look for and feel they need from thermometers - are presented here, with key themes of convenience and practicality, comfort, cost, safety, and endorsement. Further qualitative data and analysis will be reported separately in a future publication.

Electronic Axillary thermometer

Parents' experiences with the axillary thermometers were described in almost all cases using negative language, with over half describing their child as appearing "uncomfortable" during its use. For some the appearance of the device contributed to the child's negative experience.

"He freaked out at that one.. Because he thought it was a needle.. so he thought he was going to have an injection." I8, Mother, age 31-40, 2 children

Parents raised concerns about the practicality of this device, however, if it were the medically recommended thermometer, they would be willing to persevere in using it.

"From a practical and user point of it, you know, it just seemed impractical, but if it's the best then so be it, you know, there was no harm" I1, Father, 2 children.

Tympanic thermometer

Parents' views on tympanic thermometers were more neutral than axillary thermometers, with several parents using them at home. They commonly expressed concern that thermometer performance might be affected by their failure to operate it correctly.

"I'm sure if I got it, if I was trained I might know better but I'm never sure if I've put it far enough into her ear or not far enough in or, I mean I'd always had a normal reading from it but I'm not confident that I wouldn't miss something by not being able to use it perfectly." I17, Mother, age 31-40, 1 child

NCITs

Parents highlighted the practicality and convenience of NCITs. The feedback regarding placement from the devices were reported to be helpful to ensure "correct" usage and interpretation.

"It was as if nothing was happening he was just sitting on my lap, he was perfectly calm, so they were, I was quite impressed with those actually, I didn't know they existed." I10, Father, 1 child, with history of fever this episode

“With the blue one there was a light, I thought that was quite helpful so you knew [um] exactly the distance you needed to be... I guess that also helps with the accuracy of reading because you know you’re taking it the right distance” I15, Mother, age 31-40, 1 child

Features of an ideal thermometer

Parents described five key attributes of thermometers that would influence their likelihood of using them: convenience of use, comfort for their child, and the cost, safety and endorsement of the device. These themes were reflected throughout the descriptions of their and their child’s experiences with the different devices.

Parents also displayed interest in the cost of devices for themselves and for the NHS.

“The non-contact.. it’s a lot easier there’s less parts to exchange on it, whereas with the underarm you’ve got to wipe it, with the ear one you’ve got to replace the caps so, you know, it all comes down to cost. Every patient comes in and you have to replace one cap for each patient you do an ear thermometer, you know, it all adds up doesn’t it.” I13 Father, 1 child

However, some parents required reassurance regarding the reliability and safety of novel, or less familiar, technology.

“I think it’s mainly because you’re kind of used to...like that physical contact for one or two minutes, so if you kind of use anything that is very different from the usual way of doing things then it always comes to your mind is this really reliable.” I14 Father, age 31-40, 2 children, with history of fever this episode

“I didn’t mind doing it, I did ask a couple of times if it was safe to use all of them... it was just because your colleague said it was infrared and they’re new they’re trialling them so I thought oh I’ll ask the question it is on a child.” I4, Mother, age 31-40, 2 children

DISCUSSION

Summary of main findings

In this methods comparison study we found that although the mean difference between NCIT and other thermometer measurements was only moderate, the upper and lower limits of agreement were more than 1°C. This has the potential to adversely affect clinical decision making, given that guidance relates to thresholds of temperature.(4) This also exceeds the limits of agreement of +/- 0.5 °C which are commonly believed to be acceptable.(21) The proportion classified as febrile out of all those with an electronic axillary measurement of 38°C and over was low for both NCITs. We found that reproducibility of measurements with the NCITs was good. Fewer attempts were required and fewer failures were reported for NCITs, and performance was similar for the cheaper and more expensive brands tested.

The majority of parents rated all devices as acceptable, although parental satisfaction and child discomfort appears to be better for the NCITs. The axillary thermometer was least liked by parents and children, and resulted in the most failed readings. When interviewed, parents expressed a preference for the practicality and comfort afforded by NCITS, and were predominantly negative about the user experience of axillary thermometers, felt to be more intrusive and have the potential to cause distress or discomfort, particularly if their child was unwell. Parents considered the convenience of use, the comfort of their child, cost, safety and endorsement when evaluating thermometers, however, there was a willingness to adopt whichever device was medically recommended.

Strengths and limitations

To our knowledge this is the first methods comparison study of peripheral thermometers to be conducted in children attending primary care with acute illness. 401 children were recruited, making this one of the larger studies to address this question. However, there are some limitations to our study. Firstly, uncertainty exists over the accuracy of the electronic axillary and tympanic methods we included as usual care. This study found wide limits of agreement between tympanic and axillary thermometers in this population. A recent systematic review(22) found 18 studies comparing tympanic thermometers and 19 studies comparing axillary thermometers to core temperature measurement (which was rectal measurement in the majority of cases) in children, with no studies conducted in primary care. Pooled estimates of mean difference was -0.43 (95% Limits of Agreement -1.40 to 0.55) for axillary and -0.15 (95% Limits of Agreement -0.67 to 0.37) for tympanic, although

heterogeneity was very high and findings varied when different brands were considered separately. The authors concluded that neither method met acceptable limits of agreement compared to core temperature measurements. Therefore it is impossible to be certain which of the thermometers was closest to the core temperature; NCITs could be better performing than established methods.

Furthermore, the diagnostic accuracy of axillary and tympanic thermometers for fever using core body temperature measurement as the reference standard is variable. A systematic review found five paediatric studies reporting sensitivities ranging from 14% to 63% for electronic axillary measurement and 13 paediatric studies reported sensitivity ranging from 0.23 to 0.87 for tympanic measurement of fever of 38°C or more. This complicates the interpretation of our estimates of the sensitivity and specificity of NCITs.

Secondly, as just over 10% of our participants were febrile on study entry, the confidence intervals around our estimate of sensitivity are wide. Our research team approached consecutive children arriving at the GP, but the parents who agreed to participate may have been those with children they felt to be less unwell which may have resulted in under-recruitment of febrile children. However over 30% of recruited children had received medication for fever in the previous 6 hours, and by limiting our recruited population to children with recent onset of illness we included those most likely to require temperature measurement as part of their assessment. Finally, the temperature measurements were performed by study personnel who inevitably developed expertise in using the equipment, which means that failure rates and accuracy may differ from what would be expected in practice staff or parents. However, all thermometers were simple to use and frequent use in primary care settings would ensure similar expertise developed.

Comparison with other literature

There have been no systematic reviews evaluating the accuracy of NCITs. However, the poor performance of electronic axillary and tympanic thermometers compared to core body temperature measurements has been highlighted in a number of reviews.(6,22–25) The mean difference of -0.14 and -0.16 °C between NCITs and electronic axillary thermometers found in this study is within the range demonstrated by other studies comparing NCITs to electronic axillary(26), tympanic(9) and electronic(11,12) and mercury rectal thermometers(10). The only study reporting comparable data for the Thermofocus brand compared to it to a mercury in-glass thermometer used in the axilla in ambulatory paediatrics and found greater agreement, with an overall mean difference of 0.07°C and limits of agreement of -0.62°C (-0.67 to -0.47) and 0.76°C (0.61 to 0.91) (8). However they used the average of 2 axillary measurements and 3 NCIT measurements in each child which could have

contributed to the improved performance. We found that both NCITS had very low sensitivity for fever as detected by axillary thermometry. Two studies using Thermoflash and Beurer NCITs have demonstrated low sensitivity for fever as detected by rectal thermometry.(11,12) However five studies,(9–11) two using the Thermofocus,(8,12) have found much higher sensitivity for fever.

While parents' knowledge and beliefs about fever and temperature management in children have previously been described,(26,27) there is little description of the impact or experience of different thermometers. Parental concerns regarding practicality and comfort were highlighted in a study exploring attitudes to rectal thermometers.(28)

Implications for future research and practice

In children under five, temperature measured by NCITs can vary by over 1°C from measurements made by axillary and tympanic thermometers, which is a potentially clinically significant variability. Given the uncertainty over the accuracy of electronic axillary and tympanic thermometers for core body temperature it is hard to draw firm conclusions about the likely impact on practice if NCITs were introduced as standard care.

If we assume that axillary thermometry is an accurate reference standard, then the moderate agreement between NCITs and axillary and our finding that a different population of children would be classified as febrile by NCITs compared to electronic axillary thermometers means that decisions about the best pathway of care for a child could vary depending on which thermometer is used (4). This in turn suggests that primary care clinicians should be cautious in using this technology.

However, we also found wide limits of agreement between the electronic axillary and tympanic thermometers, both of which are advocated in guidance, and existing studies cast doubt over the accuracy of either method compared to core body temperature measurement ((22)). Furthermore, parents were more positive about the benefits of NCITs than other types of devices in both rating scales and qualitative interviews, and there was good agreement between repeated measurements in both NCITs. A high quality study comparing NCITS, axillary and tympanic measurements to core body temperature measurements could help clarify this, but the invasive nature of core body temperature measurement will make this a challenge to conduct in a primary care population.

Therefore, clinicians need to be cautious about the accuracy of any peripheral thermometry approach and ensure that the management decisions they make are using this data as part of a holistic assessment. There is clear potential for technological innovation in this field to develop more accurate methods of peripheral thermometry to support clinical decision making.

FIGURE LEGENDS

Figure 1: Bland Altman plot for agreement between Thermofocus NCIT thermometer and electronic axillary thermometer. Solid line: mean difference between the two methods; dashed lines: upper and lower limits of agreement; dash-dotted line: line of no difference.

Figure 2: Bland Altman plot for agreement between Firhealth NCIT thermometer and electronic axillary thermometer. Solid line: mean difference between the two methods; dashed lines: upper and lower limits of agreement; dash-dotted line: line of no difference.

REFERENCES

1. Hay AD, Heron J, Ness A. The prevalence of symptoms and consultations in pre-school children in the Avon Longitudinal Study of Parents and Children (ALSPAC): A prospective cohort study. *Fam Pract*. 2005;22(4):367–74.
2. Fleming DM, Smith GE, Charlton JR, et al. Impact of infections on primary care--greater than expected. *Commun Dis public Heal*. 2002;5(1):7–12.
3. Armon K, Stephenson T, Gabriel V, et al. Determining the common medical presenting problems to an accident and emergency department. *Arch Dis Child*. 2001;84:390–2.
4. NICE. Fever in Under 5's: Assesment and Initial Management. Clin Guidel [Internet]. 2017;(May 2013). Available from: <https://pathways.nice.org.uk/pathways/fever-in-under-5s/symptoms-and-signs-of-specific-illnesses-in-children-with-fever#content=view-node%3Anodes-kawasaki-disease>
5. Omron. 20-SECOND FLEXIBLE DIGITAL SAFETY INFORMATION : [Internet]. p. 206. Available from: <http://ec1.images-amazon.com/media/i3d/01/A/man-migrate/MANUAL000039800.pdf>
6. Craig J V., Lancaster GA, Taylor S, et al. Infrared ear thermometry compared with rectal thermometry in children: A systematic review. *Lancet*. 2002;360(9333):603–9.
7. 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;
8. 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;20:1311–8.
9. Ng DK, Chan C-H, Lee RS, Leung LC. Non-contact infrared thermometry temperature measurement for screening fever in children. *Ann Trop Paediatr*. 2005;25(4):267–75.
10. 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(4):471–6.
11. Allegaert K, Casteels K, van Gorp I, Bogaert G. Tympanic, infrared skin, And temporal artery scan thermometers compared with rectal measurement in children: A real-life assessment. *Curr Ther Res - Clin Exp* [Internet]. 2014;76:34–8. Available from: <http://dx.doi.org/10.1016/j.curtheres.2013.11.005>
12. Paes BF, Vermeulen K, Brohet RM, et al. Accuracy of tympanic and infrared skin thermometers in children. *Arch Dis Child*. 2010;95:974–8.

13. Selent MU, Molinari NM, Baxter A, et al. Mass Screening for Fever in Children. Vol. 29, *Pediatric Emergency Care*. 2013. 305–313 p.
14. E Osio C, Carnelli V. Comparative study of body temperature measured with a non-contact infrared thermometer versus conventional devices: The first Italian study on 90 pediatric patients. Vol. 59, *Minerva pediatrica*. 2007. 327–336 p.
15. Fortuna EL, Carney MM, Macy M, et al. Accuracy of Non-contact Infrared Thermometry Versus Rectal Thermometry in Young Children Evaluated in the Emergency Department for Fever. *J Emerg Nurs* [Internet]. 2010;36(2):101–4. Available from: <http://dx.doi.org/10.1016/j.jen.2009.07.017>
16. Sollai S, Dani C, Berti E, et al. Performance of a non-contact infrared thermometer in healthy newborns. *BMJ Open*. 2016;6:e008695.
17. Greenes DS, Fleisher GR. Accuracy of a noninvasive temporal artery thermometer for use in infants. *Arch Pediatr Adolesc Med* [Internet]. 2001;155(3):376–81. Available from: <http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L32199186%0Ahttp://ca3cx5qj7w.search.serialssolutions.com?sid=EMBASE&issn=10724710&id=doi:&article=Accuracy+of+a+noninvasive+temporal+artery+thermometer+for+use+in+infants&stitle=Arch>
18. Wong D, Baker C. Pain in Children: Comparison of Assessment Scales. *Pediatr Nurs*. 1988;14(1):9–17.
19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* [Internet]. 1986;327(8476):307–10. Available from: <https://www-users.york.ac.uk/~mb55/meas/ba.pdf>
20. Bland M. How can I decide the sample size for a study of agreement between two methods of measurement? [Internet]. 2004. Available from: <https://www-users.york.ac.uk/~mb55/meas/sizemeth.htm>
21. Geijer H, Udumyan R, Lohse G, Nilsagård Y. Temperature measurements with a temporal scanner: Systematic review and meta-analysis. *BMJ Open*. 2016;6:e009509.
22. Niven DJ, Gaudet JE, Laupland KB, et al. Accuracy of peripheral thermometers for estimating temperature: A systematic review and meta-analysis. *Ann Intern Med*. 2015;163:768–77.
23. Craig J V, Lancaster GA, Williamson PR, Smyth RL. Temperature measured at the axilla compared with rectum in children and young people: systematic review. *Bmj*.

2000;320(7243):1174–8.

24. Zhen C, Xia Z, Long L, Pu Y. Accuracy of infrared ear thermometry in children: A meta-analysis and systematic review. *Clin Pediatr (Phila)*. 2014;53(12):1158–65.
25. Zhen C, Xia Z, Ya Jun Z, 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):114–26.
26. 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(2):180–6.
27. Kelly M, Sahm LJ, Shiely F, et al. Parental knowledge, attitudes and beliefs on fever: A cross-sectional study in Ireland. *BMJ Open*. 2017;7:e015684.
28. Kelly M, McCarthy S, O’Sullivan R, et al. Drivers for inappropriate fever management in children: a systematic review. *Int J Clin Pharm*. 2016;38(4):761–70.
29. Kai J. Parents’ perceptions of taking babies’ rectal temperature. *BMJ [Internet]*. 1993;307:660–2. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1679010/pdf/bmj00038-0032.pdf>

DETAILS OF ETHICAL APPROVAL: South Central – Berkshire Research Ethics Committee, 17/SC/0068.

DETAILS OF FUNDING: This project was funded by the HTA Programme 16/45/01. AvDB, GH and GE were supported by the National Institute for Health Research (NIHR) Community Healthcare MedTech and In Vitro Diagnostics Co-operative. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

ROLE OF THE SPONSOR: The sponsor had no role in study design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit the paper for publication.

COMPETING INTERESTS STATEMENT: "All authors have completed the Unified Competing Interest form (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work."

COPYRIGHT STATEMENT: "The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above."

CONTRIBUTIONS: GH, JV, FAI, KW, SF, MG, EM, and AvdB contributed to the planning of the study. All authors contributed to the conduct and reporting of the work described in this article. GH is responsible for overall content as guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

TRANSPARENCY DECLARATION: “The lead author (the manuscript’s guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained”.

DATA SHARING STATEMENT: No data available.

ACKNOWLEDGEMENTS: nil

	N (%)
Age (years, median and IQR)	1.6 (0.79-3.38)
Gender (boys)	203 (50.62)
Ethnicity	
- White British	280 (69.83)
- White other	38 (9.48)
- Mixed	27 (6.73)
- Pakistani	21 (5.24)
- Other Asian	11 (2.74)
- African	8 (2.00)
- Indian	5 (1.25)
- Chinese	5 (1.25)
- Bangladeshi	4 (1.00)
- Caribbean	1 (0.25)
- Black British	1 (0.25)
Mother's age (years, median and IQR)	32 (29-36)
Number of siblings (median and IQR)	1 (0-1)
Parent believed child to be febrile at point of assessment	119 (29.75)
Fever medication in past 6 hours	134 (33.50)
Parent report of fever duration (days, median and IQR)	1.5 (0.5-3)
Illness duration (days, median and IQR)	3 (2-7)
Recruitment site:	
OOH primary care	34 (8.5)
In hours primary care	367 (91.5)

Table 1: Participant characteristics

	Mean difference and 95% CI (°C)	Lower limit of agreement and 95% CI (°C)	Upper limit of agreement and 95% CI (°C)
Methods Comparison			

Thermofocus NCIT minus electronic axillary, n=371	-0.14 (-0.21 to -0.06)	-1.57 (-1.69 to -1.44)	1.29 (1.16 to 1.42)
Firhealth NCIT minus electronic axillary, n=374	-0.16 (-0.23 to -0.09)	-1.54 (-1.66 to -1.41)	1.22 (1.10 to 1.34)
Thermofocus NCIT minus tympanic, n=386	-0.10 (-0.17 to -0.03)	-1.55 (-1.68 to -1.42)	1.35 (1.22 to 1.48)
Firhealth NCIT minus tympanic, n=389	-0.10 (-0.17 to -0.03)	-1.47 (-1.59 to -1.35)	1.28 (1.16 to 1.40)
Thermofocus NCIT minus Firhealth NCIT, n=395	0.00 (-0.04 to 0.05)	-0.90 (-0.98 to -0.82)	0.91 (0.83 to 0.99)
Electronic axillary minus tympanic, n=367	0.06 (-0.02 to 0.14)	-1.49 (-1.63 to -1.34)	1.61 (1.47 to 1.75)
Reproducibility			
Thermofocus NCIT 1 st minus 2 nd reading, n=395	-0.04 (-0.07 to -0.01)	-0.56 (-0.60 to -0.51)	0.47 (0.43 to 0.52)
Firhealth NCIT 1 st minus 2 nd reading, n=397	0.01 (-0.02 to 0.04)	-0.60 (-0.65 to -0.54)	0.61 (0.56 to 0.67)

Table 1: Method Agreement results

	Thermofocus				Firhealth				Tympanic			
Sensitivity (%; 95% CI)	29.3 (16.1-45.5)				4.9 (0.6-16.5)				29.3 (16.1-45.5)			
Specificity (%; 95% CI)	96.9 (94.6-98.5)				98.9 (97.2-99.7)				93.6 (90.6-95.9)			
Positive predictive value (%; 95% CI)	52.2 (30.6-73.2)				33.3 (4.3-77.7)				34.3 (19.1-52.2)			
Negative predictive value (%; 95% CI)	92.3 (89.2-94.8)				90.1 (86.8-92.9)				92.1 (88.8-94.6)			
Likelihood ratio + (95% CI)	9.58 (4.52-20.31)				4.39 (0.83-23.24)				4.58 (2.47-8.50)			
Likelihood ratio - (95% CI)	0.73 (0.60-0.89)				0.96 (0.90-1.03)				0.76 (0.62-0.92)			
Absolute numbers (TP, FP, FN, TN)	12	11	29	349	2	4	29	356	12	23	29	337

Table 2: Diagnostic accuracy for fever defined as $\geq 38^{\circ}\text{C}$ by the electronic axillary thermometer

	Thermofocus NCIT N (%)	Firhealth NCIT N (%)	Electronic axillary N (%)	Tympanic N (%)
1 attempt required	382 (95.3)	390 (97.3)	363 (90.5)	364 (90.8)
2 attempts required	10 (2.5)	8 (2.0)	11 (2.7)	15 (3.7)
3 attempts required	4 (1.0)	1 (0.2)	2 (0.5)	9 (2.2)
No reading				
Technical error (thermometer not activating)	3 (0.8)	1 (0.2)	0 (0.0)	0 (0.0)
Technical error (other)	1 (0.2)	0 (0.0)	7 (1.7)	3 (0.7)
Lack of cooperation of the child	1 (0.2)	0	16 (4.0)	5 (1.2)
Reason not specified	0 (0.0)	1 (0.2)	2 (0.5)	0 (0.0)
Thermometer unsuitable	0 (0.0)	0 (0.0)	0 (0.0)	5 (1.2)

Table 4: Number of attempts needed to achieve first measurement and technical failures for each thermometer