

BMJ Open Temperature measurements with a temporal scanner: systematic review and meta-analysis

Håkan Geijer,^{1,2} Ruzan Udumyan,³ Georg Lohse,^{1,4} Ylva Nilsagård^{1,5}

To cite: Geijer H, Udumyan R, Lohse G, *et al*. Temperature measurements with a temporal scanner: systematic review and meta-analysis. *BMJ Open* 2016;**6**:e009509. doi:10.1136/bmjopen-2015-009509

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-009509>).

Received 23 July 2015
Revised 7 January 2016
Accepted 11 January 2016



CrossMark

¹Centre for Assessment of Medical Technology in Örebro, Region Örebro County, Örebro, Sweden

²Department of Radiology, School of Medical Sciences, Örebro University, Örebro, Sweden

³Department of Clinical Epidemiology and Biostatistics, School of Medical Sciences, Örebro University, Örebro, Sweden

⁴Örebro Rehab Center, Örebro, Sweden

⁵Department of Medicine, School of Health Sciences, Örebro University, Örebro, Sweden

Correspondence to

Dr Håkan Geijer; hakan.geijer@regionorebrolan.se

ABSTRACT

Objectives: Systematic review and meta-analysis on the diagnostic accuracy of temporal artery thermometers (TAT).

Design: Systematic review and meta-analysis. The index test consisted of temperature measurement with TAT. The reference test consisted of an estimation of core temperature.

Participants: Clinical patients as well as healthy participants, with or without fever.

Interventions: Literature search in PubMed, Embase, Cinahl and Web of Science. Three reviewers selected articles for full-text reading after which a further selection was made. Risk of bias was assessed with QUADAS-2. Pooled difference and limits of agreement (LoA) were estimated with an inverse variance weighted approach. Subgroup and sensitivity analyses were performed. Sensitivity and specificity were estimated using hierarchical models. Quality of evidence was assessed according to the GRADE system.

Primary and secondary outcome measures: The primary outcome was measurement accuracy expressed as mean difference \pm 95% LoA.

A secondary outcome was sensitivity and specificity to detect fever. If tympanic thermometers were assessed in the same population as TAT, these results were recorded as well.

Results: 37 articles comprising 5026 participants were selected. Pooled difference was -0.19°C (95% LoA -1.16 to 0.77°C), with moderate quality of evidence. Pooled sensitivity was 0.72 (95% CI 0.61 to 0.81) with a specificity of 0.94 (95% CI 0.87 to 0.97). The subgroup analysis revealed a trend towards underestimation of the temperature for febrile patients. There was a large heterogeneity among included studies with wide LoA which reduced the quality of evidence.

Conclusions: TAT is not sufficiently accurate to replace one of the reference methods such as rectal, bladder or more invasive temperature measurement methods. The results are, however, similar to those with tympanic thermometers, both in our meta-analysis and when compared with others. Thus, it seems that TAT could replace tympanic thermometers with the caveat that both methods are inaccurate.

Trial registration number: CRD42014008832.

Strengths and limitations of this study

- With 37 studies and 5026 study participants, this is the largest summary of the evidence for temperature measurements at the temporal artery.
- The sensitivity analysis did not change the overall result notably.
- A weakness is the large heterogeneity among included studies.

INTRODUCTION

Body temperature is one of the most commonly used parameters in healthcare. For this, reliable equipment must be used. There is no universal agreement on how accurate a thermometer must be, but the method is generally considered accurate and reliable if the mean difference is less than 0.2 to 0.5°C and the limits of agreement (LoA) are less than $\pm 0.5^{\circ}\text{C}$.^{1–3} Reference methods for temperature measurement have traditionally been rather invasive with measurements taken from the nasopharynx, oesophagus, pulmonary artery, brain or urinary bladder. There is thus a need to find a less invasive method for body temperature measurement as a replacement for the ‘reference’ methods.

Temperature measurement over the temporal artery (TAT, temporal artery thermometry) is a method for temperature measurement that uses infrared technology to detect the heat that is radiated from the skin surface over the temporal artery.

For many years, rectal measurements have been used as the clinical reference method with an acceptable balance between accuracy and degree of invasiveness. Recently, it has to a large degree been replaced by infrared ear thermometry, measuring at the tympanic membrane. However, this method is regarded as suboptimal, mainly because of poor repeatability and a tendency to show false low results compared with core temperature.^{4–6}

Previous literature reports have given mixed results of the value of TAT, and there

are no recent systematic reviews of the method. The purpose was thus to perform a systematic literature review and meta-analysis of the measurement accuracy of TAT compared with reference temperature. A secondary aim was to compare the accuracy of TAT and tympanic temperature measurement when both temperatures were measured on the same samples.

The study was designed as a systematic review.

METHOD AND MATERIALS

This systematic review has been registered in the PROSPERO International prospective register of systematic reviews (<http://www.crd.york.ac.uk/PROSPERO>), CRD42014008832.

Study identification

A literature search was performed by a librarian in the electronic databases PubMed/MEDLINE (search string “(temporal artery) AND (((temperature) OR thermometer) OR fever)”), Embase, Cinahl, Web of Science, The Cochrane Library, Trip, International Network of Agencies for Health Technology Assessment (INAHTA) and Centre for Reviews and Dissemination (CRD). Ongoing studies were searched via ClinicalTrials.gov. Reference lists of included studies were checked. The paper is based on the systematic search of literature published up to 29 September 2015.

Study selection and quality assessment

Three reviewers read all titles and abstracts independently. Obviously irrelevant articles were removed, whereas the full text of the potentially relevant articles was retrieved and assessed on the basis of the eligibility criteria for the inclusion in the current review. Disagreements were solved in consensus.

For selecting a study, all of these inclusion criteria should be fulfilled: (A) primary study; (B) temperature measurement at the temporal artery; (C) comparison with core temperature; (D) study performed in a health-care setting. Exclusion criteria were (A) non-human studies; (B) review articles, editorials, letter or congress abstracts; (C) insufficient data to report or calculate bias or sensitivity/specificity; (D) language other than English, French, German or one of the Nordic languages.

The subject matter was delimited according to PICO⁷ (population—intervention (index test)—comparison (reference test)—outcome) to clinical patients as well as healthy participants, with or without fever. The index test consisted of temperature measurement with TAT. The reference test consisted of an estimation of reference temperature, expressed as measurement in the nasopharynx, oesophagus, pulmonary artery, rectum, brain and urinary bladder. However, participants received verification with the same reference standard within each study.

All included studies were assessed for methodological quality by three independent reviewers according to QUADAS-2.⁸ Disagreements were solved in consensus. Most focus was laid on the domain *Flow and Timing* since the timing between temperature measurements was deemed to be the most crucial part. The process of recording the temperature consisted simply of recording a figure, so blinding was not deemed to be as important.

Outcomes

The primary outcome was measurement accuracy of the index test compared to a reference standard, expressed as pooled estimates of mean temperature difference (systematic error) and 95% LoA (random error). The secondary outcome was average summary estimates of test sensitivity (SE) and specificity (SP) at a chosen test threshold. If tympanic thermometers had been assessed in the same population as the TAT, these results were recorded as well.

Data extraction

Two reviewers independently extracted the relevant data and resolved disagreements through discussion with other reviewers.

From each included study, we retrieved information on study and patient characteristics, type of the index test thermometer, reference standard and information on comparator test, if available, and relevant statistics: mean difference (TAT—reference) and SD of the differences in temperature readings. Mean differences and SD reported in Fahrenheit were converted into Celsius. When mean differences and/or SD of the differences were not directly reported, we computed them from other reported data using standard formulae. Thus, SD of the mean difference was computed from CIs, range of differences, SD for each thermometer and the correlation coefficient, or mean difference and t-statistic. In one study, the mean difference and SD were estimated after extracting individual values from the figures. When possible, we also extracted paired estimates of sensitivity and specificity.

Data analysis

Mean difference in temperature readings

To obtain pooled estimates of systematic error (bias) and random error (LoA), we used the inverse variance weighted approach to combine individual study estimates of the mean difference and SD. More details on the techniques used in this meta-analysis can be found in Williamson *et al.*⁹

Pooled estimates of the differences and limits of agreement were calculated using a random-effects approach.¹⁰

To explore possible reasons for heterogeneity, we performed subgroup analyses. We hypothesised a priori that age, type of thermometer, presence/absence of fever and reference standard may be sources of heterogeneity across studies, and performed subgroup meta-analyses

according to these characteristics where sufficient data were available.

Several sensitivity analyses were performed in various combinations excluding studies with a high risk of bias (in the domain *Flow and Timing*); studies that used replicated data in pairs using differences for each pair of measurements and did not provide information on how they accounted for within-person correlation of observations¹¹; or studies lacking information on whether SD of the difference was corrected,^{11 12} when means of repeated measurements by each of the two methods on the same participant were used to evaluate the agreement between the two methods (see online supplementary appendix for details).

Sensitivity and specificity

We used coupled forest plots and a summary receiver operating characteristics (sROC) plot to display SE and SP estimates from individual studies, and obtained average summary estimates of SE and SP from studies that reported results at selected common positivity thresholds ($t \geq 38.0^\circ\text{C}$) using bivariate random-effects meta-analysis.¹³ The bivariate model jointly analyses pairs of SE and SP to account for the patterns of correlation between the two measures. To check the robustness of the results, we performed sensitivity analysis by excluding influential studies and outliers. We used Cook's distance to identify influential studies and standardised level-2 residuals to identify outliers.^{14 15} We did not investigate publication bias, since standard tests for publication bias are not recommended in meta-analysis of diagnostic accuracy studies.¹⁶

Statistical analysis was performed using Stata 12/SE, including the user written programmes.^{14 15} A Stata programme, has been written incorporating formulae described in Williamson *et al*⁹ to obtain the pooled estimate of systematic error and LoA utilising random-effects methods.

Quality of evidence (GRADE)

We assessed the quality of evidence for the estimation of pooled difference and LoA according to the GRADE system taking into account risk of bias, consistency, directness, precision and publication bias.¹⁷

Health economy

A simplified health economic assessment was performed, comparing TAT and tympanic measurements. The time for performing measurements was assumed to be equal for the two thermometers.³

RESULTS

The literature search resulted in 626 hits. Another 27 articles were added after a manual search of reference lists. After duplicate removal, 558 articles remained. Of these, 97 articles were selected for full-text reading. Thirty-seven of these fulfilled the inclusion and

exclusion criteria and were selected for final analysis. Of these, the decision was unanimous in 34 cases. Two reviewers agreed on two cases, and in the final included case only one reviewer initially advocated inclusion. The selection process is shown in figure 1. Study characteristics are shown in table 1.

A literature search in The Cochrane Library resulted in six hits, including two primary studies, of which one was included via the primary search.¹ The search of ClinicalTrials.gov resulted in nine studies, of which seven were completed, one cancelled and one awaiting start of recruitment. One of the completed studies has been published.⁴⁸ The search of the Trip database contributed nothing new while CRD gave three reviews but no new primary studies.

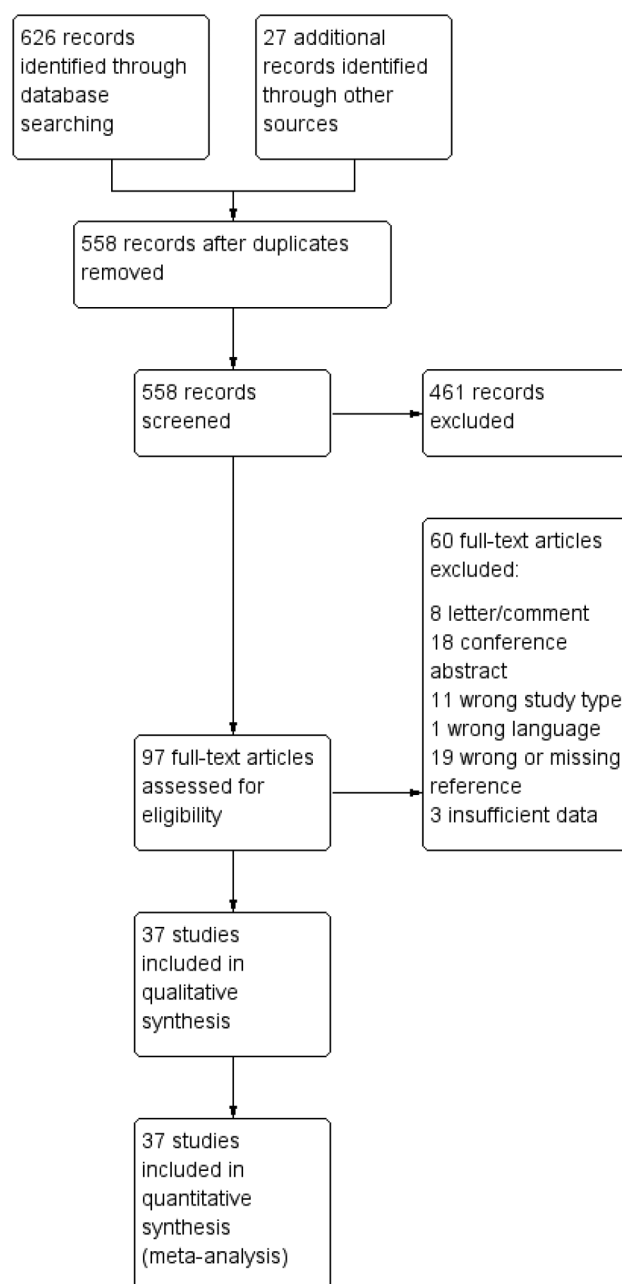


Figure 1 Study flow diagram.

Table 1 Study characteristics of the 37 included studies

Author, year, country	Inclusion criteria	Population	Febrile status	Maximum time between measurements	Temporal artery device*	Reference standard	Other comparison
Allegaert 2014, Belgium ¹⁸	Children admitted to paediatric wards	294, median age 3.2 years, range 0–17 years	Febrile and afebrile	5 min	TAT-5000	Rectal temperature (Filac 3000, Covidien, Mechelen, Belgium)	Tympanic (AccuSystem Genius2 Tympanic Infrared Ear Thermometer, Covidien, Mechelen, Belgium)
Al-Mukhaizeem 2004, Canada ¹⁹	Children undergoing elective dental surgery requiring endotracheal tube placement	80, mean age 45 months (SD 35)	2 febrile	Unclear	LXTA Temporal scanner (Exergen, Watertown, Massachusetts, USA)	Oesophageal temperature probe (TeleThermometer, YSI Incorporated, USA)	
Bahorski 2012, USA ²⁰	Infants and children presenting in emergency centre, ICU and outpatient unit	47, 43% male, age 3 to 36 months	Febrile (47%) and afebrile	Rapid sequential manner	TAT-5000	Rectal temperature (Welch-Allyn)	
Batra 2013, India ²¹	Children 2–12 years, emergency room setting	50 febrile, mean age 6.1 years, 48% male. 50 afebrile, mean age 6.15 years, 60% male	50 febrile and 50 afebrile	Unclear	Exergen TAT-2000C (Exergen)	Rectal temperature, mercury thermometer (Hicks Thermometers, Aligarh, India)	Axillary, tympanic (EQ ET 99, Equinox Overseas Private, New Delhi, India)
Callanan 2003, USA ²²	Infants under 3 months in emergency department	187 measured with both methods	Afebrile and 23 febrile	Unclear	SensorTouch TA (Exergen)	Rectal temperature (SureTemp, WelchAllyn)	
Calonder 2010, USA ²³	Adults undergoing surgery	23, mean age 55.7 years (SD 13.4), 26% male. Two measurements each	Afebrile	2 min	TAT-5000	Oesophageal probe (Smiths Medical, Dublin, Ohio, USA)	Oral
Carr 2011, USA ²⁴	Inpatients 0–24 months	40, mean age 10.9 months, 55% male	Febrile	Unclear	TAT-5000	Rectal temperature (Sure Temp, Welch Allyn Instruments)	
Drake-Brockman 2014, Australia ²⁵	Children undergoing general anaesthesia for routine elective non-cardiac surgery	200, mean age 8.44 years (SD 0.17), 59% male	Unclear	Concurrently	TAT-5000	Nasopharyngeal temperature (IntelliVue MP800, Philips, Amsterdam, Netherlands)	Skin temperature, tympanic (TermoScan 6021, Braun, Melsungen, Germany)

Continued

Table 1 Continued

Author, year, country	Inclusion criteria	Population	Febrile status	Maximum time between measurements	Temporal artery device*	Reference standard	Other comparison
Dybwik 2003, Norway ²⁶	Adult patients in intensive care	164	Afebrile and febrile	Unclear	Exergen TAT-4000 (Exergen)	Rectal temperature (Terumo C402)	
Furlong 2015, USA ²⁷	Adult patients in cardiac surgical intensive care	60, mean age 60.8 years (SD 15.2), 68% male	Febrile	Simultaneously	TAT-5000	Pulmonary artery catheter (Swan-Ganz VIP; Edwards Lifesciences, Irvine, California)	
Greenes 2001, USA ²⁸	Infants in emergency department, younger than 1 year	304	36% febrile	Unclear	LXTA Temporal scanner (Exergen)	Rectal temperature (Diatek, Welch Allyn, Skaneateles Falls, New York, USA)	Tympanic (FirstTemp Genius, Sherwood Medical, St Louis, Missouri, USA)
Greenes 2004, USA ²⁹	Infants under 1 year in emergency department given an antipyretic drug	45, mean age 210 days (range 11–335)	All febrile	Unclear	LXTA Temporal scanner (Exergen)	Rectal temperature (Diatek, Welch Allyn, Skaneateles Falls, New York, USA)	
Gunawan 2010, Indonesia ³⁰	Neonates more than 24 h old	134, mean age 36 h (SD 13 h), 52% male	Maximum 37.8°C	Unclear	TAT-5000	Rectal temperature (Clinical thermometer-CE 0197, China)	
Hamilton 2013, Argentina ³¹	Paediatric inpatients or outpatients	212, 205 completed study, 58% male	46% febrile	5 min	TAT-5000	Under 5 years rectal, over 5 years oral temperature (SureTemp Plus, Welch Allyn, Skaneateles Falls, New York, USA)	Tympanic (ThermoScan PRO 4000 IR, Braun, Kronberg, Germany)
Hebbar 2005, USA ¹	Patients in paediatric ICU	44, mean age 11.5 months (25th–75th percentile 2–34 months)	Afebrile and febrile	Unclear	LXTA Temporal scanner (Exergen)	Pulmonary or rectal temperature (Allegiance Healthcare Corporation, McGaw Park, Illinois, USA)	
Holzhauser 2009, USA ³²	Children 3–36 months presenting at emergency department	474 enrolled, 201 febrile included	Afebrile and febrile (42%)	Unclear	Exergen TAT (Exergen)	Rectal temperature (Welch Allyn, New York, USA)	
Kimberger 2007, Austria ²	Adult neurosurgical patients	35 in surgery, mean age 49 years (SD 25), 34% male; 35 in ICU, mean age 58 years (SD 19), 51% male	Afebrile and febrile	Simultaneously	TAT-5000	Bladder temperature sensor (SmithsMedical, London, UK)	

Continued



Table 1 Continued

Author, year, country	Inclusion criteria	Population	Febrile status	Maximum time between measurements	Temporal artery device*	Reference standard	Other comparison
Kirk 2009, UK ³³	16 years or older within 24 h of severe traumatic brain injury	20, median age 33 years, 80% male	Unclear	Unclear	TAT-5000	Brain temperature (ICP/temperature probe, Neurovent-PTemp, Raumedic AG, Münchberg, Germany)	Tympanic (Core-Check model 2090, IVAC, San Diego, California, USA)
Langham 2009, USA ³⁴	Adult surgical patients	50, mean age 57 years (SD 14), 48% male	Afebrile and febrile	5 min	TAT-5000	Bladder temperature (Foley catheter with thermistor, Mon-a-therm FoleyTemp, Mallinckrodt Anesthesiology, St. Louis, Missouri, USA)	Tympanic (FirstTemp Genius 3000A, Kendall, Mansfield, Massachusetts, USA)
Lawson 2007, USA ³⁵	Adult patients in intensive care with pulmonary artery catheter	60, mean age 57 years (SD 15), 67% male	Afebrile and febrile	1 min	TAT-5000	Pulmonary artery Swan-Ganz catheter (Edwards Lifesciences, Irvine, California, USA)	Tympanic (Genius Infrared Tympanic Thermometer 3000A, Sherwood Medical, St Louis, Missouri, USA)
Lee 2011, USA ³⁶	Neonatal in intensive care	34, mean age 35.7 weeks (SD 1.8), 53% male	Afebrile	2 min	TAT-5000	Indwelling rectal probe (oesophageal/rectal temperature probe, Smiths Medical ASD, Rockland, Massachusetts, USA)	Axillary
Mangat 2010, UK ³⁷	Adult surgical patients	61, mean age 66 years (SD 14), 75% male	Afebrile	Unclear	TAT-5000	Nasopharyngeal (Thermistor 400 series 9 Fr, Mallinckrodt, USA)	Tympanic (Genius 2 in core mode, Covidien, Hampshire, USA and PRO4000, Braun, Germany)
Moore 2015, USA ³⁸	Children 3 months to 4 years	239, mean age 1.5 years (SD 0.77), 53% male	41% febrile	'Immediately following'	Temporal scanner (Exergen)	Rectal (Alaris Medical Sciences, San Diego, California, USA)	
Myny 2005, Belgium ³⁹	Orally intubated patients in ICU	57, mean age 60 years (SD 14.9), 60% male	Afebrile and febrile	3 min	LXTA Temporal scanner (Exergen)	Pulmonary artery catheter (Baxter Health Care, Irvine, USA)	
Nimah 2006, USA ⁴⁰	Children under 7 years in intensive care	36, mean age 20.0 months (SD 18.6 months), 58% male	51% febrile	In a rapid manner (unclear)	SensorTouch HF370 (Philips, Chicago, Illinois, USA)	Bladder temperature (RSP Foley Catheter with 400 Series thermistor, Respiratory Support)	Tympanic (Thermoscan IRT 3020 and IRT 3520, Braun, Kronberg, Germany)

Continued

Table 1 Continued

Author, year, country	Inclusion criteria	Population	Febrile status	Maximum time between measurements	Temporal artery device*	Reference standard	Other comparison
Odinaka 2014, Nigeria ⁴¹	Children under 5 years in emergency department	156, mean age 10.8 months (SD 13.6), 52% male	Afebrile and febrile (51%)	Simultaneously	Exergen TAT-2000C (Exergen)	Products Inc, San Diego, California, USA) Rectal (mercury in glass)	
Penning 2011, Netherlands ⁴²	Children 0–18 years	198, mean age 5.1 years (SD 4.7), 61% male	Afebrile and febrile (41%)	Max 15 min after rectal	TAT-5000	Rectal temperature (Terumo C402/C202, Terumo, Tokyo, Japan)	
Reynolds 2014, USA ⁴³	Children under 4 years admitted to emergency department	52, mean age 13.5 months (SD 8.0), 60% male.	Febrile (15%) and afebrile	Unclear	TAT-5000	Rectal temperature (Sure Temp Plus 690, Welch Allyn, Skaneateles Falls, New York, USA)	Axillary temperature
Rubia-Rubia 2011, Spain ⁴⁴	Patients over 18 years old admitted to intensive care	201, mean age 59 years (SD 11), 74% male	Afebrile and febrile	Simultaneously	ThermoTouch Baby (Chicco, Grandate, Italy)	Pulmonary artery catheter	Infrared ear thermometer
Sahin 2012, Turkey ⁴⁵	Children who underwent elective lower abdominal surgery	60, mean age 1.84 years (SD 1.17), 45% male	Afebrile	5 min	PlusMRD Infrared Temporal Artery Thermometer (pM 1–802, PlusMED, Istanbul, Turkey)	Nasopharyngeal temperature (GE Datex-Ohmeda S/5, Datex-Ohmeda, Madison, Wisconsin, USA)	Axillary mercury-glass thermometer
Schuh 2004, Canada ⁴⁶	Children under 24 months in emergency department	327, mean age 9.2 months (SD 6.8)	Afebrile and febrile	Unclear	LXTA Temporal scanner (Exergen)	Rectal temperature (IVAC 2000, ALARIS Medical Systems, San Diego, California, USA)	
Siberry 2002, USA ⁴⁷	Children up to 2 years presenting for acute care visit	275, mean age 11.2 months (range 0–24), 49% male	Afebrile and febrile	Unclear	LXTA Temporal scanner (Exergen)	Rectal temperature (SureTemp, WelchAllyn)	
Singler 2013, Germany ⁴⁸	Patients ≥75 years in an emergency department	427 patients, mean age 82.7±5.1 years, 159 (37%) male	67 (15.7%) febrile	Unclear	TAT-5000	Rectal temperature (IVAC TEMP PLUS II Model 2080)	Tympanic (Braun ThermoScan ear thermometer)
Stelfox 2010, Canada ⁴⁹	Adults in intensive care	14, mean age 51 years (SD 18), 36% male	Afebrile and febrile	Rapid sequential manner	TAT-5000	Bladder temperature (Foley Catheter temperature Sensor, Smiths Group, Rockland, USA)	
Suleman 2002, USA ⁵⁰	Adult and paediatric patients recovering from	56, 30 adults (56 ±15 years old) and 26 children (3	Febrile	Simultaneously	SensorTouch (Philips)	Pulmonary catheter in adults and bladder catheter in children	

Continued



Risk of bias

The risk of bias and applicability concerns are summarised in figure 2. In general, the patient selection

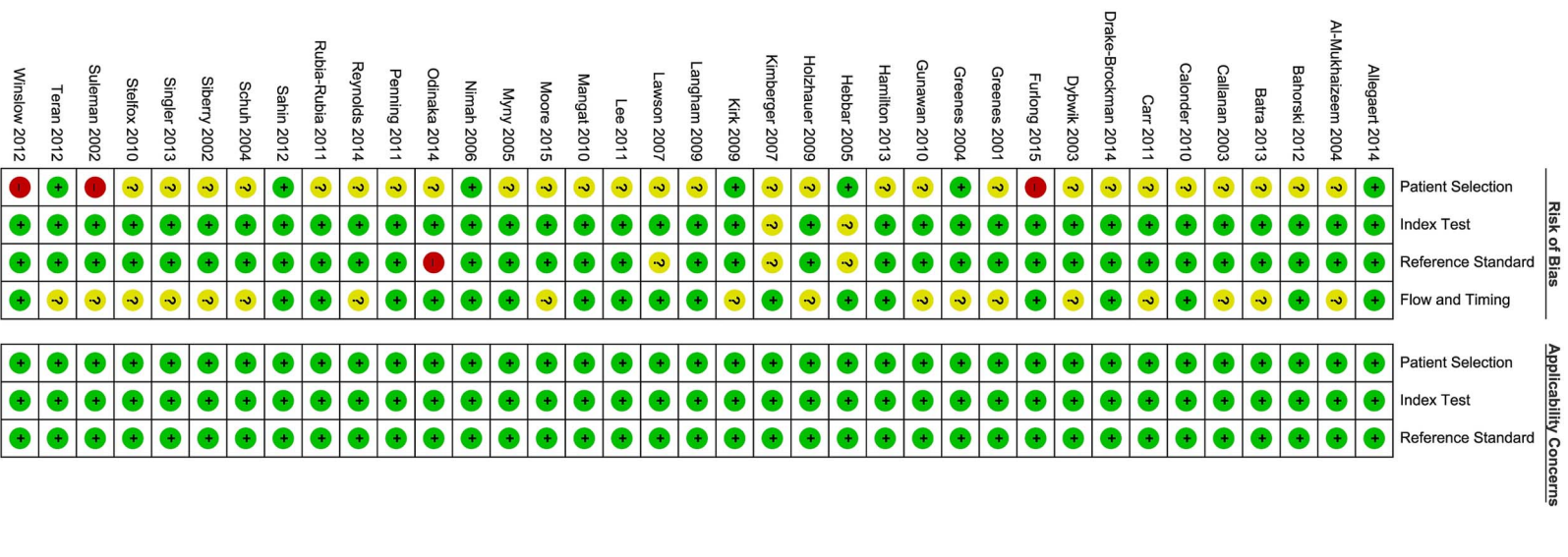


Figure 2 Risk of bias and applicability concerns summary.

Table 1 Continued

Author, year, country	Inclusion criteria	Population	Febrile status	Maximum time between measurements	Temporal artery device*	Reference standard	Other comparison
Teran 2012, Bolivia ⁵¹	cardiopulmonary bypass Children in ER and inpatient unit, 1 to 48 months	±4 years old). 15 +16 febrile of these selected 434, mean age 14.6 months, SD 10.7. 48% male	167 (38%) febrile	15 s	Exergen TAT-2000C	Rectal temperature (glass mercury thermometer)	
Winstlow 2012, USA ⁵²	Convenience sample with scheduled surgery over 18 years	64, mean age 57 years (33% male)	Afebrile	–	TAT-5000	Bladder temperature (Bardex Lubricath 400-Series and Lubri-Sil Foley Catheter, Bard, Covington, Georgia, USA)	

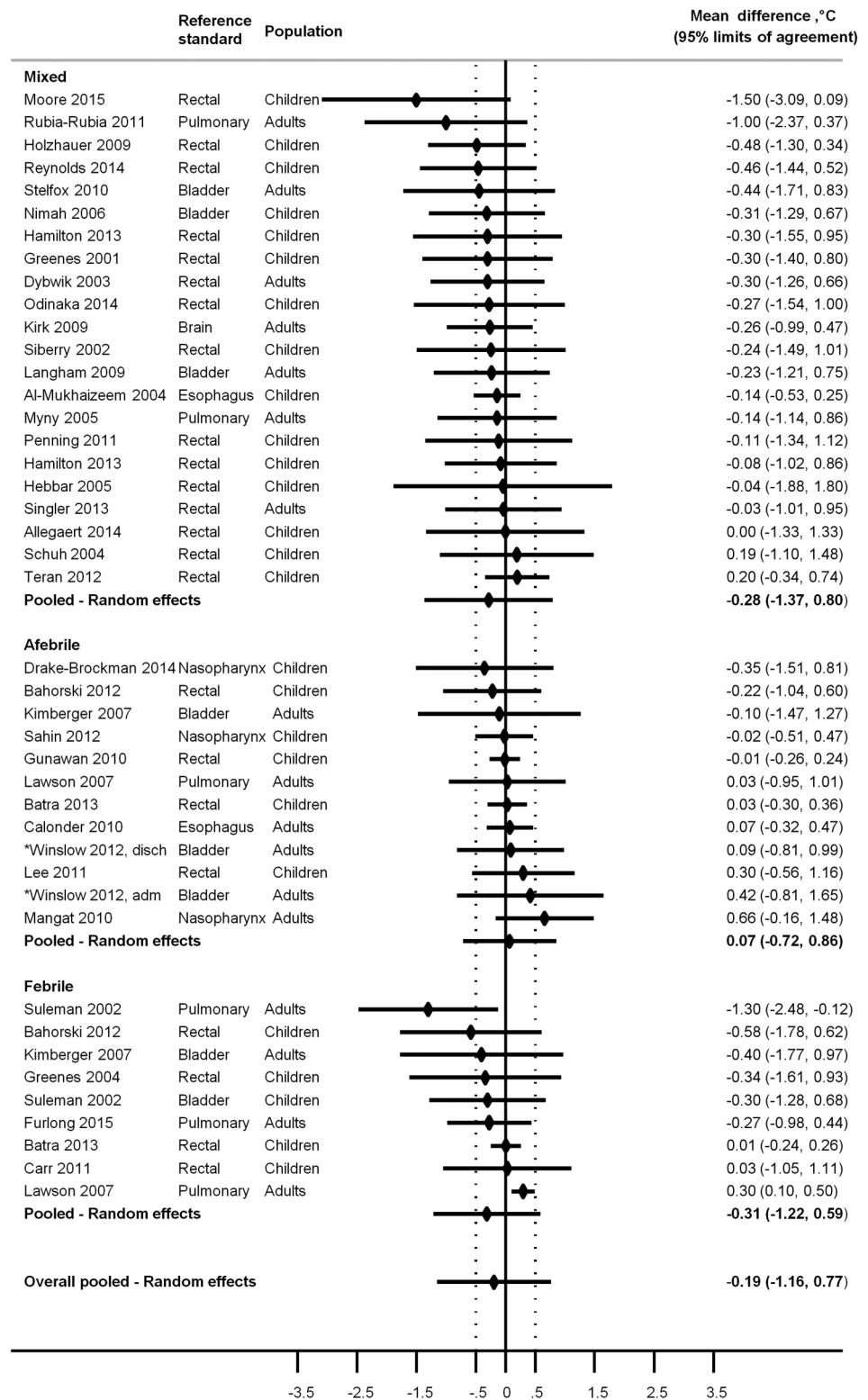
*TAT-5000: Exergen TemporalScanner TAT-5000 (Exergen, Watertown, Massachusetts, USA). TAT, temporal artery thermometers.

consisted of convenience samples that were not consecutive or randomised. Financial support was regarded as a possible source of publication bias. Seven articles reported support by grants from manufacturers.^{19 28 29 31 40 46 50} Another five studies were supported with instruments from the manufacturers.^{1 20 22 42 47}

Pooled mean difference in temperature readings

The 37 included articles comprise altogether 5026 study participants, 1301 adults and 3725 children. Thirty-six articles reported mean differences from the reference method, and some provided estimates for different sub-groups resulting in 43 comparisons. The overall random-effects pooled mean difference in temperature readings

Figure 3 Mean temperature difference (temporal artery thermometer –reference standard) and 95% limits of agreement by febrile status.



from these 43 comparisons was -0.19°C (95% LoA -1.16 to 0.77°C) (figure 3).

Subgroup and sensitivity analyses

There was a trend towards larger differences from the reference for febrile patients, with an underestimation of the temperature, mean difference -0.31°C (95% LoA -1.22 to 0.59°C), while the afebrile group was closer to the reference, mean difference 0.07°C (95% LoA -0.72 to 0.86°C) (figure 3). The results for adult and children subgroups were almost identical, mean difference -0.20°C (95% LoA -1.17 to 0.76°C) for children and -0.17°C (95% LoA -1.14 to 0.79°C) for adults (table 2). Grouping by reference standard did not show any differences. When grouping by type of TAT, the TAT-5000 thermometer (22 comparisons) had a result similar to all others.

Excluding studies with an 'Unclear' or 'High' risk of bias in the domain Flow and Timing, or studies lacking information on how they dealt with multiple measurements on the same participant, did not change results notably (pooled differences ranging from -0.09 to -0.19°C ; see online supplementary appendix for details).

Average summary estimates of SE and SP at the $\geq 38.0^{\circ}\text{C}$ cut-off value

Sixteen articles reported data on SE and SP. The SE varied between 0.26 and 0.94 while the SP varied between

0.46 and 1.00. The cut-off for test positivity ranged from ≥ 37.8 to $\geq 39.0^{\circ}\text{C}$.

We pooled the results from 14 studies (1 adult and 13 paediatric) including 1568 participants with fever, and 2566 participants without fever to estimate summary estimates of SE and SP at the $\geq 38.0^{\circ}\text{C}$ threshold. The reference test was rectal temperature in 13 studies, and bladder temperature in 1 study. SE and SP estimates and their 95% CI from each of these studies are displayed using coupled forest plots (figure 4A). The sROC plot (figure 4B) shows the 95% confidence and prediction regions. There was substantial heterogeneity for both SE and SP with greater variability in estimated SP than SE across studies. Bivariate random-effects meta-analysis produced the following summary estimates: SE 0.721 (95% CI 0.610 to 0.810), SP 0.939 (95% CI 0.865 to 0.973), positive likelihood ratio 11.8 (95% CI 5.3 to 26.1), and negative likelihood ratio 0.30 (95% CI 0.21 to 0.42). Since most studies had fewer participants with fever than without fever, estimates of SP are more precise than those of SE.

On the basis of Cook's distance, we found the studies by Teran *et al*⁵¹ and Siberry *et al*⁴⁷ to be the most influential in the meta-analysis (in descending order) (figure 5). Of these, Teran *et al* was identified as an outlier having the highest standardised residuals for SP (figure 5). After refitting the model and leaving this study out, bivariate random-effects meta-analysis produced the following summary estimates: SE 0.690 (95% CI 0.590 to 0.780) and SP 0.92 (95% CI 0.84 to 0.96).

Comparison with tympanic thermometers

Eleven articles included comparison with tympanic thermometers in the same population, comprising 1764 participants. In these articles, the mean difference from the reference method for TAT was -0.06°C (95% LoA -0.92 to 0.79°C) and for tympanic thermometers it was -0.29°C (95% LoA -1.15 to 0.57°C).

Four articles reported SE and SP for TAT and tympanic thermometers at the $\geq 38.0^{\circ}\text{C}$ threshold in the same population, 734 participants.^{18 21 28 40} The results were similar with SE 0.70 (95% CI 0.28 to 0.93) and SP 0.99 (95% CI 0.85 to 1.00) for tympanic thermometers.

Quality of evidence (GRADE)

The quality of evidence was graded for the overall result of pooled difference from the reference method with LoA. The quality level was rated down by one point due to inconsistency between the trials (point estimates ranging from -1.50 to 0.66°C). We considered that having support from manufacturers was not enough risk to downgrade on publication bias. This resulted in a moderate evidence quality ($\oplus\oplus\oplus\circ$) for a 95% LoA of -1.16 to 0.77°C (table 3).

Economic analysis

The local procurement price for the TAT is SEK 4200, and for a tympanic instrument it is SEK 895. For the

Table 2 Estimates of the pooled mean difference and 95% LoA between the temporal artery thermometer and reference standard. Random-effects meta-analysis*

	Pooled mean difference, $^{\circ}\text{C}$ (95% limits of agreement)	Number of comparisons
Overall	-0.19 (-1.16 to 0.77)	43
<i>Subgroup analysis</i>		
Reference standards†		
Rectal	-0.19 (-1.21 to 0.81)	23
Oesophagus	-0.03 (-0.43 to 0.36)	2
Bladder	-0.17 (-1.30 to 0.95)	8
Nasopharynx	0.09 (-0.73 to 0.91)	3
Pulmonary artery	-0.40 (-1.30 to 0.51)	6
Patient factors		
Children	-0.20 (-1.17 to 0.76)	26
Adults	-0.17 (-1.14 to 0.79)	17
Febrile status		
Febrile	-0.31 (-1.22 to 0.59)	9
Afebrile	0.07 (-0.72 to 0.86)	12
Mixed	-0.28 (-1.37 to 0.79)	22
Thermometer factors‡		
TAT-5000	-0.10 (-1.09 to 0.89)	22
Other	-0.27 (-1.23 to 0.67)	20

*Random-effects pooled estimates are calculated according to Williamson *et al*.⁹

†One study used the brain.

‡Thermometer type was unclear in one study.

LoA, limits of agreement; TAT, temporal artery thermometers.

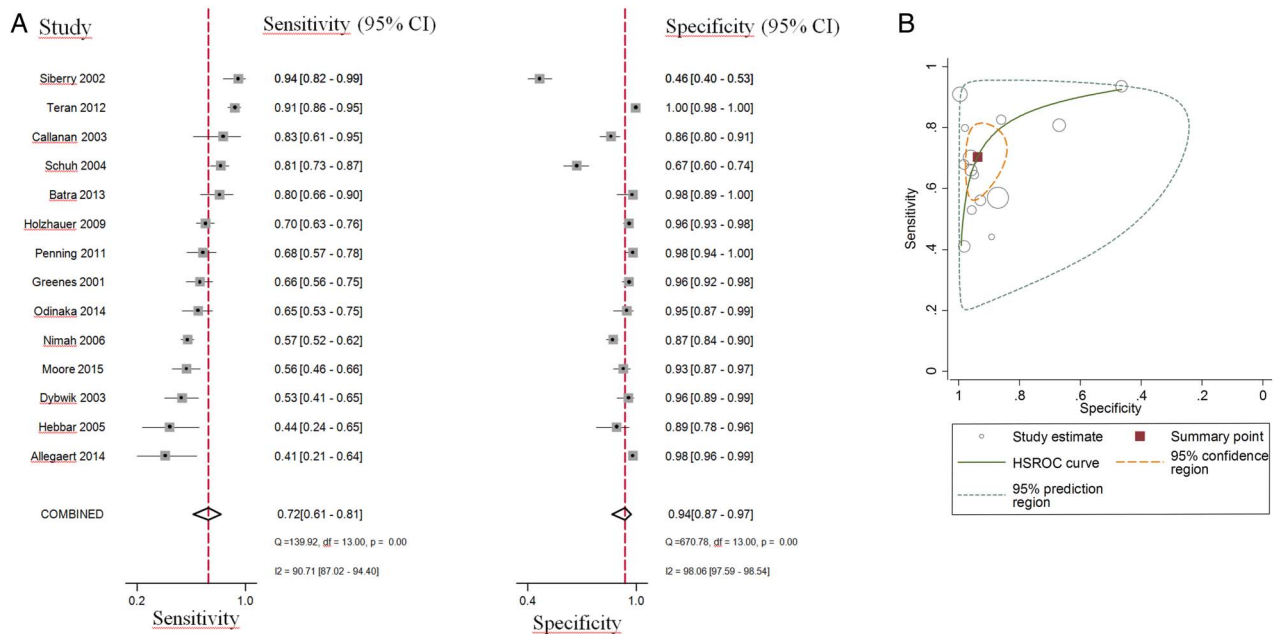


Figure 4 Accuracy of temperature measurement with a temporal artery thermometer measured through sensitivity and specificity. Pooled estimates obtained by a bivariate random-effects model (A) Coupled forest plot, (B) Summary receiver operating characteristics plot of sensitivity and specificity at $\geq 38.0^{\circ}\text{C}$ cut-off value. Each circle shows individual study estimates; inner ellipse represents 95% confidence region, and outer ellipse represents 95% prediction region for a future study.

tympanic instrument, a single-use protective cover is needed. With an interest rate of 2% and an assumed depreciation time of 6 years for the TAT and 4 years for the tympanic instrument, the cost per measurement would be equal at about 1100 measurements per year. For fewer measurements per instrument, the tympanic instrument would be cheaper.

DISCUSSION

The present meta-analysis indicates that TAT has a pooled difference from the reference of -0.19°C with 95% LoA -1.16 to 0.77°C or about $\pm 1.0^{\circ}\text{C}$. Common criteria for what is a clinically acceptable deviation from the reference temperature have been reported as LoA less than $\pm 0.5^{\circ}\text{C}$.^{1 2} TAT exceeds this level considerably, and it cannot be recommended as a replacement for one of the reference methods. The diagnostic accuracy was, however, very similar when compared with tympanic thermometers in the same participants. The subgroup analysis showed a trend towards lower temperature estimates in febrile patients, which in part may explain the rather low sensitivity of 0.72 and specificity of 0.94. In the literature, the minimum sensitivity acceptable to clinicians has been stated to be 0.9.^{32 46 47} Except for this, the performance was rather similar regardless of the reference method, adults versus children or type of instrument. The sensitivity analysis did not show any significant influence when we adjusted for study quality or statistical methods in the articles. The risk of bias analysis showed that the study populations were in general highly selected with convenience samples most common. Blinding was almost non-existent but was not judged to be a problem since most instruments give a digital figure that simply has to be recorded without interpretation. The timing between index and reference methods was, however, judged to be important since various parts of the body react differently when temperature is rising or falling.²⁹ The quality of evidence was rated as moderate due to inconsistency between the included studies. Publication bias was difficult to

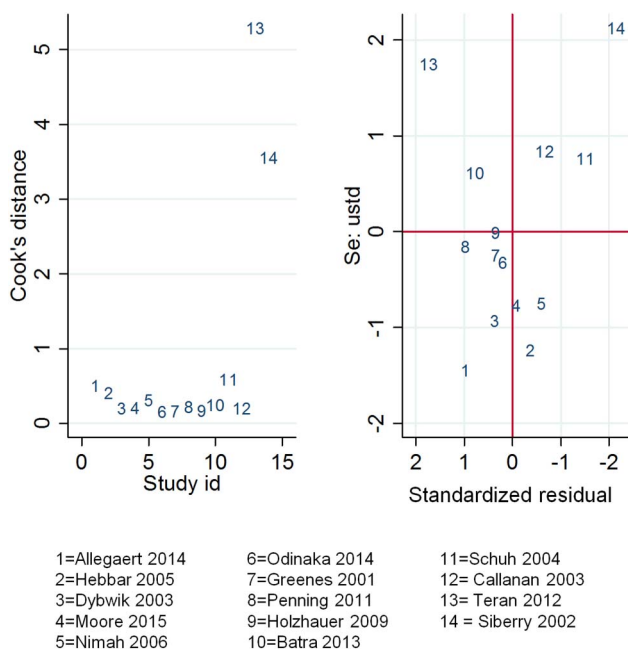


Figure 5 Influential and outlying studies.

Table 3 GRADE evidence profile

Test result	Study design	Factors that may decrease quality of evidence					Publication bias	Test property (95% limits of agreement)	Quality of evidence (GRADE)
		Risk of bias	Directness	Consistency	Precision				
Point estimate with LoA, overall	Cross-sectional design, 37 studies, 5026 participants (⊕⊕⊕⊕)	0	0	−1	0	0	Pooled difference −0.19°C (−1.16°C to 0.77°C)	⊕⊕⊕⊕	

LoA, limits of agreement.

evaluate, which is common in studies on diagnostic accuracy. The annual cost for temperature measurements is not high compared to other aspects of health-care. The largest influence on cost is probably personnel cost, so an instrument with a long measurement process is probably more expensive than instruments with rapid measurements such as the TAT.

It has been shown that TAT gives less discomfort and pain to children compared with rectal and axillary instruments.^{24 28 32 36} The rectal thermometer has also been reported to be frightening and psychologically harmful for children and there is always a risk of perforation and infection.^{53 54} Long-term risks are not known, but rectal temperature measurements could together with other painful, stressful and integrity insulting procedures add to psychological suffering for the child. Another fact in favour of TAT is that the patient does not need to be awake for temperature measurement. If the most important issue is to have high accuracy and repeatability but the method is uncomfortable and integrity insulting, the frequency of temperature measurements should be reduced as much as possible.

The present systematic review is with 37 studies and 5026 study participants the largest summary of the evidence for temperature measurements at the temporal artery. Its strength is that the sensitivity analysis did not change the overall result notably. A weakness is the large heterogeneity among included studies.

Temperature measurements with TAT have been evaluated in a health technology assessment report from Scotland⁵⁵ where it was considered as not exact enough when compared with a reference standard. A recent meta-analysis by Niven *et al*⁵⁶ came to the same conclusion; they, however, included only 12 articles. When comparing with tympanic measurements, the results point in various directions. Barnason *et al*⁵⁷ show evidence supporting the use in non-febrile adults and children 3 years and older, with clearer evidence supporting oral temperature measurements. Other reviews found no evidence supporting the use of TAT.^{58 59} Tympanic thermometer measurements in children have been evaluated in a systematic review and meta-analysis by Zhen *et al*.⁶ A pooled difference of 0.22°C (95% LoA −0.44 to 1.30°C) was found compared with reference. They concluded that tympanic measurements cannot replace

rectal temperature measurements in these patients. Tympanic measurements have been reported as acceptable in critically ill patients in a systematic review by Jefferies *et al*,⁶⁰ but had low sensitivity and high specificity in other systematic reviews.^{4 61}

Our results indicate that TAT is not sufficiently accurate to replace one of the reference methods such as rectal, bladder or more invasive temperature measurement methods. Although inaccurate, the results are similar to those with tympanic thermometers, both in our meta-analysis and when compared with others. Thus, it seems that TAT could replace tympanic thermometers with the caveat that both methods are inaccurate. It is unlikely that further research would alter these conclusions. However, there is a need to find a refined non-invasive thermometer with high accuracy.

Acknowledgements The authors thank Margareta Landin at the Medical Library at Örebro University who performed the literature search. Lars Hagberg, PhD, performed the health economic analysis. Ronny Carlsson assisted with technical information. Mia Svantesson-Sandberg, PhD, performed the ethical analysis. Monica Hultcrantz, Agneta Pettersson and Pernilla Östlund from the Swedish Council on Health Technology Assessment (SBU) participated in the assessment of methodological quality and in rating the quality of evidence.

Contributors HG performed the literature search together with a librarian. HG, GL and YN selected articles and assessed the methodological quality. RU performed the statistical analysis. This report was mainly written by HG and was critically reviewed, revised and subsequently approved by all authors.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Hebbar K, Fortenberry JD, Rogers K, *et al*. Comparison of temporal artery thermometer to standard temperature measurements in pediatric intensive care unit patients. *Pediatr Crit Care Med* 2005;6:557–61.

2. Kimberger O, Cohen D, Illievich U, *et al.* Temporal artery versus bladder thermometry during perioperative and intensive care unit monitoring. *Anesth Analg* 2007;105:1042–7.
3. Ng KPM, FR, Heng BH. *An evaluation of non-invasive temperature measurement for inpatients: a review of the literature.* Health Technology Assessment. Singapore: National Healthcare Group, 2007:1–15.
4. Dodd SR, Lancaster GA, Craig JV, *et al.* In a systematic review, infrared ear thermometry for fever diagnosis in children finds poor sensitivity. *J Clin Epidemiol* 2006;59:354–7.
5. Huggins R, Glaviano N, Negishi N, *et al.* Comparison of rectal and aural core body temperature thermometry in hyperthermic, exercising individuals: a meta-analysis. *J Athl Train* 2012;47:329–38.
6. Zhen C, Xia Z, Long L, *et al.* Accuracy of infrared ear thermometry in children: a meta-analysis and systematic review. *Clin Pediatr (Phila)* 2014;53:1158–65.
7. Higgins J, Green S, eds. *Cochrane handbook for systematic reviews of interventions version 5.1.0.* The Cochrane Collaboration, 2011.
8. Whiting PF, Rutjes AW, Westwood ME, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
9. Williamson PR, Lancaster GA, Craig JV, *et al.* Meta-analysis of method comparison studies. *Stat Med* 2002;21:2013–25.
10. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177–88.
11. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135–60.
12. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
13. Reitsma JB, Glas AS, Rutjes AW, *et al.* Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–90.
14. Dwamena BA. midas: A program for meta-analytical integration of diagnostic accuracy studies in Stata. Division of Nuclear Medicine, Department of Radiology, University of Michigan Medical School, Ann Arbor, Michigan. 2007.
15. Harbord RM, Whiting P. metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata J* 2009;9:211–29.
16. Macaskill P, Gatsonis C, Deeks JJ, *et al.* Chapter 10: analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0.* The Cochrane Collaboration, 2010. <http://srda.cochrane.org/>
17. Guyatt GH, Oxman AD, Kunz R, *et al.* What is “quality of evidence” and why is it important to clinicians? *BMJ* 2008;336:995–8.
18. Allegaert K, Casteels K, van Gorp I, *et al.* Tympanic, infrared skin, and temporal artery scan thermometers compared with rectal measurement in children: a real-life assessment. *Curr Ther Res Clin Exp* 2014;76:34–8.
19. Al-Mukhaizeem F, Allen U, Komar L, *et al.* Comparison of temporal artery, rectal and esophageal core temperatures in children: Results of a pilot study. *Paediatr Child Health* 2004;9:461–5.
20. Bahorski J, Repasky T, Ranner D, *et al.* Temperature measurement in pediatrics: a comparison of the rectal method versus the temporal artery method. *J Pediatr Nurs* 2012;27:243–7.
21. Batra P, Goyal S. Comparison of rectal, axillary, tympanic, and temporal artery thermometry in the pediatric emergency room. *Pediatr Emerg Care* 2013;29:63–6.
22. Callanan D. Detecting fever in young infants: reliability of perceived, pacifier, and temporal artery temperatures in infants younger than 3 months of age. *Pediatr Emerg Care* 2003;19:240–3.
23. Calonder EM, Sendelbach S, Hodges JS, *et al.* Temperature measurement in patients undergoing colorectal surgery and gynecology surgery: a comparison of esophageal core, temporal artery, and oral methods. *J Perianesth Nurs* 2010;25:71–8.
24. Carr EA, Wilmoth ML, Eliades AB, *et al.* Comparison of temporal artery to rectal temperature measurements in children up to 24 months. *J Pediatr Nurs* 2011;26:179–85.
25. Drake-Brockman TF, Hegarty M, Chambers NA, *et al.* Monitoring temperature in children undergoing anaesthesia: a comparison of methods. *Anaesth Intensive Care* 2014;42:315–20.
26. Dybwik K, Nielsen EW. [Infrared temporal thermometry]. *Tidsskr Nor Laegeforen* 2003;123:3025–6.
27. Furlong D, Carroll DL, Finn C, *et al.* Comparison of temporal to pulmonary artery temperature in febrile patients. *Dimens Crit Care Nurs* 2015;34:47–52.
28. Greenes DS, Fleisher GR. Accuracy of a noninvasive temporal artery thermometer for use in infants. *Arch Pediatr Adolesc Med* 2001;155:376–81.
29. Greenes DS, Fleisher GR. When body temperature changes, does rectal temperature lag? *J Pediatr* 2004;144:824–6.
30. Gunawan M, Soetjningsih I, Kardana M. Comparison of the accuracy of body temperature measurements with temporal artery thermometer and axillary mercury thermometer in term newborns. *Paediatr Indones* 2010;50:67–72.
31. Hamilton PA, Marcos LS, Secic M. Performance of infrared ear and forehead thermometers: a comparative study in 205 febrile and afebrile children. *J Clin Nurs* 2013;22:2509–18.
32. Holzhauser JK, Reith V, Sawin KJ, *et al.* Evaluation of temporal artery thermometry in children 3–36 months old. *J Spec Pediatr Nurs* 2009;14:239–44.
33. Kirk D, Rainey T, Vail A, *et al.* Infra-red thermometry: the reliability of tympanic and temporal artery readings for predicting brain temperature after severe traumatic brain injury. *Crit Care* 2009;13:F81.
34. Langham GE, Maheshwari A, Contrera K, *et al.* Noninvasive temperature monitoring in postanesthesia care units. *Anesthesiology* 2009;111:90–6.
35. Lawson L, Bridges EJ, Ballou I, *et al.* Accuracy and precision of noninvasive temperature measurement in adult intensive care patients. *Am J Crit Care* 2007;16:485–96.
36. Lee G, Flannery-Bergey D, Randall-Rollins K, *et al.* Accuracy of temporal artery thermometry in neonatal intensive care infants. *Adv Neonatal Care* 2011;11:62–70.
37. Mangat J, Standley T, Prevost A, *et al.* A comparison of technologies used for estimation of body temperature. *Physiol Meas* 2010;31:1105–18.
38. Moore AH, Carrigan JD, Solomon DM, *et al.* Temporal artery thermometry to detect pediatric fever. *Clin Nurs Res* 2015;24:556–63.
39. Myny D, De Waele J, Defloor T, *et al.* Temporal scanner thermometry: a new method of core temperature estimation in ICU patients. *Scott Med J* 2005;50:15–18.
40. Nimah MM, Bshesh K, Callahan JD, *et al.* Infrared tympanic thermometry in comparison with other temperature measurement techniques in febrile children. *Pediatr Crit Care Med* 2006;7:48–55.
41. Odinaka KK, Edelu BO, Nwolisa CE, *et al.* Temporal artery thermometry in children younger than 5 years: a comparison with rectal thermometry. *Pediatr Emerg Care* 2014;30:867–70.
42. Penning C, van der Linden JH, Tibboel D, *et al.* Is the temporal artery thermometer a reliable instrument for detecting fever in children? *J Clin Nurs* 2011;20:1632–9.
43. Reynolds M, Bonham L, Gueck M, *et al.* Are temporal artery temperatures accurate enough to replace rectal temperature measurement in pediatric ED patients? *J Emerg Nurs* 2014;40:46–50.
44. Rubia-Rubia J, Arias A, Sierra A, *et al.* Measurement of body temperature in adult patients: Comparative study of accuracy, reliability and validity of different devices. *Int J Nurs Stud* 2011;48:872–80.
45. Sahin SH, Duran R, Sut N, *et al.* Comparison of temporal artery, nasopharyngeal, and axillary temperature measurement during anesthesia in children. *J Clin Anesth* 2012;24:647–51.
46. Schuh S, Komar L, Stephens D, *et al.* Comparison of the temporal artery and rectal thermometry in children in the emergency department. *Pediatr Emerg Care* 2004;20:736–41.
47. Siberry GK, Diener-West M, Schappell E, *et al.* Comparison of temple temperatures with rectal temperatures in children under two years of age. *Clin Pediatr (Phila)* 2002;41:405–14.
48. Singler K, Bertsch T, Heppner HJ, *et al.* Diagnostic accuracy of three different methods of temperature measurement in acutely ill geriatric patients. *Age Ageing* 2013;42:740–6.
49. Stelfox HT, Straus SE, Ghali WA, *et al.* Temporal artery versus bladder thermometry during adult medical-surgical intensive care monitoring: an observational study. *BMC Anesthesiol* 2010;10:13.
50. Suleman MI, Doufas AG, Akça O, *et al.* Insufficiency in a new temporal-artery thermometer for adult and pediatric patients. *Anesth Analg* 2002;95:67–71.
51. Teran CG, Torrez-Llanos J, Teran-Miranda TE, *et al.* Clinical accuracy of a non-contact infrared skin thermometer in paediatric practice. *Child Care Health Dev* 2012;38:471–6.
52. Winslow EH, Cooper SK, Haws DM, *et al.* Unplanned perioperative hypothermia and agreement between oral, temporal artery, and bladder temperatures in adult major surgery patients. *J Perianesth Nurs* 2012;27:165–80.
53. El-Radhi AS, Barry W. Thermometry in paediatric practice. *Arch Dis Child* 2006;91:351–6.
54. Maxton FJ, Justin L, Gillies D. Estimating core temperature in infants and children after cardiac surgery: a comparison of six methods. *J Adv Nurs* 2004;45:214–22.
55. Thompson L. What is the clinical and cost effectiveness of temporal artery thermometers compared with infrared in-ear



- thermometers for temperature measurement in routine clinical practice? Glasgow: Healthcare Improvement Scotland. Technologies scoping report 11. 2012.
56. 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.
 57. Barnason S, Williams J, Proehl J, *et al.* Emergency nursing resource: non-invasive temperature measurement in the emergency department. *J Emerg Nurs* 2012;38:523–30.
 58. Bahr SJ, Senica A, Gingras L, *et al.* Clinical nurse specialist-led evaluation of temporal artery thermometers in acute care. *Clin Nurse Spec* 2010;24:238–44.
 59. Hooper VD, Andrews JO. Accuracy of noninvasive core temperature measurement in acutely ill adults: the state of the science. *Biol Res Nurs* 2006;8:24–34.
 60. Jefferies S, Weatherall M, Young P, *et al.* A systematic review of the accuracy of peripheral thermometry in estimating core temperatures among febrile critically ill patients. *Crit Care Resusc* 2011;13:194–9.
 61. Holte TO, Vandvik PO, Elvsaas IKÖ, *et al.* *Diagnostisk nøyaktighet av øre-, munnhule-, armhule- og pannetermometer sammenliknet med rektaltermometer for å identifisere feber hos voksne pasienter innlagt i sykehus eller sykehjem. Rapport fra Kunnskapssenteret.* Oslo: Nasjonalt kunnskapssenter for helsetjenesten, 2009:1–49.