

A systematic review of the accuracy of peripheral thermometry in estimating core temperatures among febrile critically ill patients

Sarah Jefferies, Mark Weatherall,
Paul Young and Richard Beasley

Fever is common among critically ill patients, and its detection has important implications for patient management.¹ Fever occurs as a response to tissue injury and inflammation. The detection of a new fever may herald the onset of an infection, often leading to additional investigations and changes in patient management. The presence of persistent fever may indicate treatment failure, progressive disease or disordered thermoregulation, and may be associated with poor prognosis.² Accurate determination of body temperature in the febrile range is not only important in influencing intensive care unit practice, but also in the setting of clinical research, where inclusion criteria or interventions may be set at specific temperature thresholds.

Direct measurement of core temperature is regarded as the most accurate method to determine body temperature, as it is less influenced by variables such as ambient temperature and peripheral perfusion. Core temperature is best represented by the pulmonary artery catheter (PAC) thermometer, the "gold standard" for clinical thermometry, although other methods of core temperature assessment, such as with oesophageal and bladder devices, are also considered accurate.³ However, devices used to measure core temperature are generally more expensive and more invasive than peripheral thermometry, and may require a skilled operator to position.

Over the past four decades, clinical thermometry has developed to enable rapid, convenient core temperature estimation using electronic devices at peripheral sites.⁴ These electronic thermometers often have functions that convert temperatures measured at one site of the body to an estimate of the temperature at another site (eg, core, oral, rectal). Conversion algorithms for these functions vary between thermometers and are determined by the manufacturer. For all modes other than "unadjusted" (or "equal") mode, a fixed number is automatically added to the temperature taken. Clinical studies investigating the accuracy of peripheral thermometers in determining core temperature have reported differing accuracy at different core temperatures,^{5,6} and recommendations have been made for further investigation to take into account potential confounders, including temperature range variables.⁷

ABSTRACT

Background: There is uncertainty about the accuracy of peripheral thermometers in measuring temperatures within the febrile physiological range.

Objective: To determine the accuracy of peripheral thermometers in detecting febrile core temperatures among critically ill patients, and, if required, to determine a standard conversion equation to improve accuracy.

Methods: A systematic search of MEDLINE, Embase, the Cochrane Central Register of Controlled Trials and PubMed was undertaken to identify clinical trials comparing peripheral thermometry in critically ill adult patients with core temperatures $> 37.5^{\circ}\text{C}$. Our prespecified plan was to perform a meta-analysis of the clinical accuracy of mean peripheral thermometer temperature difference from core temperature and calculation of limits of agreement.

Results: Systematic review identified three studies that compared infrared tympanic, rectal or oral thermometer readings with pulmonary artery catheter core temperature readings among critically ill adults with fever. Studies were heterogeneous and all failed to report appropriate measurements of variation for the estimates of clinical accuracy, which prevented meta-analysis and limited interpretation of the results. Mean differences were within $\pm 0.2^{\circ}\text{C}$ in five of seven tympanic thermometer/mode/temperature combinations and in the one oral thermometer studied. All of three rectal thermometer/temperature combinations studied reported mean differences outside this range.

Conclusion: The identified studies suggest that in critically ill patients, tympanic and oral thermometry provide, on average, accurate measures of core temperatures within the febrile range and can be recommended for this purpose. Further studies with appropriate statistical methods are required to assess the accuracy of peripheral thermometers among critically ill patients with fever.

Crit Care Resusc 2011; 13: 194–199

In view of the uncertainty regarding the accuracy of peripheral thermometry at elevated core temperatures, and the importance of detecting fever in the ICU, we undertook a systematic review with the intention of performing a meta-analysis. We hypothesised that peripheral thermometry would tend to underestimate febrile core temperatures, and that meta-analysis would determine a standard conversion to calculate core temperature based on peripheral temperature measurements.

Methods

Search strategy

Four databases were used to identify clinical studies comparing measurements from peripheral thermometers with those obtained simultaneously from core thermometers in the critically ill. The databases used were MEDLINE (1950 to present); Embase (1947 to present); the Cochrane Central Register of Controlled Trials (1991 to present) and PubMed (1950 to present): search date, 29 December 2010. Keywords were: "thermometer*"; and "pulmonary artery" or "bladder"; limited to human and clinical trials. Potentially relevant articles that were not written in English were translated. One of us (S.J.) examined each article's title and abstract and the full article if necessary. The reference lists of all relevant articles were also reviewed, and additional hand searching was carried out.

Inclusion criteria

Clinical trials that investigated the accuracy of peripheral methods of temperature measurement compared with pulmonary artery catheter or bladder thermometry among adult patients with core temperatures $>37.5^{\circ}\text{C}$, and presented appropriate summary statistics, were included. For the purpose of this systematic review, peripheral thermometry included any method of thermometry other than pulmonary artery, oesophageal or bladder thermometry.

Exclusion criteria

- Non-human studies.
- Studies that involved iatrogenic physical temperature manipulation (eg, external warming or cooling), as this may introduce thermal gradients that are distinct from physiological fever.
- Studies with no data for febrile core temperature ranges.
- Studies that only presented data in graphical form, so that use of data for analysis would have required estimation of data points.
- Studies that examined temporal artery thermometry.

Data extraction and interpretation

Extraction of data was based on reported summary statistics. These were the mean differences, reflecting bias, and

appropriate measurements of variance. In the original publication by Bland and Altman, limits of agreement were defined as plus or minus two times the standard deviation of the difference in measurements of the same research participants measured twice only, reflecting that about 95% of the difference between future paired measurements would fall between the limits of agreement.⁸

We recognised that research designs in this area often involve measuring the same patients' temperatures on multiple occasions, often with multiple different peripheral thermometers. We therefore reviewed publications for reports of appropriate measures of variance for the mean difference that could be used to calculate approximate 95% confidence intervals for an individual predicted value. In particular, we attempted to find the reports of variance components due to measurement error that took into account multiple measurements on individual research participants. We were planning on performing a meta-analysis of the size of the mean bias and attempted to find reports of variance measurements, or appropriately calculated confidence intervals, appropriate to the mean difference.

Based on previous research, we were particularly interested in clinical accuracy, defined as a mean peripheral thermometer temperature difference from a febrile core temperature within $\pm 0.2^{\circ}\text{C}$, and limits of agreement within $\pm 0.5^{\circ}\text{C}$.⁹⁻¹² We had planned to perform an inverse variance weighted meta-analysis in critically ill patients with febrile core temperatures for clinical accuracy and limits of agreement.

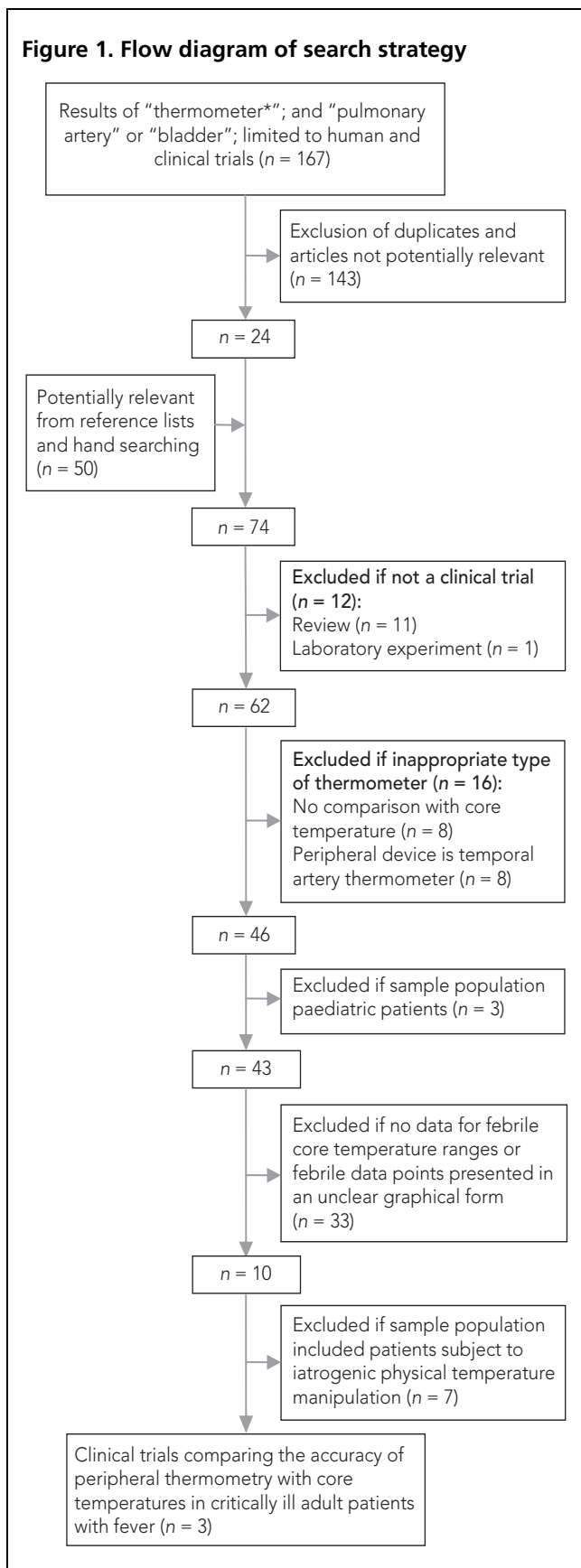
Results

Systematic review identified three studies reporting the accuracy of peripheral thermometers in detecting febrile core temperatures in critically ill patients (Figure 1).^{10,13,14} All three studies included multiple measurements, often in a poorly specified number of febrile patients, and the statistical methods either failed to account for repeated measures on the same participants or did not report appropriate measures of variation. As a result, we were unable to conduct a meta-analysis.

Characteristics of included studies

All three included articles were prospective, non-experimental, observational studies in adult critical care patients, undertaken over a decade ago (Table 1). In the three studies, details of baseline characteristics of the febrile patients, such as sex and diagnosis, were not provided. Details of the number of febrile patients studied or the number of measurements taken in each patient were also lacking.

All articles used PAC as the method of core temperature measurement. The type of peripheral electronic thermo-

Figure 1. Flow diagram of search strategy

meter and modes selected varied between studies. Milewski and colleagues studied both an infrared (IR) tympanic thermometer in its unadjusted mode and a digital electronic rectal thermometer.¹³ Rotello and colleagues investigated three types of IR tympanic thermometer (one in adjusted-to-oral mode, two in unadjusted mode) and one digital electronic rectal thermometer.¹⁴ Giuliano and colleagues investigated two types of IR tympanic thermometer (both in the core mode setting) and one oral electronic thermometer.¹⁰ In this study, most patients were intubated and oral temperatures were not collected within 30 minutes of a patient receiving mouth cares.¹⁰

In one study, the researcher operating the thermometers was blinded.¹⁴ In two studies, more than one operator was involved, with attempts made to evaluate interoperator error.^{10,13} All study protocols provided operator device training, described adequate device calibration and accounted for "draw down", in which false reductions in tympanic temperature occur when repeated readings are taken from the same ear within 2 minutes.

Two of the three studies did not describe the exclusion of ear pathological features, which can affect the accuracy of tympanic thermometry;^{10,13} and none described device cleaning, replacement of IR thermometer probe caps with each reading or the control of other variables, such as recent warming of the ear by head position on pillow.^{9,15} The use of the "ear tug" technique, which may influence tympanic measurements, varied in the three studies.

Clinical accuracy data

Table 2 summarises the mean differences reported in the included studies which compared tympanic, oral and rectal thermometry with core temperature measurements. Meta-analysis could not be performed, as all studies reported error data without accounting for the effects of repeated measures due to multiple measurements among study patients.

An assessment of clinical accuracy could therefore only be made based on the reported mean differences. Five of seven different tympanic thermometer/mode/core temperature range combinations among the studies were clinically accurate, with a mean difference within $\pm 0.2^\circ\text{C}$ of core febrile temperatures (Table 2). The two tympanic thermometer/mode/core temperature range combinations that exceeded this limit were the Thermoscan Pro-1 in unadjusted mode for core temperatures $37.6\text{--}38.0^\circ\text{C}$ (mean difference, -0.22°C from core), although clinical accuracy was demonstrated in this device/mode at $>38.0^\circ\text{C}$, and the Thermoscan HM-1 in oral mode (mean difference from core 0.24°C). Oral thermometry was used in one study and was clinically accurate with a mean difference from core of

Table 1. Characteristics of the included studies

Study	Parti- pants; no.	Peripheral thermometer (mode)	Protocol	No. of operators (blinding)	Operator training	Use of ear tug technique	Avoid draw- down	Otoscopic exclusion of ear pathology	Calibration of devices	Evaluation of other variables
Milewski et al, 1991 ¹³	Adult ICU; febrile <i>n</i> =? total <i>n</i> =9	Tympanic TM1 (unadjusted); rectal RT1 (monitor/non-predictive)	1–2-hourly single measurements for up to 48 h; RT read at 3 min from insertion	> 1 (not blinded)	Y	R ear, no tug	Y	N	Y	Individual temperature trends
Rotello et al, 1996 ¹⁴	Adult ICU; febrile <i>n</i> =9, total <i>n</i> =20	Tympanic TM2 (oral); tympanic TM3 (unadjusted); tympanic TM4 (unadjusted); rectal RT2 (monitor/non-predictive)	3 readings with each TM at 2 min intervals and a RT2 reading with the third TM, the RT2 having been placed for 6 min	1 (single blinding)	Y	R ear, with tug	Y	Y	Y	Clinical repeatability between the TM devices
Giuliano et al, 2000 ¹⁰	Adult ICU; febrile <i>n</i> =? total <i>n</i> =72	TM5 (core); TM6 (core); oral OT1 (not specified)	1 reading per device (one TM per ear) over 1 min. Repeated up to 3 times at least 20 min apart	3 (not blinded)	Y	Alternating ears, ear tug use not specified	Y	N	Y	Clinician–thermometer interaction

ICU = intensive care unit. TM = tympanic thermometer. RT = rectal thermometer. OT = oral thermometer. TM1 = ThermoScan Pro-1 (Thermoscan Inc, San Diego, Calif, USA). TM2 = ThermoScan HM-1 (Thermoscan Inc). TM3 = ThermoScan Pro-1 (Thermoscan Inc). TM4 = ThermoScan Pro-LT (Thermoscan Inc). TM5 = FirstTemp Genius II (model 3000A, Sherwood Medical, Carlsbad, Calif, USA). TM6 = ThermoScan Ear Pro-1 (model IR-1, Thermoscan Inc). RT1 = Electronic rectal thermometer (Diatek, San Diego, Calif, USA). RT2 = electronic rectal thermometer (Diatek). OT1 = SureTemp oral thermometer (model 678, Welch Allyn, San Diego, Calif, USA).

0.18° C.¹⁰ Rectal thermometry demonstrated clinical inaccuracy, overestimating core temperature measured by PAC (Table 2).

Discussion

Our systematic review suggests that for critically ill patients, on average, tympanic and oral thermometry provide accurate estimations of core temperatures within the febrile range, but that rectal thermometry is clinically inaccurate. However, because of limitations in study reports, we were unable to present a pooled estimate of clinical accuracy, limits of agreement or appropriate confidence limits.

To the best of our knowledge, no other systematic review has evaluated the accuracy of peripheral thermometry in the febrile critically ill adult patient population. Hooper and colleagues reviewed the accuracy of peripheral thermometry among adult critical care patients, including several studies that did not meet the criteria for our review.⁷ They found variable reports of accuracy among a highly heterogeneous set of studies, and concluded that there was a requirement for further investigation to account for differences in accuracy across sex, ethnicity, age and temperature range variables.⁷ In another adult critical medicine review, O’Grady and colleagues concluded that the accuracy of tympanic thermometry was “consistently poor” and ranked their recommendations for peripheral thermometry in the order of rectal, oral, and lastly tympanic.¹⁶ They did not

discuss the flaws in statistical methods or reporting, nor the heterogeneity of the studies upon which they based this advice.

Two studies that did not meet the criteria for our review further demonstrate the variability in the existing evidence. Schmitz and colleagues compared peripheral with core methods of thermometry specifically among critically ill patients who were febrile at baseline.⁶ However, as their analysis incorporated readings taken while the patients were afebrile, their research was excluded from our analysis. They reported that tympanic thermometry overestimated core temperature at PAC temperatures < 38.3° C and underestimated at PAC temperatures ≥ 38.3° C. This suggests that the comparable performance of tympanic thermometry may vary across the physiological febrile range.

In a more recent study by Moran and colleagues, the average temperature difference between PAC and tympanic thermometry was –0.36 (SD, 0.47), with reported limits of agreement of 0.56 and –1.28° C, representing a greater bias and variability from core temperature than that found with axillary thermometry.¹⁷ In this article, the results were derived from patients who experienced a range of core temperatures, from hypothermia to hyperthermia, and therefore it did not meet the criteria for our analysis. However, the authors further report that regression analyses examining temperature differences between PAC and tympanic, and PAC and axillary measurements, across the range of patient temperatures, were not statistically significant.¹⁷

Table 2. Summary data presented by the included studies of mean differences from PAC for febrile core temperature ranges in critically ill patients

Study	Febrile range(s), °C	Mean difference from PAC, °C* (n) [†]		
		Tympanic	Oral	Rectal
Milewski et al, 1991 ¹³	37.6–38.0	TM1 -0.22 (n = 41)	—	RT1 0.46 (n = 35)
	38.1–40.0	TM1 -0.01 (n = 40)	—	RT1 0.53 (n = 36)
Rotello et al, 1996 ¹⁴	37.6–39.1	TM2 0.24 (n = 27)	—	RT2 0.35 (n = 27)
		TM3 0.14 (n = 27)	—	—
		TM4 0.01 (n = 27)	—	—
Giuliano et al, 2000 ¹⁰	38.1–39.3	TM5 -0.17 (n = 50)	OT1 0.18 (n = 48)	—
		TM6 -0.05 (n = 50)	—	—

PAC = pulmonary artery catheter. TM = tympanic thermometer. RT = rectal thermometer. OT = oral thermometer. TM1 = ThermoScan Pro-1 (Thermoscan Inc, San Diego, Calif, USA). TM2 = ThermoScan HM-1 (Thermoscan Inc). TM3 = ThermoScan Pro-1 (Thermoscan Inc). TM4 = ThermoScan Pro-LT (Thermoscan Inc). TM5 = FirstTemp Genius II (model 3000A, Sherwood Medical, Carlsbad, Calif, USA). TM6 = ThermoScan Ear Pro-1 (model IR-1, Thermoscan Inc). RT1 = electronic predictive (Diatek, San Diego, Calif, USA). RT2 = Electronic predictive (Diatek). OT1 = SureTemp (model 678, Welch Allyn, San Diego, Calif, USA). * Peripheral thermometer mean –core thermometer mean; negative numbers therefore represent a mean underestimate of core temperatures. † No. of measurements.

This suggests that the bias between PAC and these two thermometry measurements does not change in the febrile range. When we considered the reported limits of agreement by inspection of the Bland–Altman plots for tympanic or axillary thermometry versus PAC, there appeared to be greater variability in differences at mean temperatures <38°C. However, this is difficult to verify without analysis of the febrile-range participants.

Inspection of the plots also strongly suggested that the limits of agreement calculated by the authors were based on the standard deviation of all the paired readings and do not take into account the multiple readings on the same participants, and so are likely to be too narrow. This is less likely to have a significant effect on the calculated mean bias. Moran and colleagues concluded that urinary thermometry was superior in accuracy to tympanic thermometry and questioned the use of tympanic thermometers in critically ill patients.¹⁷

This review is significant for several reasons. Firstly, it summarises the limited available data comparing peripheral and core temperatures among febrile critically ill patients, but indicates that tympanic and oral thermometry may be used to accurately detect febrile core temperatures among these patients. Secondly, it highlights the lack of published data on the performance of some of the brands of peripheral thermometers currently used in clinical practice. Thirdly, and perhaps most importantly, the analysis highlights a major statistical flaw in the existing published literature, discussed further below.

This systematic review was performed with a sensitive search of four major databases, with further hand searches and review of reference lists. We are confident that all

relevant articles were identified by this strategy. However, the studies included in the systematic review were heterogeneous by design and did not use statistical analysis to adequately account for variation due to repeated measures.

The original description of the Bland–Altman method assumes that each research participant in an agreement study is measured once with each measurement device.⁸ The purpose of the method is to derive approximately 95% confidence intervals for a future individual predicted value of bias. For the simple design where participants are only measured twice, this is based on plus or minus two times the standard deviation of the differences and the two methods of measurement agree if the limits of agreement are acceptable on subject-matter grounds. For comparisons between measurements of temperature, $\pm 0.5^\circ\text{C}$ has been proposed as acceptable limits of agreement.^{10–12}

The situation is more complex when participants in a study of agreement are measured more than twice. This is because there are two or more sources of variation to consider: between patients, between measurement instruments, if more than one is used, and the “leftover variation”. The leftover variation is the measurement error that is appropriate to assess agreement. In the Bland–Altman procedure, with only two measurements per participant, the between-patient component is automatically accounted for by working with the differences only. The standard method of accounting for repeated, more than two, measurements on the same patients are mixed linear models, of which repeated-measures analysis of variance is a simple example. These techniques use a variety of methods, of which the most used is restricted maximum likelihood, to estimate the variation due to measurement error. Failure to

account for repeatedly measuring the same research participants, or that different research participants are measured a different number of times, will give inappropriate estimates of variation (eg, standard deviations or standard error of the mean, and confidence intervals) for differences between different measuring techniques. This difference is likely to mean that confidence intervals are inappropriately narrow. Simply taking the standard deviation of all the differences for repeated measurements does not appropriately estimate the correct element of variation to judge agreement.

It is likely that the estimates of the mean values of bias, shown in Table 2, are close to those obtained from analysis that properly accounts for components of variation, if each participant's temperature was measured about the same number of times. However, if some participants' temperatures were measured far more than those of others, these participants are inappropriately weighted in calculating the mean and the mean difference may itself be biased.

We recommend that further studies with appropriate statistical methods be conducted to properly assess the accuracy of peripheral thermometers currently being used in critically ill adult patients with fever. Such studies need to account for the components of variation to estimate measurement error. This will require analysis that explicitly accounts for repeated measurements together with appropriate estimates of bias and precision derived from, for example, mixed linear models, with appropriate specification of random and fixed effects. For the purposes of ongoing clinical practice and trials using oral, tympanic or rectal thermometry, we advise that, on the basis of the limited available evidence, tympanic and oral thermometry methods should be regarded as equivalent to core temperature and that rectal thermometry should not be used.

Competing interests

None declared.

Author details

Sarah Jefferies, Medical Research Fellow,¹ and Medical Registrar²

Mark Weatherall, Physician,² and Associate Professor³

Paul Young, Intensivist²

Richard Beasley, Director,¹ and Physician²

1 Medical Research Institute of New Zealand, Wellington, New Zealand.

2 Capital and Coast District Health Board, Wellington, New Zealand.

3 University of Otago Wellington, Wellington, New Zealand.

Correspondence: Sarah.Jefferie s@mri.nz.ac.nz

References

- 1 Laupland KB, Shahpori R, Kirkpatrick AW, et al. Occurrence and outcome of fever in critically ill adults. *Crit Care Med* 2008; 36: 1531-5.
- 2 Circiumaru B, Baldock G, Cohen J. A prospective study of fever in the intensive care unit. *Intensive Care Med* 1999; 25: 668-73.
- 3 Lefrant JY, Muller L, de La Coussaye JE, et al. Temperature measurement in intensive care patients: comparison of urinary bladder, oesophageal, rectal, axillary, and inguinal methods versus pulmonary artery core method. *Intensive Care Med* 2003; 29: 414-8.
- 4 Crawford DC, Hicks B, Thompson MJ. Which thermometer? Factors influencing best choice for intermittent clinical temperature assessment. *J Med Eng Technol* 2006; 30: 199-211.
- 5 Erickson RS, Kirklin SK. Comparison of ear-based, bladder, oral, and axillary methods for core temperature measurement. *Crit Care Med* 1993; 21: 1528-34.
- 6 Schmitz T, Bair N, Falk M, Levine C. A comparison of five methods of temperature measurement in febrile intensive care patients. *Am J Crit Care* 1995; 4: 286-92.
- 7 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.
- 8 Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet* 1995; 346: 1085-7.
- 9 Erickson RS, Meyer LT. Accuracy of infrared ear thermometry and other temperature methods in adults. *Am J Crit Care* 1994; 3: 40-54.
- 10 Giuliano KK, Giuliano AJ, Scott SS, et al. Temperature measurement in critically ill adults: a comparison of tympanic and oral methods. *Am J Crit Care* 2000; 9: 254-61.
- 11 Cattaneo CG, Frank SM, Hesel TW, et al. The accuracy and precision of body temperature monitoring methods during regional and general anesthesia. *Anesth Analg* 2000; 90: 938-45.
- 12 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.
- 13 Milewski A, Ferguson KL, Terndrup TE. Comparison of pulmonary artery, rectal, and tympanic membrane temperatures in adult intensive care unit patients. *Clin Pediatr (Phila)* 1991; 30: 13-6; discussion 34-5.
- 14 Rotello LC, Crawford L, Terndrup TE. Comparison of infrared ear thermometer derived and equilibrated rectal temperatures in estimating pulmonary artery temperatures. *Crit Care Med* 1996; 24: 1501-6.
- 15 Davie A, Amoore J. Best practice in the measurement of body temperature. *Nurs Stand* 24: 42-9.
- 16 O'Grady NP, Barie PS, Bartlett JG, et al. Guidelines for evaluation of new fever in critically ill adult patients: 2008 update from the American College of Critical Care Medicine and the Infectious Diseases Society of America. *Crit Care Med* 2008; 36: 1330-49.
- 17 Moran JL, Peter JV, Solomon PJ, et al. Tympanic temperature measurements: are they reliable in the critically ill? A clinical study of measures of agreement. *Crit Care Med* 2007; 35: 155-64. □