Accuracy and precision of zero-heat-flux temperature measurements with the 3MTM Bair HuggerTM Temperature Monitoring System

A systematic review and meta-analysis

Aaron Conway PhD1,2,3,✉, Megan Bittner BSc2, Dan Phan BSc2, Kristina Chang MScN1, Navpreet Kamboj BScN2, Elizabeth Tipton PhD4, and Matteo Parotto MD, PhD5,6

1 Peter Munk Cardiac Centre, University Health Network, Toronto, Canada  
2 Lawrence S. Bloomberg Faculty of Nursing, University of Toronto, Toronto, Canada  
3 School of Nursing, Queensland University of Technology (QUT), Brisbane, Australia  
4 Department of Statistics, Northwestern University, Evanston, IL, United States  
5 Department of Anesthesia and Pain Management, Toronto General Hospital, UHN, Toronto, Canada  
6 Department of Anesthesia and Interdepartmental Division of Critical Care Medicine, University of Toronto, Toronto, Canada

✉ Correspondence: [Aaron Conway PhD <[aaron.conway@utoronto.ca](mailto:aaron.conway@utoronto.ca)> | +1 416 946 7112 | 585 University Ave | Toronto, ON M5G 2N2](mailto:aaron.conway@utoronto.ca)

### Abstract

**Purpose**

Zero-heat-flux thermometers provide clinicians with the ability to continuously and non-invasively monitor body temperature. These devices are increasingly being used to substitute for more invasive core temperature measurements during surgery and in critical care. The aim of this review was to determine the accuracy and precision of zero-heat-flux temperature measurements from the 3MTM Bair HuggerTM Temperature Monitoring System.

**Methods**

Medline and EMBASE were searched for studies that reported on a measurement of core or peripheral temperature that coincided with a measurement from the zero-heat-flux device. Study selection and quality assessment was performed independently using the Revised Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS-2). The Grading of Recommendations, Assessment, Development and Evaluations (GRADE) approach was used to summarize the strength of the evidence. Pooled estimates of the mean bias and limits of agreement with outer 95% confidence intervals (population limits of agreement) were calculated.

**Results**

Sixteen studies were included. The primary meta-analysis of zero-heat-flux versus core temperature consisted of 22 comparisons from 16 individual studies. Data from 952 participants with 314,137 paired measurements were included. The pooled estimate for the mean bias was 0.03°C. Population limits of agreement, which take into consideration the between-study heterogeneity and sampling error, were wide, spanning from -0.93°C to 0.98°C. The GRADE evidence quality rating was downgraded to moderate due to concerns about study limitations. Population limits of agreement for the sensitivity analysis restricted to studies rated as having low risk of bias across all the domains of the QUADAS-2 were similar to the primary analysis.

**Conclusion**

The range of uncertainty in the accuracy of a thermometer should be taken into account when using this device to inform clinical decision-making. Clinicians should therefore consider the potential that a temperature measurement from a 3MTM Bair HuggerTM Temperature Monitoring System could be as much as 1°C higher or lower than core temperature. Use of this device may not be appropriate in situations where a difference in temperature of less than 1°C is important to detect.

Clinical trial number: Not applicable

Keywords: temperature, measurement, zero-heat-flux, surgery, anesthesia

## Declarations

Availability of data and material (data transparency): All data used in the meta-analyses is available [here](https://zhf-info.netlify.app).

Code availability: All data used in the meta-analyses is available [here](https://github.com/awconway/zhf-review) and archived [here](https://doi.org/10.5281/zenodo.3817610).

Abbreviated title: Accuracy of zero-heat-flux thermometers

Funding statement: Support was provided solely from institutional and/or departmental sources. MP is supported by an Early Investigator Merit Award from the Department of Anesthesia, University of Toronto and Toronto General Hospital

Conflicts of interest: The authors declare no competing interests.

Author contributions: AC: Study design, analysis, wrote manuscript; MB: Study design, data extraction, risk of bias assessment, data analysis, contributed to writing manuscript; DP: Study design, data extraction, risk of bias assessment, revised the manuscript for important intellectual content; NK: Study design, revised the manuscript for important intellectual content; KC: Study design, revised the manuscript for important intellectual content; ET: Study design, data analysis, revised the manuscript for important intellectual content; MP: Study design, revised the manuscript for important intellectual content.

**Glossary of terms:**

GRADE = Grading of Recommendations, Assessment, Development and Evaluations

QUADAS-2 = Revised Quality Assessment of Diagnostic Accuracy Studies tool

LoA = Limits of Agreement

CI = Confidence intervals

SD2 = Variance

τ2 = Tau-squared

## Introduction

Continuous core temperature monitoring is commonly used within perioperative and intensive care settings. Thermoregulatory dysfunction is associated with the induction of anesthesia and can lead to adverse outcomes including cardiac arrhythmia, altered hemostasis and increased risk of surgical site infection.[1–4] The pulmonary artery catheter is the reference standard for continuous core temperature monitoring, however, the invasive nature of this method renders an increased risk of bloodstream infections and damage to surrounding tissue.[5] Although accurate relative to pulmonary artery temperature measurements, surrogate measures of core temperature at esophageal, rectal, and bladder locations remain mildly invasive interventions.

Zero-heat-flux thermometry is an alternative non-invasive method that allows for continuous temperature monitoring from the skin surface. Unlike other commonly used non-invasive thermometers such as infrared aural canal and temporal devices, the zero-heat-flux device does not apply a correction factor. As such, it is classified as an indirect *measurement* of core temperature, not an indirect *estimate*.[6] Zero-heat-flux thermometers consist of a thermal insulator positioned between two thermistors, which are covered by a heating component. The heater is initially warmed to equilibrate with skin surface temperature. By eliminating heat loss to the environment, skin surface temperature closely reflects the temperature at a depth of about 1 to 2 cm below the location of the thermometer.[7] It has been confirmed that zero-heat-flux thermometers can provide an accurate substitute for core temperature monitoring.[8] A comprehensive review of this method of temperature measurement, which was originally developed in the 1970s, is available.[9]

A new generation of zero-heat-flux thermometers are now approved for use within perioperative and intensive care settings. The accuracy of temperature measurements from this latest generation of devices has been investigated in multiple studies over the last five years. The 3MTM Bair HuggerTM Temperature Monitoring System (3MTM, St Paul, MN) was the zero-heat-flux device used in most of these studies. This device was previously known as the SpotOn Temperature Monitoring System. It is a single-use, disposable sensor, which is recommended to be positioned on the lateral surface of the forehead, as this is one of the locations in the body where subcutaneous temperature approximates core temperature. Appraisal and synthesis of the results would aid clinicians in deciding the appropriate circumstances in which this particular zero-heat-flux thermometer may be used. We aimed to determine if the 3MTM Bair HuggerTM Temperature Monitoring System has clinically acceptable accuracy and precision relative to established core and peripheral temperature measurement devices. Accuracy is defined as the average difference between temperature measurements from the zero-heat-flux and comparator device and precision as the variance (standard deviation) in the differences.

## Methods

A systematic review was conducted. The primary comparison for this review was zero-heat-flux temperature measured using the 3MTM Bair HuggerTM Temperature Monitoring System versus temperature measured from a core site, which we defined as temperature measured at either an arterial, esophageal, bladder, rectal, nasopharyngeal or oropharyngeal site.

### Inclusion criteria

Observational studies that reported temperature measurements from a 3MTM Bair HuggerTM Temperature Monitoring System and comparator thermometer were included. Studies involving a case control design were excluded due to potential for overestimation of the intervention performance. Studies were excluded if conducted on non-human subjects or outside of a clinical healthcare setting. No publication date restrictions were applied. Published conference abstracts were included if there was enough information reported to appraise the quality of the study. There were no language restrictions applied during the search.

### Data sources and searches

Published studies were found by searching Medline and EMBASE to January 2020. The Cochrane-recommended search strategy combining terms for the ‘target condition’ and ‘index test’ was used.[10] This search strategy is an efficient approach for systematic reviews of diagnostic test accuracy studies.[11] We also conducted forward citation searching, by using Google Scholar to search the citations of the first article published on the accuracy of the 3MTM Bair HuggerTM Temperature Monitoring System.[7] The search strategies used for each database are available in the appendix. Selection of studies was undertaken independently by two reviewers.

### Data extraction and quality assessment

Information was extracted regarding study characteristics (author, year of publication, country, design, sample size, clinical setting, numbers studied and analyses for each outcome), population characteristics (inclusion and exclusion criteria) and temperature measurement characteristics (placement of sensor, timing and methods of measurements). The outcomes that were extracted included the mean bias (eg, accuracy) and variance (eg, SD, precision) in temperature measurement between the zero-heat-flux and comparator thermometers. We also extracted information about how repeated measurements were handled. In particular we assessed whether studies: (1) analysed each pair of data separately; (2) treated each pair of data as independent; or (3) used either analysis of variance or a random effects model as a way to control for the dependent nature of the repeated measures data.[12]

Two reviewers independently assessed the risk of bias for the included studies using the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2).[13] Reviewers rated the risk of bias for patient selection, conduct of the zero-heat-flux measurements, conduct of the comparator thermometer measurements, and timing and flow (eg, timing of zero-heat-flux and established core temperature measurements, dropouts) as ‘high’, ‘low’ or ‘unclear’ risk of bias. We worked to minimize the risk of publication bias by conducting a comprehensive search of multiple databases as well as an international clinical trial registry.[14] Statistical approaches for detection of reporting bias were not conducted due to lack of validated methods.[15] Simulations have revealed that tests for detecting funnel plot asymmetry will result in publication bias being incorrectly identified too often.[16]

In order to rate the quality of evidence, we applied the Grading Quality of Evidence and Strength of Recommendation methodology.[17] Evidence was downgraded in accordance with study limitations, inconsistency and imprecision. There were no circumstances in which evidence was downgraded for indirectness as this systematic review only included relevant studies. Although the possibility of publication bias was not excluded, this bias was not formally assessed as it was not considered sufficient enough to reason downgrading the quality of evidence.

### Data synthesis and analysis

The objective for the meta-analysis was to estimate the population limits of agreement between temperature measurements from the 3MTM Bair HuggerTM Temperature Monitoring System and established comparator thermometers. A framework for meta-analysis of Bland-Altman method comparison studies based on limits of agreement approach was used.[18] This method was selected because it parallels the approach used in primary Bland-Altman studies, whereby an estimate is generated for the pooled limits of agreement in the population (not just in the samples studied). The ‘population limits of agreement’ is more broad than the limits of agreement commonly reported in the meta-analyses of Bland-Altman studies.[18] Pooled limits of agreement were calculated using , where is the average bias across studies, is the average within-study variation in differences and is the variation in bias across studies.

Estimations of and were made using a weighted least-squares model (similar to a random-effects approach), with the associated standard errors estimated using robust variance estimation. Robust variance estimation was used alternatively to model-based standard errors as some studies included in the systematic review used repeated-measures designs without accommodating for the correlation between measurements.[19–21] We used the method-of-moments estimator from DerSimonian & Laird [22] for the parameter.

Measures of uncertainty were included in our meta-analyses by calculating the outer 95% confidence intervals for pooled limits of agreement. We also accounted for repeated measurements if they were not properly adjusted for in individual studies. This was achieved by using weights proportional to the number of participants, not the total number of measurements. The R statistical program was used to conduct all analyses.[23] All data and R code used in the meta-analyses is available [here](https://github.com/awconway/zhf-review) and archived [here](https://doi.org/10.5281/zenodo.3817610).

Prior to conducting the meta-analyses, the results from each study were converted into a standard format, with bias meaning comparator thermometer temperature measurement minus the 3MTM Bair HuggerTM Temperature Monitoring System measurement in degrees Celsius (°C). In several studies, results were reported for multiple groups of participants, therefore in the meta-analysis each of these groups was treated as a separate ‘comparison’.[24–27] Other studies reported multiple sets of results, whereby analyses were conducted between zero-heat-flux and various comparator devices used on the same participant. These instances were also treated as a separate ‘comparison’ if the comparator devices were a part of separate meta-analysis groups.[27–30] One study used both nasopharyngeal and oropharyngeal temperature as comparators.[28] Both combined and separate estimates were reported. We used the combined estimate in our primary analysis and the estimate for just the nasopharyngeal temperatures in a subgroup analysis. One study reported intraoperative, postoperative and overall results for the same participants[7]. Only the paired measurements from the overall results were included in the main and low risk bias analyses, leaving paired measurements exclusively from the intraoperative and postoperative timepoints to be included in respective meta-analyses subgroups. The conventionally cited clinically acceptable agreement between zero-heat-flux and comparator devices is 0.5°C.[7] It was deemed that outer confidence bounds for 95% limits of agreement between zero-heat-flux and core temperature measurements (termed as ‘population limits of agreement’) outside of these bounds would be clinically unacceptable.

A sensitivity analysis for the primary comparison (zero-heat-flux versus temperature measurement at core site) was performed based on risk of bias, whereby ‘unclear risk of bias’ was treated as ‘high risk’ and ‘high risk of bias’ studies were excluded from analyses. We also conducted a sensitivity analysis for the primary comparison (zero-heat-flux versus temperature measurement at core site) excluding studies that received funding from industry. Probe insertion distance for temperature measurements from the nasopharynx and oropharynx may impact accuracy.[31] For this reason, we also conducted a separate analysis excuding these comparators and an analysis of the studies that compared temperature measurements from the 3MTM Bair HuggerTM Temperature Monitoring System with nasopharyngeal temperature. As clinicians would be interested in the accuracy of zero-heat-flux relative within the clinical setting in which they use it, we conducted subgroup analyses for the primary comparison according to clinical setting of the study (either intraoperative or intensive care unit).

## Results

### Study selection and description

Sixteen studies were included (Figure 1). Two studies reported only in abstract form were not included and assigned as ‘studies awaiting classification’ because there was insufficient information provided.[32, 33]

A summary of the characteristics of included studies is displayed in Table 1. The primary comparison of the 3MTM Bair HuggerTM Temperature Monitoring System versus core temperature measurements consisted of 22 comparisons from 16 individual studies.[7, 24–30, 34–41] In total, data from 952 participants with 314,137 paired measurements were included in this comparison (two studies did not report the total number of measurements included in their analysis). The sensitivity analysis for the primary comparison with only studies that were assessed as low risk of bias across all domains included 12 comparisons from 7 studies that enrolled 566 participants with 238,361 paired measurements (two studies did not report the total number of measurements included in their analysis). There were 5 studies that compared zero-heat-flux to core temperature measurements in ICU patients, comprising 7 comparisons with 155,598 measurements from 246 participants. Another 12 studies were included that compared zero-heat-flux to core temperature measurements in patients undergoing surgery, comprising 16 comparisons with 158,539 measurements from 809 participants. Nasopharyngeal thermometers were used as the comparator devices in 4 studies, with 108,180 paired measurements from 293 participants. Supplementary information presented in the study [webpage](https://zhf-info.netlify.app) provides comprehensive summaries of the studies included in each of these meta-analyses.

All comparisons reported adherence to the 3MTM Bair HuggerTM Temperature Monitoring System manufacturer instructions and placed the sensor on the forehead of the participants. One included study also reported results for comparisons where the zero-heat-flux thermometer was not placed on the forehead.[41] We did not include these comparisons in our meta-analysis.

Risk of bias assessments for each study are presented in Figure 2. High risk of bias associated with patient selection for 13 (43%) comparisons from 9 (56%) studies, conduct of zero-heat-flux and comparator measurements in 10 (33%) comparisons from 8 (50%) studies and 12 (40%) comparisons 9 (30%) studies, respectively (mostly due to zero-heat-flux measurements being taken with knowledge of the comparator measurements and vice versa) and participant flow for 13 (43%) studies. In 19 (50%) comparisons from 8 (50%) studies, the authors had declared a conflict of interest or receipt of funding or equipment from the manufacturer of the zero-heat-flux device under evaluation.

### Primary comparison

Table 2 presents results of the primary, sensitivity and all subgroup analyses. The pooled estimate for the mean bias between the 3MTM Bair HuggerTM Temperature Monitoring System and core temperature measurements was 0.03°C. Population limits of agreement, which take into consideration the between-study heterogeneity and sampling error, were wide, spanning from -0.93°C to 0.98°C (314,137 measurements; 952 participants; 16 studies). The amount of between-study heterogeneity is displayed graphically in the density plot in Figure 3. The quality of evidence for the primary comparison was downgraded to moderate quality due to concerns about study limitations.

### Sensitivity analyses

Population limits of agreement for the sensitivity analysis restricted to studies rated as having low risk of bias across all the domains of the QUADAS-2 were similar to the primary analysis (238,361 measurements; 566 participants; 7 studies). The mean bias was again 0°C with population limits of agreement spanning from -0.92°C to 0.93°C. A further sensitivity analysis excluding studies that received industry funding revealed population limits of agreement that were considerably wider than the primary analysis (-1.29°C to 1.36°C). The evidence rating for this sensitivity analysis was downgraded to low quality though, due to concerns about study limitations, inconsistency and imprecision.

To investigate if results were sensitive to our decision about which temperature monitoring sites to include in the primary analysis, we conducted a separate analysis excluding any naso/oropharyngeal comparisons. Results were similar to the primary analysis (mean bias 0.03°C with population limits of agreement from -1 to 1.06°C). The mean bias between 3MTM Bair HuggerTM Temperature Monitoring System and nasopharyngeal temperature measurements in meta-analysis of results from 4 studies was 0.01°C. Population limits of agreement were -0.94°C to 0.97°C. We downgraded the quality of evidence to low, again due to concerns about imprecision and study limitations.

## Subgroup analyses

We conducted two subgroup analyses for the primary comparison according to the clinical setting in which the study was conducted. In the subset of studies conducted in the ICU, the mean bias was -0.04°C with population limits of agreement between -1.47°C and 1.39 °C. The GRADE rating for quality of evidence was downgraded to very low quality due to study limitations, inconsistency and imprecision. In the subset of studies that evaluated the use of 3MTM Bair HuggerTM Temperature Monitoring System during surgery found a mean bias of 0.02°C with population limits of agreement from -0.95°C to 0.98°C. The quality of evidence was rated as moderate quality due to concerns about study limitations.

## Discussion

This systematic review showed that a temperature measurement from a 3MTM Bair HuggerTM Temperature Monitoring System device could be as much as about 1°C higher or lower than core temperature. These results may have important implications for practice. It was reassuring that results of our sensitivity analysis restricted to studies assessed to be at low risk of bias using the QUADAS-2 tool were similar. As such, it is vital for healthcare professionals considering using this device to first determine if differences in temperature smaller than this magnitude would be important for the given clinical situation. If so, then it may not be appropriate to substitute the 3MTM Bair HuggerTM Temperature Monitoring System in place of a core thermometer.

Our estimates of the accuracy of measurements from the 3MTM Bair HuggerTM Temperature Monitoring System are similar to results from a previous meta-analysis that compared other peripheral thermometers, such as sublingual and temporal artery devices, with core temperature measurements.[42] However, this previous meta-analysis used a statistical approach that did not incorporate the magnitude of heterogeneity in results between studies or sampling error. As such, it is possible that the zero-heat-flux thermometer may still be more precise than other peripheral thermometers.

The accuracy of 3MTM Bair HuggerTM Temperature Monitoring System has been evaluated in various intraoperative contexts as well as in the intensive care unit setting. Our subgroup analyses did not indicate that the accuracy of this temperature monitoring device was any more effective in a particular setting and revealed insights to direct future research. There were fewer participants included in the subgroup of studies conducted in the Intensive Care Unit compared to those conducted during surgery. As a result, population limits of agreement for the Intensive Care Unit subgroup were broad. Additional studies are required to increase confidence in the accuracy of 3MTM Bair HuggerTM Temperature Monitoring System device in this setting, where continuous temperature monitoring is often required. Also of note, only two studies included in our systematic review included pediatric patients. Additional studies to evaluate the accuracy of this device in children may be warranted to increase confidence.

It should be noted that we did not include studies that used the Temple Touch Pro because it is not strictly a zero-heat-flux device. This is a new thermometer that is similar to the zero-heat-flux device in that it is placed cutaneously, but the underlying technology is different.[43] Likewise, studies that evaluated the Dräger Tcore device were not included because this review focused specifically on the 3MTM Bair HuggerTM Temperature Monitoring System.

Many studies in this review analysed a large number of measurements of temperature with relatively small sample sizes. Importantly, the approach we used for our meta-analysis takes this into account. By using robust variance estimation, weights for pooling estimates in the meta-analysis become proportional to the number of patients, not the total number of measurements.[18]

### Limitations

We did not extract data on adverse events due to temperature monitoring with the 3MTM Bair HuggerTM Temperature Monitoring System. The possibility of publication bias cannot be ruled out, although the evidence suggests this may not be as serious of a problem for studies that are not randomized controlled trials.[15] Our focus for the meta-analysis was on calculating population limits of agreement, which incorporate the variation in bias between studies into the estimates. For this reason, we did not use meta-regression or tests for interaction between subgroups as a way to investigate sources of heterogeneity. It is important to note that this review did not assess the clinical utility of temperature monitoring using the zero-heat-flux device. The evidence from this review should be considered in the context of other information about the reliability and ease of use of this device. For example, the 3MTM Bair HuggerTM Temperature Monitoring System permits use of the same temperature monitoring device to be used throughout the whole perioperative care pathway. As such, replacing indirect estimates of core temperature, such as infrared tympanic and temporal devices, may be an advantage of the 3MTM Bair HuggerTM Temperature Monitoring System. Indirect estimates of core temperature are not recommended for use in surgical patients, yet are commonly used because of convenience.[6]

Finally, our initial protocol was focused on the accuracy of zero-heat-flux temperature measurements in general, not the 3MTM Bair HuggerTM Temperature Monitoring System specifically. A decision was made after data extraction and analysis was conducted to revise our inclustion criteria and include only studies that reported on the accuracy of the 3MTM Bair HuggerTM Temperature Monitoring System for this particular report. This decision was made because of the potential that differences in the technology integrated into zero-heat-flux devices produced by different manufacturers (i.e. 3MTM Bair HuggerTM Temperature Monitoring System and the Dräger Tcore) will influence estimates about the accuracy of zero-heat-flux devices overall. Although revising the inclusion criteria must be considered a limitation, this decision permitted a more targeted evaluation of the 3MTM Bair HuggerTM Temperature Monitoring System and easier interpretation of results.

## Conclusion

Substantial differences between temperature measurements from core and the 3MTM Bair HuggerTM Temperature Monitoring System were identified in this meta-analysis. Clinicians should consider the range of uncertainty in the accuracy of the zero-heat-flux thermometer when using this device to inform their decision-making. As such, there may be circumstances where use of this device would not be appropriate because a difference in temperature of 1°C is important to detect for the given clinical situation.

# References

1. Frank SM, Fleisher LA, Breslow MJ, et al (1997) Perioperative maintenance of normothermia reduces the incidence of morbid cardiac events: A randomized clinical trial. JAMA 277:1127–1134

2. Kurz A, Sessler DI, Lenhardt R (1996) Perioperative normothermia to reduce the incidence of surgical-wound infection and shorten hospitalization. New England Journal of Medicine 334:1209–1216

3. Michelson AD, MacGregor H, Barnard MR, et al (1994) Reversible inhibition of human platelet activation by hypothermia in vivo and in vitro. Thrombosis and Haemostasis 72:633–640

4. Rohrer MJ, Natale AM (1992) Effect of hypothermia on the coagulation cascade. Critical Care Medicine 20:1402–1405

5. Hadian M, Pinsky MR (2006) Evidence-based review of the use of the pulmonary artery catheter: Impact data and complications. Critical Care 10:S8

6. Riley C, Andrzejowski J (2018) Inadvertent perioperative hypothermia. BJA Education 18:227–233

7. Eshraghi Y, Nasr V, Parra-Sanchez I, et al (2014) An evaluation of a zero-heat-flux cutaneous thermometer in cardiac surgical patients. Anesthesia & Analgesia 119:543–549

8. Sessler M Daniel I. (2008) Temperature Monitoring and Perioperative Thermoregulation. Anesthesiology: The Journal of the American Society of Anesthesiologists 109:318–338. <https://doi.org/10.1097/ALN.0b013e31817f6d76>

9. Yamakage M, Namiki A (2003) Deep temperature monitoring using a zero-heat-flow method. Journal of anesthesia 17:108–115

10. Mann R, Gilbody SM (2012) Should methodological filters for diagnostic test accuracy studies be used in systematic reviews of psychometric instruments? A case study involving screening for postnatal depression. Systematic Reviews 1:9

11. Preston L, Carroll C, Gardois P, et al (2015) Improving search efficiency for systematic reviews of diagnostic test accuracy: An exploratory study to assess the viability of limiting to medline, embase and reference checking. Systematic Reviews 4:82

12. Myles PS, Cui J (2007) I. Using the bland–altman method to measure agreement with repeated measures. BJA: British Journal of Anaesthesia 99:309–311. <https://doi.org/10.1093/bja/aem214>

13. Whiting PF, Rutjes AW, Westwood ME, et al (2011) QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. Annals of Internal Medicine 155:529–536

14. Glasziou P, Irwig L, Bain C, Colditz G (2001) Systematic reviews in health care: A practical guide. Cambridge:Cambridge University Press

15. Begg CB (2005) Systematic reviews of diagnostic accuracy studies require study by study examination: First for heterogeneity, and then for sources of heterogeneity. Journal of Clinical Epidemiology 58:865

16. Deeks JJ, Macaskill P, Irwig L (2005) The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. Journal of Clinical Epidemiology 58:882–893

17. Schünemann HJ, Oxman AD, Brozek J, et al (2008) Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ 336:1106–1110

18. Tipton E, Shuster J (2017) A framework for the meta-analysis of bland–altman studies based on a limits of agreement approach. Statistics in Medicine 36:3621–3635

19. Hedges LV, Tipton E, Johnson MC (2010) Robust variance estimation in meta-regression with dependent effect size estimates. Research Synthesis Methods 1:39–65

20. Tanner-Smith EE, Tipton E, Polanin JR (2016) Handling complex meta-analytic data structures using robust variance estimates: A tutorial in r. Journal of Developmental and Life-Course Criminology 2:85–112

21. Tipton E (2015) Small sample adjustments for robust variance estimation with meta-regression. Psychological Methods 20:375

22. DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. Controlled Clinical Trials 7:177–188

23. R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

24. Boisson M, Alaux A, Kerforne T, et al (2018) Intra-operative cutaneous temperature monitoring with zero-heat-flux technique (3M spoton) in comparison with oesophageal and arterial temperature: A prospective observational study. European Journal of Anaesthesiology (EJA) 35:825–830

25. Dahyot-Fizelier C, Lamarche S, Kerforne T, et al (2017) Accuracy of zero-heat-flux cutaneous temperature in intensive care adults. Critical Care Medicine 45:e715–e717. <https://doi.org/10.1097/ccm.0000000000002317>

26. Schell-Chaple HM, Liu KD, Matthay MA, Puntillo KA (2018) Rectal and bladder temperatures vs forehead core temperatures measured with SpotOn monitoring system. American Journal of Critical Care 27:43–50. <https://doi.org/10.4037/ajcc2018865>

27. Mäkinen M-T, Pesonen A, Jousela I, et al (2016) Novel zero-heat-flux deep body temperature measurement in lower extremity vascular and cardiac surgery. Journal of Cardiothoracic and Vascular Anesthesia 30:973–978. <https://doi.org/10.1053/j.jvca.2016.03.141>

28. West N, Cooke E, Morse D, et al (2019) Zero-heat-flux core temperature monitoring system: An observational secondary analysis to evaluate agreement with naso-/oropharyngeal probe during anesthesia. Journal of Clinical Monitoring and Computing. <https://doi.org/10.1007/s10877-019-00411-y>

29. Pesonen E, Silvasti-Lundell M, Niemi TT, et al (2018) The focus of temperature monitoring with zero-heat-flux technology (3M bair-hugger): A clinical study with patients undergoing craniotomy. Journal of Clinical Monitoring and Computing 33:917–923. <https://doi.org/10.1007/s10877-018-0227-z>

30. Iden T, Horn E-P, Bein B, et al (2015) Intraoperative temperature monitoring with zero heat flux technology (3M SpotOn sensor) in comparison with sublingual and nasopharyngeal temperature. European Journal of Anaesthesiology 32:387–391. <https://doi.org/10.1097/eja.0000000000000232>

31. Lim Hyungsun KD-C Kim Boram (2016) A comparison of the temperature difference according to the placement of a nasopharyngeal temperature probe. Korean J Anesthesiol 69:357–361. <https://doi.org/10.4097/kjae.2016.69.4.357>

32. Cullen S, Brown J (2015) Measuring core temperature-an audit:47. Anaesthesia 70 (S3):11–101

33. Vendrell JM, Lacambra BM, Bassas PE, et al (2016) Zero-heat-flux cutaneous thermometer for core temperature measurements in extreme situations. Journal of Neurosurgical Anesthesiology 228:S33

34. Carvalho H, Najafi N, Poelaert J (2019) Intra-operative temperature monitoring with cutaneous zero-heat- flux-thermometry in comparison with oesophageal temperature: A prospective study in the paediatric population. Pediatric Anesthesia. <https://doi.org/10.1111/pan.13653>

35. G’omez-Romero F, Fern’andez-Prada M, Fern’andez-Su’arez F, et al (2019) Intra-operative temperature monitoring with two non-invasive devices (3M spoton and dräger tcore) in comparison with the swan-ganz catheter. Cirugı'a Cardiovascular 26:191–196. <https://doi.org/10.1016/j.circv.2019.06.002>

36. Idei M, Nomura T, Yamamoto N, et al (2016) An evaluation and accuracy of new zero-heat-flux thermometer (3M SpotOn) in pediatric intensive care patients. Intensive Care Medicine Experimental 4:Supplement 1

37. Jack JM, Ellicott H, Jones CI, et al (2019) Determining the accuracy of zero-flux and ingestible thermometers in the peri-operative setting. Journal of Clinical Monitoring and Computing 33:1113–1118. <https://doi.org/10.1007/s10877-019-00252-9>

38. Kato H, Kawashima S, Mimuro S, et al (2015) An evaluation of deep-forehead temperature (spoton) in ICU patients after cardiac surgery. Intensive Care Medicine Experimental 3: <https://doi.org/10.1186/2197-425x-3-s1-a111>

39. Kollmann AC, Brogly N, Alsina E, et al (2019) Validation of the zero-heat-flux thermometer (spoton) in major gynecological surgery to monitor intraoperative core temperature: A comparative study with esophageal core temperature. Minerva Anestesiologica 85:351–357

40. Morettini E, Turchini F, Tofani L, et al (2019) Intraoperative core temperature monitoring: Accuracy and precision of zero-heat flux heated controlled servo sensor compared with esophageal temperature during major surgery: The ESOSPOT study. Journal of Clinical Monitoring and Computing. <https://doi.org/10.1007/s10877-019-00410-z>

41. Tachibana S, Chida Y, Yamakage M (2019) Using the bair hugger temperature monitoring system in neck and chest regions: A pilot study. JA Clinical Reports 5: <https://doi.org/10.1186/s40981-019-0252-z>

42. Niven DJ, Gaudet JE, Laupland KB, et al (2015) Accuracy of peripheral thermometers for estimating temperature: A systematic review and meta-analysis. Annals of Internal Medicine 163:768–777

43. Evron S, Weissman A, Toivis V, et al (2017) Evaluation of the temple touch pro, a novel noninvasive core-temperature monitoring system. Anesthesia & Analgesia 125:103–109

# Figure legend

Fig. 1 PRISMA Flow Diagram

Fig. 2 Risk of bias assessments for included studies

Fig. 3 Comparisons between core and zero-heat-flux thermometers within and across studies

# Appendix

## Medline search strategy

1. zero-heat-flux
2. zero-flux
3. 1 OR 2
4. (SpotON) OR (Temple Touch Pro)
5. 3 OR 4
6. Accuracy OR precision OR reliability OR validity OR validation OR standard deviation
7. Bias OR mean difference OR limit of agreement OR Bland Altman
8. 6 OR 7
9. (exp “diagnostic errors” / OR exp “sensitivity and specificity” / OR (accura\* OR reliability\* OR target\* OR utilt\* OR discriminat\* OR differentiat\*)
10. 8 OR 9
11. 5 AND 10

## EMBASE search strategy

1. zero-heat-flux
2. zero-flux
3. 1 OR 2
4. (SpotON) OR (Temple Touch Pro)
5. 3 OR 4
6. Accuracy OR precision OR reliability OR validity OR validation OR standard deviation
7. Bias OR mean difference OR limit of agreement OR Bland Altman
8. 6 OR 7
9. (‘diagnostic accuracy’/de OR ‘diagnostic test accuracy study’/de OR ‘diagnostic error’/exp OR ‘diagnostic value’/de OR ‘sensitivity and specificity’/de OR ‘predictive value’/de OR (accura\* OR reliabilit\* OR target\* OR utilit\* OR discriminat\* OR differentiat\*)
10. 5 AND 10

## Table 1: Study Characteristics

| Study | | | | Participants | | Temperature measurements | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | Age (years)a |  |  | Participants |  | Measurements | Range (°C) |
| 2019 | Carvalho |  | Pediatric Surgery |  | 5.5 (3.9) | Esophageal | 54 |  | 1664 |  |  |
| Gomez-Romero |  | Elective surgery |  | 69 (10.8) | Pulmonary artery | 41 |  | 289 |  |  |
| Jack |  | Elective Surgery |  | 48 [42-68] | Esophageal | 29 |  | 2511 |  |  |
| Kollmann Camaiora |  | Gynecology Surgery |  | 51.5 (15.2) | Esophageal | 66 |  | 401 |  |  |
| Morettini |  | Elective Surgery |  | 67.1 (12.1) | Esophageal | 99 |  | 1346 |  | 33.2-38 |
| Tachibana |  | Laparoscopy Surgery |  | 49.3 (15.9) | Esophageal | 10 |  | 303 |  | 35.5-38 |
| West |  | Elective surgery |  | 60 [49-69] | Naso or oropharyngeal | 194 |  | 132721 |  |  |
| 2018 | Boisson |  | Abdominal Surgery |  | 64 [55-69] | Axillary artery | 22 |  | 1708 |  | 33-40.2 |
| Axillary artery | 22 |  | 72 |  |
| Esophageal | 49 |  | 3245 |  |
| Esophageal | 49 |  | 101 |  |
| Pesonen |  | Craniotomy |  | 59 [44-67] | Bladder | 29 |  |  |  |  |
| Schell-Chaple |  | Febrile ICU Patients |  | 57 (15) | Bladder | 9 |  | 180 |  | 36.6-39.9 |
| Rectal | 29 |  | 580 |  |
| 2017 | Dahyot-Fizelier |  | ICU |  | 54 (5) | Iliac artery | 7 |  | 1850 |  | 33-39.7 |
| Esophageal | 52 |  | 61298 |  |
| 2016 | Idei |  | Pediatric ICU |  |  | Rectal | 26 |  | 53495 |  |  |
| Makinen |  | Vascular Surgery |  | 64 (13) | Esophageal | 15 |  |  |  |  |
| Cardiac Surgery | Pulmonary artery | 15 |  |  |  |
| 2015 | Iden |  | Elective Surgery |  | 47 (13) | Nasopharyngeal | 83 |  | 249 |  |  |
| Kato |  | ICU |  | 66 (21) | Pulmonary artery | 20 |  | 16407 |  | 35.6-39.2 |
| 2014 | Eshraghi |  | Cardiac Surgery |  | 67 (9) | Pulmonary artery | 103 |  | 35717 |  | 32-39.3 |
| amean (standard deviation) or median [interquartile range] | | | | | | | | | | | |

### Table 2: Results of meta-analyses

|  |  |  |  |  |  |  |  |  |  | Population LoA | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Studies | Comparisons | Participants | Measurements | Mean biasa | SD2ab | τ2ac | Lower 95% LoAad | Upper 95% LoAad | Outer CI for lower 95% LoAade | Outer CI for upper 95% LoAade |
| Primary | 16 | 22 | 952 | 314,137 | 0.03 | 0.11 | 0.04 | -0.75 | 0.80 | -0.93 | 0.98 |
| Low risk studies | 7 | 12 | 566 | 238,361 | 0.00 | 0.10 | 0.01 | -0.69 | 0.69 | -0.92 | 0.93 |
| Core (excluding nasopharyngeal and oropharyngeal) | 14 | 20 | 675 | 181,167 | 0.03 | 0.12 | 0.05 | -0.80 | 0.86 | -1.00 | 1.06 |
| Nasopharyngeal | 4 | 4 | 293 | 108,180 | 0.01 | 0.10 | 0.01 | -0.63 | 0.65 | -0.94 | 0.97 |
| Studies not funded by industry | 8 | 9 | 354 | 74,665 | 0.03 | 0.13 | 0.10 | -0.94 | 1.01 | -1.29 | 1.36 |
| Intensive Care Unit | 5 | 7 | 246 | 155,598 | -0.04 | 0.10 | 0.17 | -1.08 | 1.00 | -1.47 | 1.39 |
| Surgery | 12 | 16 | 809 | 158,539 | 0.02 | 0.13 | 0.01 | -0.72 | 0.76 | -0.95 | 0.98 |
| aUnits are °C | | | | | | | | | | | |
| bVariance | | | | | | | | | | | |
| cMeasure of heterogeneity | | | | | | | | | | | |
| dLoA = Limits of Agreement | | | | | | | | | | | |
| eCI = Confidence Intervals | | | | | | | | | | | |