BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Evaluation of a wearable wireless continuous temperature monitoring device with artificial intelligence, iThermonitor®, for patients in surgical wards: a prospective comparative study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-039474
Article Type:	Original research
Date Submitted by the Author:	21-Apr-2020
Complete List of Authors:	Liu, Yuwei; Sichuan University West China Hospital, Department of general surgery Liu, Changqing; Sichuan University West China Hospital, Operating room Gao, Min; Sichuan University West China Hospital, Department of biliary surgery Wang, Yan; Sichuan University West China Hospital, Department of biliary surgery Bai, Yangjing; Sichuan University West China Hospital, Department of cardiac and macrovascular surgery Gong, Renrong; Sichuan University West China Hospital, West China School of Nursing Xu, Ruihua; Sichuan University West China Hospital, Department of biliary surgery
Keywords:	Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Adult surgery < SURGERY, Physiology < BASIC SCIENCES, Telemedicine < BIOTECHNOLOGY & BIOINFORMATICS

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

Evaluation of a wearable wireless continuous temperature monitoring device with artificial intelligence, iThermonitor®, for patients in surgical wards: a prospective comparative study

Yuwei Liu[a,b], Changqing Liu [c], Min Gao [d], Yan Wang [d], Yangjing Bai [e], Renrong Gong [a,b,*], Ruihua Xu [c,*]

[a] *Department of general surgery, West China Hospital of Sichuan University, Chengdu, P.R.China*

[b] *West China School of Nursing, Sichuan University, Chengdu, P.R.China*

[c] *Department of operating room, West China Hospital of Sichuan University, Chengdu, P.R.China*

[d] *Department of biliary surgery, West China Hospital of Sichuan University, Chengdu, P.R.China*

[e] *Department of cardiac and macrovascular surgery, West China Hospital of Sichuan University, Chengdu, P.R.China*

Key words: Monitoring, Physiologic; Body Temperature; Wearable Electronic Devices

Word count: 3795.

* Corresponding authors at: No. 37 Guoxue Lane, Wuhou, Chengdu, Sichuan, China.

E-mail address: gongrenrong@wchscu.cn (R. Gong), xuruihua@wchscu.cn (R. Xu)

Tel: Renrong Gong, 18980601527; Ruihua Xu, 18980601458

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

ABSTRACT

Objectives: To evaluate a new generation of wireless axillary thermometer, the iThermonitor®, and to ascertain its feasibility for perioperative continuous body temperature monitoring in surgical patients.

Setting: Department of biliary surgery, operating room and post anesthesia care unit of a tertiary university teaching hospital in Chengdu, China.

Participants: 526 adult surgical patients were consecutively enrolled.

Design: A prospective observational study. Axillary temperature was continuously recorded by a noninvasive wireless thermometer iThermonitor® (WT705, Raiing Medical, Beijing, China), throughout the whole perioperative period. Temperature of the contra lateral armpit was measured as reference with mercury thermometers at 8:00, 12:00, 16:00 and 20:00 every day.

Outcome measures: The primary outcomes were the accuracy and precision of iThermonitor®. Secondary outcomes were the validity of detecting fever and the feasibility of continuous wearing. Pairs of temperatures were evaluated by student *t*-test, Pearson`s correlation and Bland-Altman plot.

Results: 3621 pairs of body temperatures were obtained. The temperatures measured by two methods agreed overall, with a mean difference of 0.03°C±0.36°C and a moderate correlation (r=0.755, P＜0.001). The 95% limits of agreement ranged from -0.66°C to 0.72°C, only 5.16% of the points were outside the 95% limits of agreement. The Intra-class correlation coefficient was 0.753. The continuous temperature monitoring captured more mild fevers than intermittent observation (117/526 vs.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

91/526, *P*= < 0.001), detected fever up to 4.35 hours earlier, and captured higher peak temperature (0.29°C ± 0.27°C, 95%CI: 0.26-0.31). The sensitivity was 92.31% and the specificity was 92.41% to detect fever. All subjects felt that wearing the iThermonitor® was more or less comfortable.

Conclusions: The iThermonitor® WT705 was sufficiently accurate and feasible for continuous temperature monitoring in surgical patients. Dynamically reflecting the individual trends of body temperature throughout the whole perioperative period improves fever detection.

Strengths and limitations of this study

- A new generation of wireless dermal thermometer, iThermonitor®, was sufficiently accurate for remote continuous monitoring in postsurgical patients.
- We highlighted the value of continoulsly monitoring body temperature (including core temperature) covering the whole perioperative period, which could dynamically reflect the individual trends of body temperature and improves fever detection.
- The main limitation is that we only compared the accuracy of iThermonitor® to mercury thermometers in detecting axillary temperature. Adding a set of core body temperature would help us better understand which devices are closer to the core temperature when deviations occurs.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

INTRODUCTION

Body temperature is one of the most foundational vital signs of patients. Surgical patients are typically exposed to cold environments, administrations of unwarmed intravenous fluids, bacteria invasions, and anesthetic drugs which may impair the thermoregulatory system [1], leading to perturbations of body temperature. Accurately monitoring the body temperature is essential to prevent hypothermia, detect infectious complications for surgical patients [2 3].

No ideal device to measure body temperature has been found yet [4], in terms of accuracy, availability, affordability, and ability of continuous monitoring across different clinical settings [5]. Peripheral thermometers measuring temperature from tympanic membrane, temporal artery, oral cavity, forehead or other parts are considered to be not stable and accurate enough [6-8]. Inserts of a temperature probe to esophageal, pulmonary artery, nasopharynx, rectum or bladder could precisely and continuously detect the core temperature [9], but these invasive devices increased the risk of infection, and were only used for intensive care units patients and surgical patients under anesthesia when necessary [10 11]. Until today, there are still urgent needs of thermometers to accurately and continuously monitor body temperature in clinical practice.

Wearable technology is changing the way we measure body temperature and perform clinical care [12]. In recent years, several wireless dermal wearable thermometers increased the feasibility of continuous body temperature monitoring outside of the critical care setting [13]. However, only a small proportion of wearable

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

devices have been CE marked (class Ⅱ or above) or are FDA approved as medical device [14]. The options that have enough convincing evidences to support their accuracy for patients at risk for complications in a clinical environment are still few [15]. Wearable device that only reflecting the surface skin temperature has been proved significant bias and poor correlation with oral temperature [16]. Zsuzsanna Balla et al. [13] found seven wireless dermal thermometers reflecting core temperature through internet searching, and tested four of them (only FeverSmart, iThermonitor®WT701, Quest Temp Sitter, and Thermochron iButton were commercially available). The results indicated that they were not reliable and accurate enough for most types of clinical research, although the iThermonitor® WT701 systems had the least unsatisfactory correlation to the rectal thermometer than other devices. Moreover, surgical patients were typically transferred between multiple units of care (i.e., Surgical Wards, Operating Room, Post anesthesia Care Unit). Whether these devices are capable of continuous temperature monitoring across different clinical scenarios remains unclear, since maintain the accuracy has to deal with different challenges as scenarios change. For instance, the tissue perfusion, physical activity, length of wearing time and compliance differs from awake patients and those who under anesthesia. Battery life and internal storage space also limited their application.

A new generation of noninvasive wireless dermal thermometer, iThermonitor® WT705, carried advanced versions of machine learning algorithms, may satisfy the ambition of continuously monitoring the body temperature (including core temperature) in different clinical settings. A previous version of iThermonitor® for

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

intraoperative monitoring has been reported well accuracy in representing core body temperature (distal esophagus temperature) from the axilla in patients under anesthesia [17]. However, the performances of version for ward/home have not been tested before. Therefore, we conducted a prospective study to assess the accuracy and feasibility of iThermonitor® WT705 in awake patients in surgical wards, and its potential for continuous monitoring of body temperature throughout the whole perioperative period at real clinical settings.

METHODS

Study design, subjects, and setting

A prospective comparative descriptive study was planned to evaluate the axillary wireless sensor iThermonitor® for continuous temperature monitoring by comparing it to mercury-contained thermometers. Patients admitted to the department of biliary surgery in West China Hospital of Sichuan University were consecutively recruited in this study, from August to December, 2019. The inclusion criteria was only that the patient signed the informed consent. Patients with any impediment to attach the temperature sensor iThermonitor® under their axillary were excluded. Finally, 526 patients were enrolled and all of them signed their informed written consent. The room temperature was between 22°C and 26°C.

Instruments

The study instruments were: (1) Wireless noninvasive dermal thermometer iThermonitor® (model WT705, Raiing Medical Company, Beijing, China), with

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

accuracy of ±0.1°C (5°C-40°C). This is a battery-operated reusable electronical device with 30 days of battery life. The price is also sufficiently affordable for single patient use, as executed in our study. The US Food and Drug Administration approved it as a class Ⅱ medical device. The iThermonitor® would be securely attached to the axilla (shaved if necessary) of patients with adhesive tape provided by manufacturer. The sensor would record the axillary temperature once every 4 seconds, then output the temperature every minute after calculation. The iThermonitor® WT705 is an updated version that equipped with a more powerful chip carrying two versions of patented machine learning algorithms to maintain the accuracy both in the operating room and ward/home. The version for operating room use would gauge the core temperature from axilla, after correcting changes in tissue perfusion caused by anesthetics. Another version for ward/home use tested in our study would reflect the axillary temperature itself (to adapt to clinical decision-making habits), after adjustment for daily activities, body posture changes, adhesive tapes loose and a more variable ambient temperature. The algorithms and the data output frequency can be pre-seted as needed. The temperature data with time and scenario stamps would be transmitted wirelessly via Bluetooth or WiFi to the central computer that installed in the nurse station. The dynamic temperature curves of all patients were visualized on the screen. All the iThermonitor® sensors were proofreaded for accuracy according to the manufacturer`s standard before use. (2) Mercury thermometers, with accuracy of ±0.1°C, over the range of 35°C - 42°C. Mercury thermometers would be calibrated per week by comparing to a high-precision industrial mercury thermometer as the

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

standard within the same constant temperature water tank. Mercury thermometers with temperature deviations greater than 0.2°C would not be used. We chose the mercury thermometer as reference because it was the most robust device in reflecting axillary temperature [18]. We also want to know whether the iThermonitor® could be an ideal substitute choice for mercury thermometer, which may have implications for countries looking for alternative methods to mercury thermometers.

Data collection

Demographics, age, sex and BMI were extracted from the medical records. The iThermonitor® sensor would be attached to the dried and cleaned skin in the armpit region from the first day of admission, then continuously recorded the body temperature throughout the whole perioperative period until discharge. Patients were instructed to keep wearing it, except for a CT scan or a shower. A nurse checked the skin of axillary everyday to assess whether the patient had local skin allergy or other adverse reactions. We also observed the compliance of patients wearing the iThermonitor® sensor. Patients were asked to complete a questionnaire with a scale from 0 to 5 to evaluate the tolerability of iThermonitor® on the day of discharge.

The mercury thermometer would be kept under the patient`s other armpit for 10 minutes after cleaned and dried, at specific time intervals at 8:00AM, 12:00AM, 16:00PM and 20:00PM every day. Two specially trained nurses were assigned to measure the temperature. One nurse recorded the actual time for reading the mercury thermometer (accurate to minutes), then another nurse checked the wireless temperature of the patients at the same time from the central monitoring station. If the

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

difference between the concurrent temperature measured by the two devices was over than 1°C, the measurement would be repeated immediately with a mercury thermometer, and whether the iThermonitor® was correctly wore would be checked. If there was a confirmed error due to improper measurement, this set of data would be excluded. If we confirmed that the difference was really exists, this set of data would be taken into analysis. We also assigned a nurse to ensure the continuity of temperature monitoring in the operating room (OR) and post anesthesia care unit (PACU).

Main end points

The following endpoints are evaluated:

- Accuracy. The accuracy was indicated by the mean of the difference (also called bias) and the standard deviation, as calculated by temperature recorded by iThermonitor® minus that of mercury thermometers. A priori, an absolute difference of 0.5°C was considered as clinically acceptable [17 19].

- Precision. Precision (also called reliability) was indicated by the 95% limits of agreement and Intra-class correlation coefficient. An ICC greater than 0.7 would be considered as well accepted agreement between the two temperature measurement methods.

- Validity. Validity was evaluated by the sensitivity and specificity in detecting fever. Axillary temperature of 38°C was considered as a cut-off value of fever [7]. We also compared the peak temperature recorded by the two methods, and the time of fever detected. For the iThermonitor®, when the body temperature exceeded 38°C (lasting at least 5 minutes) for the first time, the moment would be compared to the time that fever first detected by mercury thermometers.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

- Feasibility. The feasibility of continuous temperature monitoring was assessed with comfort score and possible adverse events. The comfort score ranked from 0 to 5, with 0 meaning most uncomfortable and intolerable, 5 meaning very comfortable. Any adverse event, such as skin blisters or ulcer caused by iThermonitor®, or breakages of mercury thermometers, would be recorded.

Statistics

Data were analyzed by Python (Version 3.5.1) and MedCalc (Version 19.1.3) software. Mean and standard deviations of temperature were calculated for iThermonitor® and mercury thermometers. The student *t*-test was utilized for paired samples. Pearson`s correlation analysis and Bland-Altman plot were used to evaluate the relationship between the two sets of temperatures. $P< 0.05$ was considered as statistically significant.

Sample Size Considerations

The sample size module for Bland-Altman plot of MedCalc software was conducted. We expected the mean of differences was 0.03°C, with a standard deviation of 0.23°C, according to our previous pilot study. This calculation set a maximum allowed difference of 0.5°C, a type Ⅰ error rate (Alpha) of 0.05, and a power of 80%. Finally 3292 pairs of data were deemed sufficient to adequately detect a difference between the iThermonitor® and mercury thermometer.

Patient and public involvement

No patients or members of the public were involved in the design of this study, in the implementation of the study or in result dissemination.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

RESULTS

Characteristics of the patients

We enrolled 526 patients. All of them completed the study and were included in the final analysis, no missing values need to be processed. Temperature curves with time stamps across different clinical scenarios were recorded for each patient. Figure 1 shows the example of a patient measurement. Among the 526 patients, there were 197 (37.5%) males and 329 (62.5%) females, with an average age of 53.52±14.20 years (over the range 15-86 years). The cumulative monitoring duration was 1768 days, with an average of 5.57±3.62 days for each patient. Demographic characteristics are shown in Table 1.

Figure 1. Example of a patient`s temperature curve with time stamps across different clinical scenarios

Table1. Characteristics of the patients (n=526)

Variable	Summary
Age (years)	53.85±14.20 (15-86)
Gender	
Male (%)	197(37.5%)
Female (%)	329(62.5%)
BMI (kg/m^2)	23.03±3.12(15.21-37.78)
Average monitor duration (days)	5.57±3.62(1-22)
Cumulative monitor duration (days)	1768

Data are presented as n (%) of patients, or mean±SD (range) for continuous variables.
Abbreviation: BMI, body mass index.

Precision

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

A total of 3621 pairs of body temperatures were obtained. A scatter plot shows a relatively strong linear correlation (r =0.755, P<0.001) between the two groups of temperature (Appendix figure-A). The Bland-Altman plot shows that the 95% limits of agreement were relatively narrow, with the upper limit at 0.72 °C, and the lower limit at -0.66 °C, only 5.16% (187/3621) of the points were outside the 95% limits of agreement (Figure 2). The intra-group correlation coefficient (ICC) was 0.753, indicating that the temperatures measured by the two methods were well agreed overall.

Figure 2. Bland-Altman plot of iThermonitor® axillary temperatures against mercury thermometer temperatures. ICC=0.753, 95% limits of agreement from -0.66°C to 0.72°C, only 5.16% of the points were outside the 95%limits of agreement.

Accuracy

As showing in Table 2, the mean iThermonitor® temperature was 36.61°C±0.49°C, ranged from 34.8°C to 39.6°C, while the mean temperature measured by mercury thermometers was 36.58°C±0.52 °C, ranged from 35.0°C to 39.9°C. The mean of difference (bias) was 0.031±0.353°C, ranged from -1.40°C to 1.80°C. The bias within ±0.5°C were accounted for 87.68%, while 99.17% were within ±1.0°C. According to the temperatures measured by the mercury thermometers, we divided the patient's body temperature into three adjacent intervals: hypothermia (less than 36.0°C), normal body temperature (36.0°C -37.9°C), and fever (≥38.0°C). The bias was the smallest (0.006°C ± 0.332°C, 95%CI: -0.006, 0.017) within the normal temperature range, whereas the largest bias (0.380°C±0.398 °C) existed when the temperature is

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

below 36°C. Although the differences in partial intervals were statistically significant, the bias of all intervals were within 0.5°C, the clinically acceptable standard defined a priori.

Table 2 Comparisons of temperatures measured by iThermonitor® and mercury thermometers in different intervals.

Range	Pairs of temperature	iThermonitor® (°C)	Mercury Thermometer (°C)	Bias (°C)*	95%CI of Bias		*t*	*P* #
					Lower	Upper		
< 36.0°C	275	36.09±0.40	35.71±0.18	0.380±0.398	0.333	0.428	15.839	< 0.001
36.0-37.9°C	3285	36.629±0.421	36.624±0.416	0.006±0.332	-0.006	0.017	0.971	0.332
≥38.0°C	61	38.11±0.54	38.27±0.36	-0.165±0.351	-0.255	-0.076	-3.682	< 0.001
Total	3621	36.61±0.49	36.58±0.52	0.031±0.353	0.020	0.043	5.326	< 0.001

Temperature intervals were divided according to the temperatures measured by mercury thermometers. Bias were calculated by iThermonitor® temperature minus mercury thermometer temperature. Data presented as mean±SD. *P* values wereobtained by Student`s *t* tests for paired samples, since bias were normally distributed.

Comparison of the two methods in detecting fever

Among 526 patients enrolled, 84 patients were simultaneously detected fever by two methods. However, 33 patients with body temperature slightly higher than 38°C were only detected by the iThermonitor® during our observation period, suggesting that continuous temperature monitoring could capture more mild fevers than intermittent observation. The sensitivity was 92.31% and the specificity was 92.41% to detect fever, indicating a relatively strong consistency of the two methods (Kappa=0.761, $P < 0.001$).

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 3 Consistency of iThermonitor® and mercury thermometer in the diagnosis of fever

		Mercury Thermometers		Total	Kappa	*P*
		Patients With Fever	Patients Without fever			
iThermonitor®	Patients With Fever	84	33	117	0.761	$P^1 < 0.001$
	PatientsWithout fever	7	402	409	-	$P^2 < 0.001$
	Total	91	435	526		

A patient with body temperature ≥38°C during the perioperative period was considered as 1 case of fever. P^1 was obtained by Kappa consistency test, indicating the existence of consistency. P^2 was obtained by paired chi-square test, indicating that there was significant difference between the two methods in detecting fever.

Comparison of average peak temperature

The average peak temperature captured by intermittent measurement with mercury thermometers was 37.26°C ± 0.56°C, whereas the continuous monitoring with iThermonitor® detected higher peak temperature of 37.55°C ± 0.59°C. A mild but statistically significant difference of 0.29°C±0.27°C (range: -0.45°C- 1.26°C, 95%CI: 0.26°C-0.31°C) was noted between the peak temperatures recorded by the two methods (Figure 3).

Figure 3. Peak temperature recorded by iThermonitor® and mercury thermometer.

Comparison of the earliest time of fever detected

The continuous monitoring with iThermonitor® detected fever earlier than intermittent measurement with mercury thermometers. A mean time interval of 4.35 hours was observed, with a minimum difference of -0.92 hours and maximum

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

difference of 25.34 hours.

Feasibility of continuous temperature monitoring

During 1768 monitoring days, 4.37% (23/526) of the patients complained of slight itching during the wearing period. Local skin redness was observed in two patients and then lessened after we changed the iThermonitor® to the contra lateral axilla. In terms of comfort, 109 of the first enrolled patients rated the comfort score. Sum to 21 patients ranked as 2, 81 patients ranked as 3, 6 patients ranked as 4 and 1 patients ranked as 5. All subjects felt that wearing the iThermonitor® were more or less comfortable and did not affect daily activities, indicating a well acceptable compliance of the iThermonitor®.

DISCUSSION

We conducted a prospective study to evaluate the performance of wireless dermal temperature sensor, iThermonitor® WT705, for continuous temperature monitoring in surgical patients. We choose iThermonitor® because it was one of the best performed devices due to relatively high sensitivity and robustness in reflecting core temperature than other available wireless dermal thermometers [13].

Principal Findings in precision

In our study, the algorithms were pre-setted for ward/home use. The results showed a small bias (0.03°C±0.35°C), moderate correlation and relatively narrow 95% limits of agreement between the axillary temperatures recorded by iThermonitor® and mercury thermometers. The mean of difference didn`t vary much at different temperature

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

intervals, except when the temperatures measured by mercury thermometers were below 36°C. At this interval, the iThermonitor® temperatures were higher and showed the largest bias (0.380°C±0.398°C), much more than other temperature intervals. One of the possible explanations might be that the mercury thermometers could not be accurate enough in reflecting low body temperatures when the patient did not maintain the best measurement posture for enough time. Overall, such difference was considered to be acceptable, as defined a priori, the iThermonitor® appears accurate and precise enough for clinical use.

Compared To Similar Studies

Different reference standards and algorithm versions resulted in inconsistent accuracy of iThermonitor® in previous reports. Pei et al.[17] tested the intraoperative version of iThermonitor® WT701, the axillary temperature recordings well represents core temperature in adults under anesthesia, with a mean difference of only 0.14°C±0.26°C (esophageal minus axillary) and 95% limits of agreement from -0.38°C to 0.66°C. Our work adds evidence to support the applicability of iThermonitor® in surgical wards, as a continuation or extension of intraoperative core temperature monitoring. As to the version for ward and home, it was reported that the iThermonitor® WT701 rendered a lower average temperature than rectal temperature (bias -0.77°C±0.53°C) [13]. Our study showed the iThermonitor® WT705 was better at reflecting axillary temperature than rectal temperature. This was sufficiently reasonable, since the algorithms for ward/home were originally established for reflecting axillary temperature in consideration of clinical habits. Comparisons to rectal temperature

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

were actually inappropriate. Even though the axillary temperature was not recognized as core temperature, it was adequate for guiding clinical decision in general wards.

As to the feasibility, all subjects in our study felt that wearing the iThermonitor® were more or less comfortable and did not affect daily activities. The iThermonitor® also showed highly easy to use for continuous monitoring children` body temperature at home [20]. It indicates that the iThermonitor® also had the potential to provide continuous temperature monitoring for patients after discharge.

Clinical Implications

The advantages and feasibility of continuous vital signs monitoring in general wards are attracting increasing attention [10] [21]. As claimed by previous study, the iThermonitor® was marginally superior in following the individual trends than in assessing absolute temperatures [13]. We further demonstrated that continuous temperature monitoring was better at capturing the peak of fever, and could detect fever 4.35 hours earlier than intermittent temperature monitoring. This finding may have important therapeutic implications. Dakappa et al. [11] also noted higher peak temperature using continuous tympanic temperature recording device (TherCom®) than using conventional recording of temperature, with a significant difference of 1.52°C. Another research reported 180 minutes earlier by wearable digital thermometers (TremTraq, with a battery life of 72 hours) to detect increases in body temperature than by standard monitoring [21]. Furthermore, maintaining the continuity of the body temperature monitoring process across different clinical scenarios enable us to install a real world database of the patient's perioperative body temperature,

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

which would bring us more findings in exploring the regularity of perioperative temperature fluctuation.

The mercury-containing thermometer has been widely used for hundreds of years because of its stable performances in reflecting temperature [18]. However, these glass based thermometers are fragile to leak mercury out, which is one of the top ten chemicals of major public health concern [22]. Although the World Health Organization has called for the phase out of mercury fever thermometers by 2020 [22], they are still widely used in many countries [5 23 24]. Gaps are evident in practices on promoting mercury-free thermometers. The lack of an ideal alternative device for temperature measurement maybe the main reason. Using iThermonitor® instead of mercury thermometers to reduce medical mercury emissions are beneficial to the patients, health personnel and public health [23 24].

Limitations

Our study has several limitations. First, we only compared the iThermonitor® to mercury thermometers in detecting axillary temperature. Adding a set of core body temperatures as the gold standard would help us better understand which devices are closer to the core temperature when deviations occurs. Second, we only repeated the measurement when the bias were over than 1°C. A more strict supervision of repeating the measurement when the temperature recorded by mercury thermometers are below 36°C, may improve the performance. Third, we did not consider whether there is a difference in armpit temperature between dominant and non-dominant arm. However, publication did not show that such differences existed [25].

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

CONCLUSIONS

The iThermonitor® WT705 was sufficiently accurate and feasible for continuous temperature monitoring in surgical patients. Dynamically reflecting the individual trends of body temperature throughout the whole perioperative period can help us better detect fever and may promote early warning of perioperative complications.

Contributors Yuwei Liu has participated in the conception, statistical analysis and written the manuscript. Changqing Liu was in charge of body temperature management in the OR and PACU. Min Gao and Yan Wang performed temperature measurement. Yangjing Bai participated in the design of the study. Renrong Gong conceived the study and reviewed & edited the manuscript. Ruihua Xu administrated the project. All authors read and approved the manuscript.

Funding This study was funded by the Sichuan University West China Nursing Discipline Development Special Fund Project (NO.HXHL19031).

Competing interest None declared.

Ethics approval Ethical approval was obtained from the Biomedical Research Ethics Committee of West China Hospital of Sichuan University (No. 2019-447).

Clinical trial registration The study was registered at Chinese Clinical Trail Registry (ChiCTR1900024549) on July 5, 2019. The full study protocol can be accessed at: *www.chictr.org.cn*.

Data sharing statement Data are available upon reasonable request.

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, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http:// creativecommons. org/ licenses/ by- nc/ 4. 0/

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

REFERENCES

1. Sessler DI. Opioids and postoperative shivering. *J Clin Anesth* 2016;31:42-43.
2. Middleton S, McElduff P, Drury P, et al. Vital sign monitoring following stroke associated with 90-day independence: A secondary analysis of the QASC cluster randomized trial. *Int J Nurs Stud* 2019;89:72-79.
3. Larach MG, Brandom BW, Allen GC, et al. Malignant Hyperthermia Deaths Related to Inadequate Temperature Monitoring, 2007–2012. *Anesthesia & Analgesia* 2014;119(6):1359-66.
4. Sollai S, Dani C, Berti E, et al. Performance of a non-contact infrared thermometer in healthy newborns. *BMJ Open* 2016;6(3):e008695.
5. Halder N, Peshin SS, Pandey RM, et al. Awareness assessment of harmful effects of mercury in a health care set-up in India: A survey-based study. *Toxicol Ind Health* 2015;31(12):1144-51.
6. Opersteny E, Anderson H, Bates J, et al. Precision, Sensitivity and Patient Preference of Non-Invasive Thermometers in a Pediatric Surgical Acute Care Setting. *J Pediatr Nurs* 2017;35:36-41.
7. Niven DJ, Gaudet JE, Laupland KB, et al. Accuracy of Peripheral Thermometers for Estimating Temperature. *Annals of Internal Medicine* 2015;163(10)
8. Geijer H, Udumyan R, Lohse G, et al. Temperature measurements with a temporal scanner: systematic review and meta-analysis. *BMJ Open* 2016;6(3):e009509.
9. 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. *Journal of PeriAnesthesia Nursing* 2010;25(2):71-78.
10. Downey CL, Chapman S, Randell R, et al. The impact of continuous versus intermittent vital signs monitoring in hospitals: A systematic review and narrative synthesis. *Int J Nurs Stud* 2018;84:19-27.
11. Dakappa PH, Bhat GK, Bolumbu G, et al. Comparison of Conventional Mercury Thermometer and Continuous TherCom((R)) Temperature Recording in Hospitalized Patients. *J Clin Diagn Res* 2016;10(9):OC43-OC46.
12. Liao Y, Thompson C, Peterson S, et al. The Future of Wearable Technologies and Remote Monitoring in Health Care. *Am Soc Clin Oncol Educ Book* 2019;39:115-21.
13. Zsuzsanna Balla H, Theodorsson E, Strom JO. Evaluation of commercial, wireless dermal thermometers for surrogate measurements of core temperature. *Scand J Clin Lab Invest* 2019;79(1-2):1-6.
14. Muzny M, Henriksen A, Giordanengo A, et al. Wearable sensors with possibilities for data exchange: Analyzing status and needs of different actors in mobile health monitoring systems. *Int J Med Inform* 2020;133:104017.
15. Breteler MJMM, Huizinga E, van Loon K, et al. Reliability of wireless monitoring using a wearable patch sensor in high-risk surgical patients at a step-down unit in the Netherlands: a clinical validation study. *BMJ Open* 2018;8(2):e020162.
16. Izmailova ES, McLean IL, Bhatia G, et al. Evaluation of Wearable Digital Devices in a Phase I Clinical Trial. *Clin Transl Sci* 2019;12(3):247-56.
17. Pei L, Huang Y, Mao G, et al. Axillary Temperature, as Recorded by the iThermonitor

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

WT701, Well Represents Core Temperature in Adults Having Noncardiac Surgery. *Anesth Analg* 2018;126(3):833-38.

18. Wright WF, Mackowiak PA. Origin, Evolution and Clinical Application of the Thermometer. *Am J Med Sci* 2016;351(5):526-34.
19. Lu SH, Dai YT, Yen CJ. The effects of measurement site and ambient temperature on body temperature values in healthy older adults: a cross-sectional comparative study. *Int J Nurs Stud* 2009;46(11):1415-22.
20. Kakarmath SS, de Redon E, Centi AJ, et al. Assessing the Usability of an Automated Continuous Temperature Monitoring Device (iThermonitor) in Pediatric Patients: Non-Randomized Pilot Study. *JMIR Pediatr Parent* 2018;1(2):e10804.
21. Abbasi J. Wearable Digital Thermometer Improves Fever Detection. *JAMA* 2017;318(6):510.
22. WHO calls for the phase out of mercury fever thermometers and blood pressure measuring devices by 2020. Available at: https://www.who.int/news-room/detail/11-10-2013-who-calls-for-the-phase-out-of-mercury-fever-thermometers-and-blood-pressure-measuring-devices-by-2020 [accessed 11 October 2013].
23. Li P, Yang Y, Xiong W. Impacts of Mercury Pollution Controls on Atmospheric Mercury Concentration and Occupational Mercury Exposure in a Hospital. *Biol Trace Elem Res* 2015;168(2):330-4.
24. Chadli-Debbiche A, Allani R, Jouini R, et al. Mercurial hospital waste: management, becoming and recommendations. *Tunis Med* 2018;96(5):273-80.
25. 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(7):872-80.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

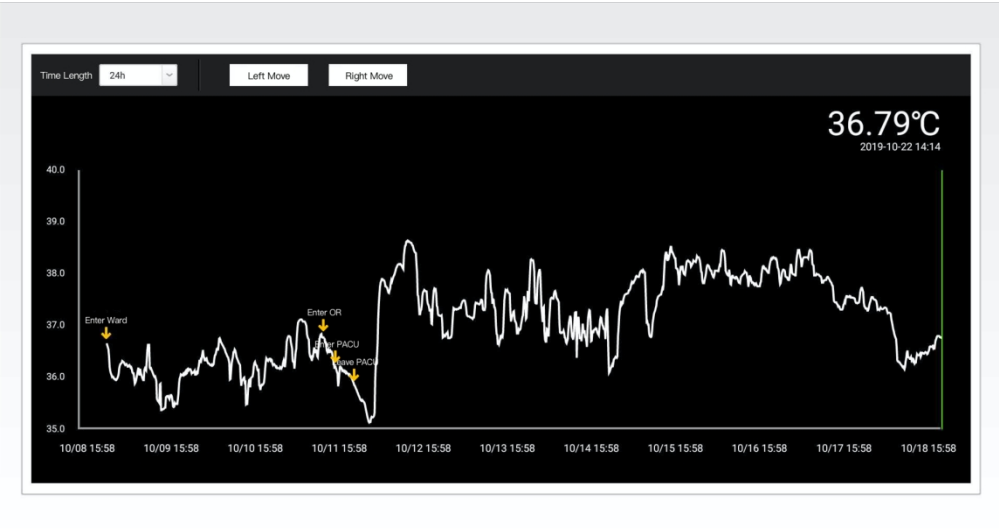


Figure 1. Example of a patient`s temperature curve with time stamps across different clinical scenarios

675x359mm (72 x 72 DPI)

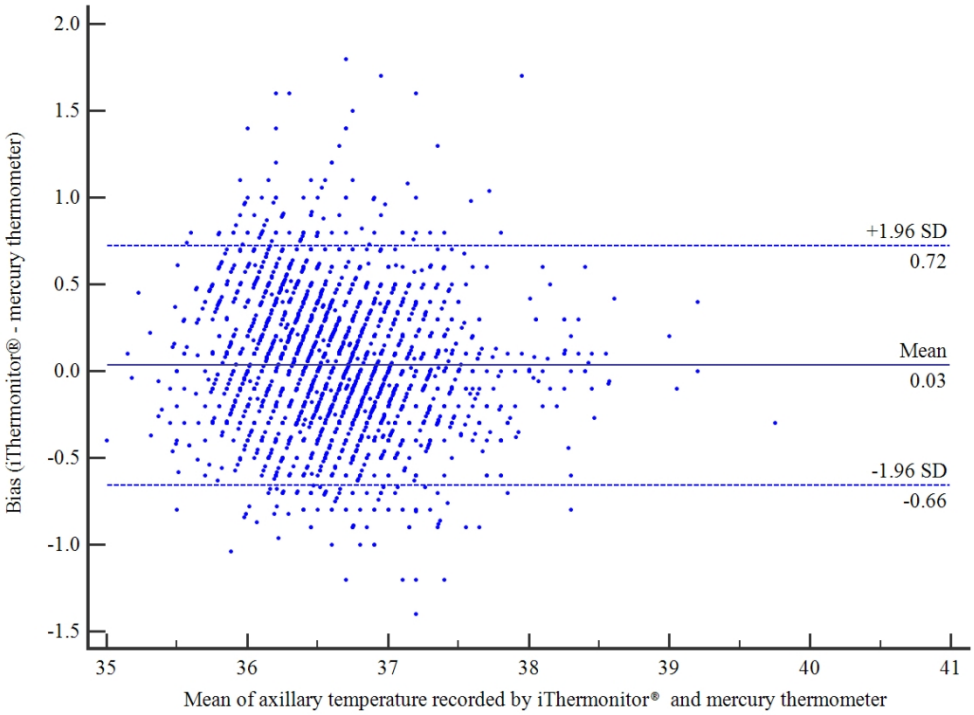


Figure 2. Bland-Altman plot of iThermonitor® axillary temperatures against mercury thermometer temperatures

423x317mm (72 x 72 DPI)

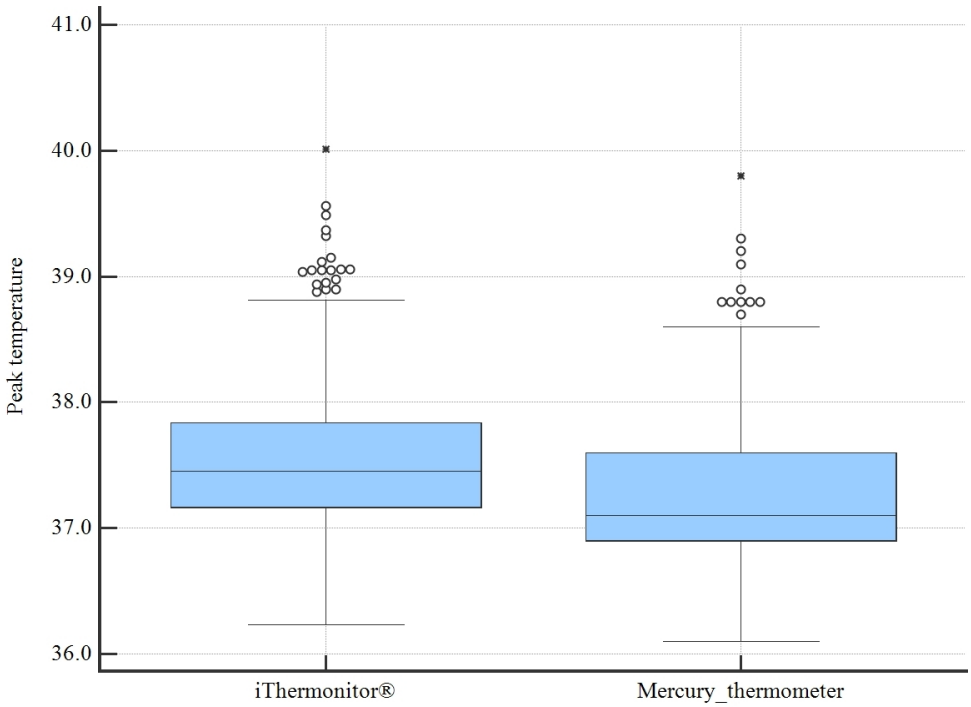
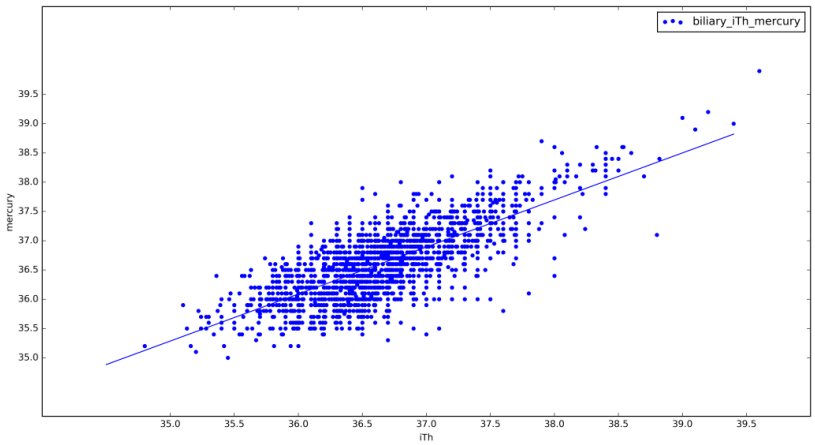


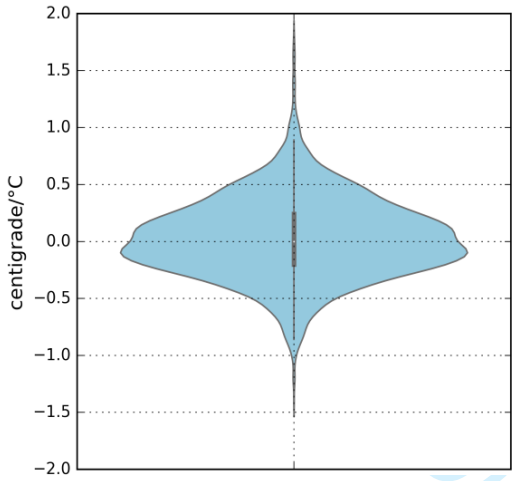
Figure 3. Peak temperature recorded by iThermonitor® and mercury thermometer.

101x76mm (300 x 300 DPI)

APPENDIX



Appendix figure A. Scatter plot of iThermonitor® axillary temperatures (x-axis) against mercury thermometer temperatures (y-axis). Pearson`s correlation coefficient r =0.755, P<0.001.



Appendix figure B. Distribution of the difference of axillary temperatures, calculated by iThermonitor® minus mercury thermometer.

Appendix table A. Comfort scores of the iThermonitor® wearing

Score	Description	Count (%)
1	Most uncomfortable and intolerable	0 (0)
2	Mild discomfort and tolerable	21 (19.3)
3	No obvious feeling and do not affect daily activities	81 (74.3)
4	Comfortable	6 (5.5)
5	Very comfortable	1 (0.9)

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	Page 3
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	Page 2-3
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	Page 4-5
	4	Study objectives and hypotheses	Page 6
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	Page 6
Participants	**6**	Eligibility criteria	Page 6
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	Page 6
	8	Where and when potentially eligible participants were identified (setting, location and dates)	Page 6
	9	Whether participants formed a consecutive, random or convenience series	Page 6
Test methods	**10a**	Index test, in sufficient detail to allow replication	Page 7-8
	10b	Reference standard, in sufficient detail to allow replication	Page 7-8
	11	Rationale for choosing the reference standard (if alternatives exist)	Page 8
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	Page 9
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	Page 9
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	Page 9
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	Page 9
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	Page 10
	15	How indeterminate index test or reference standard results were handled	Page 10
	16	How missing data on the index test and reference standard were handled	Page 11
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	Page 13
	18	Intended sample size and how it was determined	Page 10
RESULTS			
Participants	**19**	Flow of participants, using a diagram	Page 11
	20	Baseline demographic and clinical characteristics of participants	Page 11
	21a	Distribution of severity of disease in those with the target condition	Page 13
	21b	Distribution ofalternative diagnoses in those without the target condition	Page 13
	22	Time interval and any clinical interventions between index test and reference standard	Page 11
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	Page 14
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	Page 14
	25	Any adverse events from performing the index test or the reference standard	Page 15
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	Page 18
	27	Implications for practice, including the intended use and clinical role of the index test	Page 17-18
OTHER INFORMATION			
	28	Registration number and name of registry	Page 19
	29	Where the full study protocol can be accessed	Page 19
	30	Sources of funding and other support; role of funders	Page 19

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

For peer review only

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the abilityof one or more medical tests to correctly classify study participants as having a**target condition.**This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**.The reference standardisthe best available method for establishing the presence or absence ofthe target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against thoseof the reference standardcan be used to estimate the**sensitivity** of the index test(the proportion of participants *with* the target conditionwho have a positive index test), and its **specificity** (the proportion *without* the target conditionwho have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test.Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiveroperatingcharacteristic(ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage testis used before an existing test; an add-on test is used after an existing test.

Besides diagnosticaccuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis.The STARD list was not explicitly developed for these other outcomes, statistics, and studytypes, although mostSTARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.

For peer review only

BMJ Open

Evaluation of a wearable wireless device with artificial intelligence, iThermonitor® WT705, for continuous temperature monitoring for patients in surgical wards: a prospective comparative study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-039474.R1
Article Type:	Original research
Date Submitted by the Author:	08-Sep-2020
Complete List of Authors:	Liu, Yuwei; Sichuan University West China Hospital, Department of general surgery Liu, Changqing; Sichuan University West China Hospital, Operating room Gao, Min; Sichuan University West China Hospital, Department of biliary surgery Wang, Yan; Sichuan University West China Hospital, Department of biliary surgery Bai, Yangjing; Sichuan University West China Hospital, Department of cardiac and macrovascular surgery Xu, Ruihua; Sichuan University West China Hospital, Department of biliary surgery Gong, Renrong; Sichuan University West China Hospital, West China School of Nursing
Primary Subject Heading:	Nursing
Secondary Subject Heading:	Surgery
Keywords:	Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Adult surgery < SURGERY, Physiology < NATURAL SCIENCE DISCIPLINES, Telemedicine < BIOTECHNOLOGY & BIOINFORMATICS

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

Evaluation of a wearable wireless device with artificial intelligence, iThermonitor® WT705, for continuous temperature monitoring for patients in surgical wards: a prospective comparative study

Yuwei Liu,[1,2] Changqing Liu,[3] Min Gao,[2] Yan Wang,[2] Yangjing Bai,[4] Ruihua Xu,[1, 2] Renrong Gong[1,5]

[1] *West China School of Nursing, Sichuan University, Chengdu, P.R.China*

[2] *Department of biliary surgery, West China Hospital of Sichuan University, Chengdu, P.R.China*

[3] *Department of operating room, West China Hospital of Sichuan University, Chengdu, P.R.China*

[4] *Department of cardiac and macrovascular surgery, West China Hospital of Sichuan University, Chengdu, P.R.China*

[5] *Department of surgery, West China Hospital of Sichuan University, Chengdu, P.R.China*

Key words: Monitoring, Physiologic; Body Temperature; Wearable Electronic Devices

Correspondence to

Renrong Gong; gongrenrong@wchscu.cn

RuihuaXu; xuruihua@wchscu.cn

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

ABSTRACT

Objectives: To evaluate a new generation of a noninvasive wireless axillary thermometer with artificial intelligence, the iThermonitor® (WT705, Raiing Medical, Beijing, China), and to ascertain its feasibility for perioperative continuous body temperature monitoring in surgical patients.

Setting: Department of biliary surgery, operating room, and post-anesthesia care unit of a university teaching hospital in Chengdu, China.

Participants: A total of 526 adult surgical patients were consecutively enrolled.

Design: This was a prospective observational study. Axillary temperatures were continuously recorded by the iThermonitor® throughout the whole perioperative period. The temperatures of the contra lateral armpit were measured as references with mercury thermometers routinely at 8:00, 12:00, 16:00, and 20:00 every day.

Outcome measures: The outcomes were the accuracy and precision of the temperatures measured by the iThermonitor®, the validity to detect fever, and the feasibility of continuous wearing. Pairs of temperatures were evaluated by student *t*-test, Pearson`s correlation, and repeated-measured Bland-Altman plot.

Results: A total of 3621 pairs of body temperatures were obtained. The temperatures measured by the iThermonitor® agreed with those by mercury thermometers overall, with a mean difference of 0.03°C±0.36°C and a moderate correlation (r=0.755, $P<$ 0.001). The 95% limits of agreement ranged from -0.63°C to 0.73°C, with 5.11%of the differences outside the 95% limits of agreement. The Intra-class correlation coefficient was 0.753. The continuous temperature monitoring captured more fevers

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

than intermittent observation (117/526 vs. 91/526, $P < 0.001$), detected fever up to 4.35 hours earlier, and captured higher peak temperature (0.29°C ± 0.27°C, 95%CI: 0.26-0.31). All subjects felt that wearing the iThermonitor® was more or less comfortable and did not affect their daily activities.

Conclusions: The iThermonitor® is promising for continuous remote temperature monitoring in surgical patients. However, further developments still need for this device to improve its precision, especially for temperature detections in skinny patients and those with lower body temperature.

Strengths and limitations of this study

- A wearable wireless device with artificial intelligence, iThermonitor® WT705, was evaluated for continuous temperature monitoring in surgical patients.
- Axillary temperature was remotely monitored in different clinical scenarios throughout the whole perioperative period.
- More algorithm training and developments were still need for this device to improve its precision, especially for temperature detections in skinny patients and those with lower body temperature.
- Only axillary temperatures were detected for the evaluation of the iThermonitor®.
- The validity of the device were not tested in the ICU and / or in patients with compromised hemodynamic which might change the skin perfusion and temperature.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

INTRODUCTION

Body temperature is one of the most foundational vital signs of patients. Surgical patients are typically exposed to cold environments, administrations of unwarmed intravenous fluids, bacteria invasions and anesthetic drugs which may impair the thermoregulatory system,[1] leading to perturbations of body temperature. Accurately monitoring the body temperature is essential to prevent hypothermia, detect infectious complications for surgical patients.[2 3]

No ideal device has been developed yet[4] to continuously monitor body temperature across different clinical settings with satisfactory accuracy, availability and affordability.[5] Peripheral thermometers measuring temperature from the tympanic membrane, temporal artery, oral cavity, forehead, or other parts are considered to be not stable and accurate enough.[6-8] Inserts of a temperature probe to the esophageal, pulmonary artery, nasopharynx, rectum, or bladder could precisely and continuously detect the core temperature,[9] but these invasive devices increased the risk of infection and were only used for the patients in intensive care units and the surgical patients under anesthesia when necessary.[10 11] Until today, there are still urgent needs of thermometers to accurately and continuously monitor body temperature in clinical practice.

Wearable technology is changing the way that body temperatures have been measured and clinical cares have been performed.[12] In recent years, several wireless dermal wearable thermometers increased the feasibility of continuous body temperature monitoring outside of the critical care setting.[13] However, only a small proportion of

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

wearable devices have been CE marked (class Ⅱ or above) or FDA approved as medical devices.[14] There are still few options with convincing evidences to support their usages in clinical environments.[15] Wearable devices reflecting skin temperatures have been proved to have strong bias and poor correlations with oral temperatures,[16] or tympanic temperatures.[17] Zsuzsanna Balla *et al*.[13] found seven wireless dermal thermometers reflecting core temperatures through internet searching, and tested four of them (FeverSmart, iThermonitor®WT701, Quest Temp Sitter, and Thermochron iButton), which were commercially available. The results indicated that they were not reliable and accurate enough for most types of clinical researches, although the iThermonitor® WT701 systems had the least unsatisfactory correlation to rectal thermometers. Moreover, surgical patients were typically transferred between multiple units of cares (i.e., Surgical Wards, Operating Room, Post-anesthesia Care Unit). Because of the challenges during the scenarios changes such as tissue perfusion, physical activity, length of wearing time, and different compliances between awake patients and those under anesthesia, it remains unclear whether these devices are capable of continuous temperature monitoring across different clinical scenarios. In addition, battery life and internal storage space limited their application.

A new generation of a noninvasive wireless axillary thermometer, iThermonitor® WT705, was developed with advanced versions of machine learning algorithms for the needs of continuous remote monitoring of body temperature (including core temperature) in different clinical settings. Although the previous version of iThermonitor® has been reported with accuracy in representing core temperature

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

(distal esophagus temperature) in patients under anesthesia,[18] the performances of the latest version have not been tested in surgical wards yet. Therefore, we conducted a prospective study to assess the accuracy and feasibility for iThermonitor® WT705 to test body temperature in awake patients in surgical wards, and its potential for continuous monitoring of body temperature throughout the whole perioperative period at real clinical settings.

METHODS

Study design, subjects, and setting

This was a prospective comparative descriptive study to evaluate the iThermonitor® for continuous temperature monitoring by comparing it with mercury thermometers. Patients admitted to the department of biliary surgery in West China Hospital of Sichuan University were consecutively recruited in this study, from August to December 2019. The inclusion criteria was that the patient signed the informed consent. Patients with any impediment to wear the iThermonitor® in their axillae were excluded. Finally, 526 patients were enrolled and all of them signed their informed written consent.

Instruments

The study instruments were: (1) Wireless noninvasive dermal thermometers iThermonitor® (model WT705, Raiing Medical Company, Beijing, China), with accuracy of ±0.1°C (5°C-40°C). This is a battery-operated reusable electronic device with 30 days of battery life. The US Food and Drug Administration approved it as a

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

class Ⅱ medical device. The iThermonitor® was securely attached to the axilla (shaved if necessary) of patients with adhesive tapes provided by the manufacturer (Figure 1). The iThermonitor® WT705 was equipped with a more powerful chip developed from the previous version of WT701, which enables it to store more data with time and clinical scenario stamps. The sensor would record the axillary temperature once every 4 seconds, then transmit the raw data wirelessly to its associated signal repeater (cHub, Raiing Medical Company, Beijing, China). The cHub was attached to the bedside, with two versions of patented machine learning algorithms running inside. One version was for patients in the operating room to estimate core temperatures based on axillary temperatures. The other version was for patients in the ward and/or home to test the axillary temperatures, as used in the present study. The algorithms were designed to improve the accuracy of temperature tests by adjusting for possible interferences including anesthetics, daily activities, body posture changes, adhesive tape loose and/or ambient temperature, which could be pre-set as needed, together with the data output frequency. The average temperature per minute was transmitted wirelessly via Bluetooth or WiFi to the central computer in the nurse station. The dynamic temperature curves of all patients were visualized on the screen (online supplementary appendix figure 1). All the iThermonitor® sensors were proofread for accuracy according to the manufacturer`s standard before uses. (2)Mercury-in-glass thermometers (Riyue Medical Company, Chongqing, China), with accuracy of ±0.1°C, over the range of 35°C-42°C. Mercury thermometers were calibrated per week by comparing to a high-precision industrial

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

mercury thermometer (HX-290, Chuangji Instruments Company, Hebei, China). The mercury thermometers with temperature deviations greater than 0.2°C were not used. The iThermonitor® could be tested as a substitute choice for mercury thermometers since efforts had been for countries to find alternative methods to mercury thermometers.

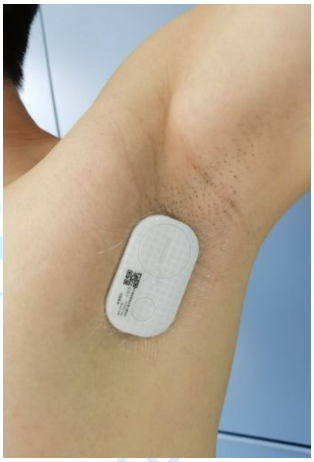


Figure 1 The wear of the iThermonitor® sensor in the axilla.

Data collection

Age, gender, and BMI were extracted from the medical records. The iThermonitor® sensor was attached to the dried and cleaned skin in the armpit region from the first day of admission to continuously record the body temperature throughout the whole perioperative period, except during a CT scan or a shower. The axillary skin was checked by registered nurses everyday for local skin allergy or other adverse reactions. Patients were asked to complete a questionnaire (online supplementary appendix Table 1) to evaluate the tolerability of the iThermonitor® on the day of discharge.

Among the temperatures measured by mercury thermometers routinely at 8:00 AM, 12:00 AM, 16:00 PM and 20:00 PM every day, those at 8:00 AM and 16:00 PM

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

were selected and paired with the temperatures tested by the iThermonitor®, because two specially trained registered nurses were assigned to measure the temperature at day shift. One nurse measured a temperature with a mercury thermometer and recorded it together the time of the measurement. The other nurse read the temperature tested by the iThermonitor® of the same patient at the same time from the central monitoring station. Meanwhile, the activity, state of consciousness, and armpits sweating were assessed and recorded.

Once the difference between the concurrent temperatures measured by the two devices was over than 1°C, the measurement was repeated immediately with a mercury thermometer, and the wear of the iThermonitor® was checked. When improper measurements or/and wears were confirmed, the corresponding sets of data were excluded. Only the differences were confirmed to really exist, the data were included for further analyses.

Main end points

The following endpoints are evaluated:

- Accuracy. The accuracy was expressed as the mean of the difference (also called bias) and the standard deviation, as calculated by temperature recorded by iThermonitor® minus that by mercury thermometers. A priori, an absolute difference of 0.5°C was considered to be clinically acceptable.[18 19]

- Precision. Precision (also called reliability) was tested by the 95% limits of agreement (95% LoA) and Intra-class correlation coefficient (ICC). As in previous studies,[17 18] the 95% LoA within ± 0.5°C and ICC greater than 0.7 wouldbe considered as well accepted precision.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

- Validity. Taking recordings of mercury thermometers as references, the validity of the iThermonitor® was evaluated by the sensitivity and specificity to detect fever. Axillary temperatures of 38°C were considered as a cut-off value of fever, according to the Brighton Collaboration Fever Working Group`s definition.[20] The peak temperatures recorded by the two methods were also compared. In addition, the first time for the iThermonitor® to detect the body temperature exceeded 38°C (lasting at least 5 minutes) was compared with the first time for mercury thermometers to detect fever.

- Feasibility. The feasibility of continuous temperature monitoring was assessed with a comfort score (online supplementary appendix table 1) and possible adverse events. Any adverse event, such as skin blisters or ulcers caused by iThermonitor®, or breakages of mercury thermometers, were included.

Statistics

Data were analyzed by Python (Version 3.5.1) and MedCalc (Version 19.1.3) software. Means and standard deviations of temperatures were calculated for the iThermonitor® and mercury thermometers. Quantitative data are expressed as mean ± standard deviation. The Student`s *t*-test for matched pairs, Pearson`s correlation analysis and repeated-measured Bland-Altman plot were used to evaluate the relationship between the two sets of temperatures. Possible factors associated with the accuracy were estimated by multiple linear regression analysis. $P< 0.05$ was considered statistically significant.

Sample Size Considerations

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

The sample size module for Bland-Altman plot of MedCalc software was conducted to estimate the sample size. An expected mean of differences was set as 0.03°C, with a standard deviation of 0.23°C, according to our previous pilot study. This calculation set a maximally allowed difference of 0.5°C (usually recognized as clinically significant),[17 18] a type Ⅰ error rate (Alpha) of 0.05, and a power of 80%. Finally, 3292 pairs of data were deemed sufficient to detect a difference between the iThermonitor® and mercury thermometers.

Patient and public involvement

No patients or members of the public were involved in the design of this study, the implementation of the study, or the result dissemination.

RESULTS

Characteristics of the patients

A sum of 526 patients were enrolled. All of them completed the study and were included in the final analysis. No missing values needed to be processed. Temperature curves with time stamps across different clinical scenarios were recorded for each patient (online supplementary appendix figure 1). Among the 526 patients, there were 197 (37.5%) males and 329 (62.5%) females, with an average age of 53.47±14.46 years (over the range of 15-86 years). The cumulative monitoring duration was 1768 days, with an average of 3.37±2.95 days for each patient. Patients were allowed to ambulate inside the hospital although most patients remained at their bedsides for daily activities, due to their surgeries. Room temperature was 24.2°C±1.3°C

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

(22°C-26°C), maintained by the central air conditioning system in the hospital. Demographic characteristics are shown in Table 1.

Table 1 Description of the patients

Variable	Summary
Patient Characteristics (N_1=526)	
Age (years)	53.47±14.46 (15-86)
Gender	
Male (%)	197 (37.5%)
Female (%)	329 (62.5%)
BMI (kg/m^2)	23.21±3.09 (15.21-37.78)
Average monitor duration (days)	3.37±2.95 (1-22)
Cumulative monitor duration (days)	1768
Patient status when measuring temperature (N_2=3621)	
Mobility	
Stay in bed	3490 (96.38%)
Off-bed activities	131 (3.62%)
Consciousness	
Awake	3572 (98.65%)
Sleep	49 (1.35%)
Sweating in the axilla	26 (0.72%)

Data are presented as n (%) for categorical variables or mean±SD for continuous variables. Abbreviation: BMI, body mass index. N_1: number of patients, N_2: number of paired temperature data sets.

Accuracy and precision of the iThermonitor®

A total of 3621 pairs of body temperatures were obtained. The mean temperature measured by the iThermonitor® was 36.61°C±0.49°C, ranged from 34.8°C to 39.6°C, while the mean temperature measured by mercury thermometers was 36.58°C±0.52°C, ranged from 35.0°C to 39.9°C. The mean of difference (bias) between the two

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

methods was 0.03±0.35°C, ranged from -1.40°C to 1.80°C. The biases within ±0.5°C, the clinically acceptable standard defined by a priori, were accounted for 87.68%, while 99.17% were within ±1.0°C. A scatter plot shows a relatively strong linear correlation (r=0.755, P<0.001) between the two groups of temperatures (online supplementary appendix figure 2). The Bland-Altman plot shows that the 95% limits of agreement were broader than the predefined range, with the upper limit at 0.73°C, and the lower limit at -0.63 °C. Meanwhile, 5.11% (185/3621) of the points were outside the 95% LoA (Figure 2). The intra-group correlation coefficient (ICC) was 0.753, indicating that the temperatures measured by the two methods were moderately agreed overall.

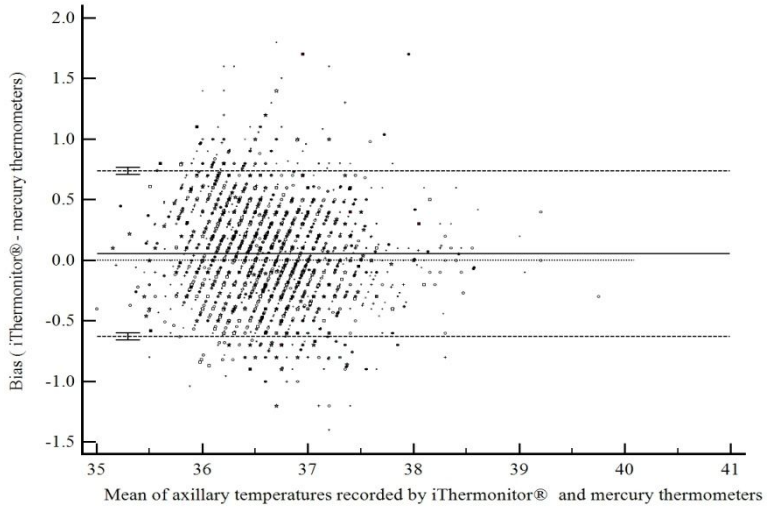


Figure 2 Repeated-measured Bland-Altman plot of iThermonitor® axillary temperatures against mercury thermometer temperatures. ICC=0.753, 95% limits of agreement were from -0.63°C to 0.73°C. 5.11% of the points were outside the 95%limits of agreement.

Factors associated with the difference between the two methods

Possible factors associated with the difference of readings between the iThermonitor® and mercury thermometers were evaluated by the multiple linear regression model (Table 2). Readings of mercury thermometers, BMI, and male gender were negatively

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

correlated with the difference. Age, mobility, consciousness, length of wearing, and sweating in the axilla were not significantly associated with the difference.

Table 2 Relationships of considered factors with the differences between iThermonitor® temperatures and mercury thermometer temperatures *

Factor considered	*B*	95%CI	*Bs*	*t*	*P* value
Reading of mercury thermometers	-0.295	-0.315 : -0.275	-0.433	-28.737	0.000
Female vs. male	-0.090	-0.111 : -0.069	-0.126	-8.453	0.000
BMI (kg/m^2)	-0.013	-0.016 : -0.009	-0.111	-7.399	0.000
Age (year)	0.001	0.000 : 0.001	0.028	1.870	0.062
Length of wearing (day)	0.001	-0.003 : 0.003	-0.002	-0.130	0.897
Stay in bed vs. off-bed activities	-0.020	-0.075 : 0.034	-0.011	-0.728	0.467
Sweating in the axilla	0.117	-0.005 : 0.238	0.028	1.884	0.060
Sleep vs. awake	-0.021	-0.110 : 0.067	-0.007	-0.466	0.641

Estimated by multiple linear regression analysis with the dependent variable as the difference of iThermonitor® temperatures minus mercury thermometer temperatures, with all the factors entered the regression. *R^2=0.215, F=123.474, P<0.001.

The effects of the associated factors were further evaluated on the differences between iThermonitor® temperatures and mercury thermometer temperatures (Table 3). The differences were significant when the readings of mercury thermometers were below 36.0°C or ≥38.0°C, but not significant between 36.0°C and 37.9°C. As to the effects of genders, the iThermonitor® readings were significantly lower in male subjects, but higher in female subjects than mercury thermometer readings. Besides, there were significant differences in patients with BMI<18.5 kg/m^2 or ranged from 18.5 kg/m^2 to 23.9 kg/m^2, but not ranged from 24.0 kg/m^2 to 27.9 kg/m^2 or ≥28.0 kg/m^2.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 3 The effects of associated factors on the temperature differences calculated by iThermonitor® readings minus mercury thermometer readings.

Factors	*N**	iThermonitor® (°C)	Mercury Thermometer(°C)	Bias (°C)	95%CI of Bias	*t*	*P*†
Readings of mercury thermometers							
Hypothermia (<36.0°C)	275	36.09±0.40	35.71±0.18	0.38±0.40	0.33 :0.43	15.839	0.000
Normal (36.0-37.9°C)	3285	36.63±0.42	36.62±0.42	0.01±0.33	-0.01 : 0.02	0.979	0.327
Fever (≥38.0°C)	61	38.11±0.54	38.27±0.36	-0.17±0.35	-0.26 : -0.08	-3.682	0.000
Gender							
Male	1480	36.57±0.45	36.60±0.49	-0.03±0.35	-0.05 : -0.02	-3.610	0.000
Female	2141	36.64±0.51	36.57±0.54	0.08±0.35	0.06 : 0.09	10.003	0.000
BMI(kg/m^2)							
Low (<18.5)	187	36.74±0.42	36.60±0.47	0.14±0.34	0.09 : 0.19	5.368	0.000
Normal (18.5-23.9)	2156	36.66±0.50	36.63±0.53	0.04±0.34	0.02: 0.05	4.869	0.000
Overweight (24.0-27.9)	1018	36.51±0.46	36.50±0.50	0.01±0.36	-0.02 : 0.03	0.653	0.514
Obesity (≥28.0)	260	36.51±0.45	36.50±0.46	0.01±0.43	-0.04 : 0.07	0.529	0.598
Total	3621	36.61±0.49	36.58±0.52	0.03±0.35	0.02: 0.04	5.326	0.000

* Number of paired temperature data sets. †Estimated by Student`s *t* test for matched pairs.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Comparisons of the two methods in detecting fever

Among the 526 patients, 117 of them were detected to have fever by continuous recordings with the iThermonitor®, while only 91 patients were detected by intermittent readings with mercury thermometers (Chi-square test for paired sample, $P<0.001$). A total of 124 patients were observed to have fevers by the iThermonitor® or mercury thermometers, and 84 subjects by both of the two methods.

Comparisons of average peak temperatures

The average peak temperature captured by intermittent measurements with mercury thermometers was 37.26°C ± 0.56°C, whereas the continuous monitoring with the iThermonitor® detected a higher peak temperature of 37.55°C ± 0.59°C. A mild but statistically significant difference of 0.29°C ± 0.27°C (range from-0.45°C to 1.26°C, 95%CI: 0.26°C-0.31°C) was noted between the peak temperatures recorded by the two methods (Figure 3).

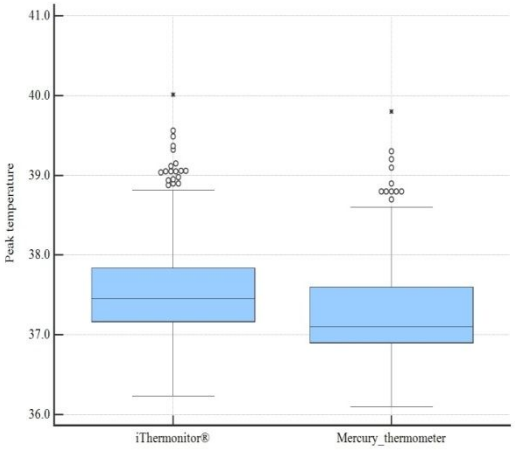


Figure 3 Peak temperatures recorded by the iThermonitor® and mercury thermometers.

Comparison of the earliest time to detect fevers

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

The continuous monitoring with the iThermonitor® detected fevers earlier than intermittent measurement with mercury thermometers. A mean time interval of 4.35 hours was observed, with a minimum difference of -0.92 hours and a maximum difference of 25.34 hours.

Feasibility of continuous temperature monitoring

During 1768 monitoring days, 4.37% (23/526) of the patients complained of slight itching during the wearing period. Local skin redness was observed in two patients, but all eviated after the iThermonitor® was moved to the contra-lateral axilla. The comfort score from 1 to 5 was used to evaluate the feasibility, with 1 meaning most uncomfortable and intolerable, 5 meaning very comfortable. In 109 patients enrolled in August and September 2019, 21 of them selected the score of 2, 81 patients selected the score of 3, six patients selected the score of 4, and one patient selected the score of 5. All subjects felt that wearing the iThermonitor® was more or less comfortable and did not affect their daily activities, indicating well acceptable compliance of the iThermonitor®.

DISCUSSION

This was a prospective study to evaluate the performance of iThermonitor® WT705, a wireless dermal temperature sensor, for continuous temperature monitoring in surgical patients. The iThermonitor® was selected because its performance had not been tested before for continuous temperature monitoring in surgical patients, although it was one

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

of the best performed devices due to relatively high sensitivity and robustness in reflecting core temperature among the available wireless dermal thermometers.[13]

Principal Findings in accuracy and precision

The iThermonitor® was compared with mercury thermometers. An overall small bias (0.03°C±0.35°C, Table 3) and moderate correlation (Figure 2) were observed between the two devices. Nevertheless, the analysis of 95% limits of agreement indicated that the comparisons with the bias less than ±0.5°C only accounted for 87.68%. Exploring the factors showed that this discrepancy was associated with the readings of mercury thermometers, genders and BMI. The largest bias (0.38°C±0.40°C) existed when the readings of mercury thermometers were below 36°C, with the readings of the iThermonitor® higher than mercury thermometers (Table 3). Errors might exist in manual measurements with mercury thermometers if the patient didn`t maintain the proper measurement posture for enough time, leading to false body temperature readings below 36°C. A more strict supervision and repeated measurements may improve the results when the readings of mercury thermometers are below 36°C. Moreover, not every reading of mercury thermometers was double-checked. This might also lead to the biases. Besides, a significantly larger bias was also noticed in skinny patients (BMI<18.5 kg/m^2), but no significant differences in overweight or obesity patients (Table 3). This result supports the earlier findings by Rubia-Rubia[21], who tested four axillary thermometers and found that the difference of readings between axillary thermometers and pulmonary artery temperatures increased when the weight decreased. A possible explanation was that skinny patients might be difficult

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

to have the thermometers properly attached to the axilla, leading to the errors in the readings of axillary thermometers. Moreover, larger biases were observed in female patients (Table 3). Evidences indicated that there were sex differences in thermoregulatory mechanisms. Lu *et al*[22] reported higher oral temperatures in older women than those in men. Rubia-Rubia *et al*[21] found the male sex increased the bias between temperatures measured by digital axillary thermometer and core temperatures, although they were unable to find a satisfactory explanation.

Compared To Similar Studies

Different reference standards and algorithm versions resulted in inconsistent accuracy of iThermonitor® in previous reports. Pei *et al*.[18] tested the intraoperative version of iThermonitor® WT701, and found that axillary temperature recordings well represented core temperatures in adults under anesthesia with a mean difference of only 0.14°C±0.26°C (esophageal minus axillary) and 95% limits of agreement from -0.38°C to 0.66°C. It was also reported that iThermonitor® WT701 rendered lower average temperatures than rectal temperatures (bias -0.77°C±0.53°C).[13] In addition, the iThermonitor® was highly easy to be used to continuously monitor children`s body temperatures at home.[23] In the present study, the novel version of iThermonitor® WT705 was selected and the algorithms were pre-setted for ward/home use. The results demonstrated that the iThermonitor® could be used to continuously monitor temperatures in surgical patients although improvements were still needed. Moreover, all subjects in the study felt that wearing the iThermonitor® was more or less comfortable and did not affect daily activities. This work adds evidence to support the

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

applicability of iThermonitor® in surgical wards as a continuation or extension of intraoperative temperature monitoring.

Clinical Implications

The advantages and feasibility of continuous vital signs monitoring in general wards are attracting increasing attention.[10 24] As to body temperature monitoring, detecting the time to have fevers and the peak of fever are important for diagnoses and clinical decision-making. Dakappa *et al*.[11] noted higher peak temperature with a significant difference of 1.52°C, using a continuous tympanic temperature recording device (TherCom®) than using mercury thermometers three times a day. Another research reported 180 minutes earlier by wearable digital thermometers (TremTraq, with a battery life of 72 hours) to detect increases in body temperature than by standard monitoring.[24] As claimed by a previous study, the iThermonitor® was marginally superior in following the individual trends than in assessing absolute temperatures.[13] The present study demonstrated that continuous temperature monitoring with the iThermonitor® was better at capturing the peak of fever, and could detect fever 4.35 hours earlier than intermittent temperature monitoring. These findings may have important therapeutic implications. Furthermore, maintaining the continuity of the body temperature monitoring across different clinical scenarios helps to install a real-world database of the patient's perioperative body temperature, which would provide more information in exploring the regularity of perioperative temperature fluctuation.

The mercury-containing thermometer has been widely used for hundreds of years

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

because of its stable performances in reflecting temperature.[25] However, these glass-based thermometers are fragile to leak mercury out, which is one of the top ten chemicals of major public health concern.[26] Although the World Health Organization has called for the phase out of mercury fever thermometers by 2020,[26] they are still widely used in many countries.[5 27 28] Gaps are evident in practices on promoting mercury-free thermometers. The lack of an ideal alternative device for temperature measurement is an important reason. Using the iThermonitor® instead of mercury thermometers to reduce medical mercury emissions is beneficial to the patients, health personnel and public health.[27 28]

Limitations

There were several limitations in the present study. First, only axillary temperatures were detected using the iThermonitor® and mercury thermometers. Adding a set of core body temperatures would help better understand the validity of the device to monitor body temperatures. Second, the validity of the device were not tested in the ICU and/or in patients with compromised hemodynamic which might change the skin perfusion and temperature. Third, armpits were randomly selected for temperature measurements and the difference of armpit temperature was not taken into account between the dominant and non-dominant arm, it might not exist.[21] Besides, an axillary temperature of 38℃ was set as a fever reference in the present study. However, the most convincing cut-off value remains unclear, due to varying definitions of fever.[7 29 30]

CONCLUSIONS

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

It is promising for the iThermonitor® to continuously monitor temperatures in surgical patients. This device can improve fever detection by dynamically reflecting the individual trends of body temperature throughout the whole perioperative period. However, more algorithm training is still needed for this device to improve its accuracy, especially when it is used in hypothermia or fever patients, female patients, or skinny or even normal weight patients.

Contributors: Yuwei Liu participated in the conception, statistical analysis, and manuscript draft preparation. Changqing Liu was in charge of body temperature management in the OR and PACU. Min Gao and Yan Wang performed temperature measurements. Yangjing Bai participated in the design of the study. Ruihua Xu administrated the project. Renrong Gong conceived, designed and managed the running of the study, interpreted the data, and revised and finally approved the manuscript. All authors read and approved the manuscript.

Funding: This study was funded by the Sichuan University West China Nursing Discipline Development Special Fund Project (NO.HXHL19031).

Competing interest: None declared.

Ethics approval: Ethical approval was obtained from the Biomedical Research Ethics Committee of West China Hospital of Sichuan University on July 4, 2019 (No. 2019-447).

Clinical trial registration: The study was registered at Chinese Clinical Trail Registry (ChiCTR1900024549) on July 5, 2019. The full study protocol can be accessed at*www.chictr.org.cn*.

Data sharing statement: Data are available upon reasonable request.

Open access: This is an open access article distributed in accordance with theCreative Commons Attribution Non Commercial (CC BY-NC4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on differentterms, provided the original work isproperly cited, appropriate credit is given, any changes made indicated, and the useis non-commercial. See: http:// creativecommons. org/ licenses/ by- nc/ 4. 0/

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

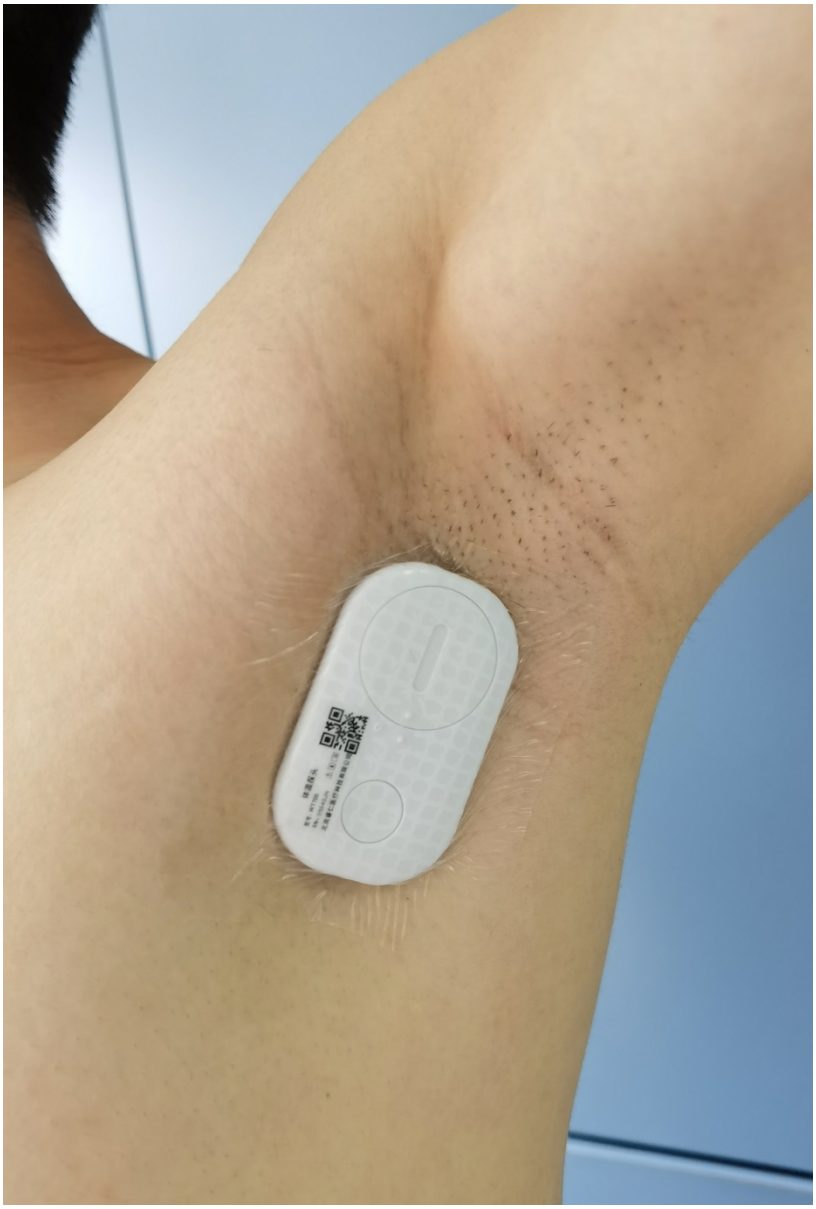
REFERENCES

1. Sessler DI. Opioids and postoperative shivering. *J Clin Anesth* 2016;31:42-43.
2. Middleton S, McElduff P, Drury P, et al. Vital sign monitoring following stroke associated with 90-day independence: A secondary analysis of the QASC cluster randomized trial. *Int J Nurs Stud* 2019;89:72-79.
3. Larach MG, Brandom BW, Allen GC, et al. Malignant Hyperthermia Deaths Related to Inadequate Temperature Monitoring, 2007–2012. *Anesthesia & Analgesia* 2014;119(6):1359-66.
4. Sollai S, Dani C, Berti E, et al. Performance of a non-contact infrared thermometer in healthy newborns. *BMJ Open* 2016;6(3):e008695.
5. Halder N, Peshin SS, Pandey RM, et al. Awareness assessment of harmful effects of mercury in a health care set-up in India: A survey-based study. *Toxicol Ind Health* 2015;31(12):1144-51.
6. Opersteny E, Anderson H, Bates J, et al. Precision, Sensitivity and Patient Preference of Non-Invasive Thermometers in a Pediatric Surgical Acute Care Setting. *J Pediatr Nurs* 2017;35:36-41.
7. Niven DJ, Gaudet JE, Laupland KB, et al. Accuracy of Peripheral Thermometers for Estimating Temperature. *Annals of Internal Medicine* 2015;163(10)
8. Geijer H, Udumyan R, Lohse G, et al. Temperature measurements with a temporal scanner: systematic review and meta-analysis. *BMJ Open* 2016;6(3):e009509.
9. 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. *Journal of PeriAnesthesia Nursing* 2010;25(2):71-78.
10. Downey CL, Chapman S, Randell R, et al. The impact of continuous versus intermittent vital signs monitoring in hospitals: A systematic review and narrative synthesis. *Int J Nurs Stud* 2018;84:19-27.
11. Dakappa PH, Bhat GK, Bolumbu G, et al. Comparison of Conventional Mercury Thermometer and Continuous TherCom((R)) Temperature Recording in Hospitalized Patients. *J Clin Diagn Res* 2016;10(9):OC43-OC46.
12. Liao Y, Thompson C, Peterson S, et al. The Future of Wearable Technologies and Remote Monitoring in Health Care. *Am Soc Clin Oncol Educ Book* 2019;39:115-21.
13. Zsuzsanna Balla H, Theodorsson E, Strom JO. Evaluation of commercial, wireless dermal thermometers for surrogate measurements of core temperature. *Scand J Clin Lab Invest* 2019;79(1-2):1-6.
14. Muzny M, Henriksen A, Giordanengo A, et al. Wearable sensors with possibilities for data exchange: Analyzing status and needs of different actors in mobile health monitoring systems. *Int J Med Inform* 2020;133:104017.
15. Breteler MJMM, Huizinga E, van Loon K, et al. Reliability of wireless monitoring using a wearable patch sensor in high-risk surgical patients at a step-down unit in the Netherlands: a clinical validation study. *BMJ Open* 2018;8(2):e020162.
16. Izmailova ES, McLean IL, Bhatia G, et al. Evaluation of Wearable Digital Devices in a

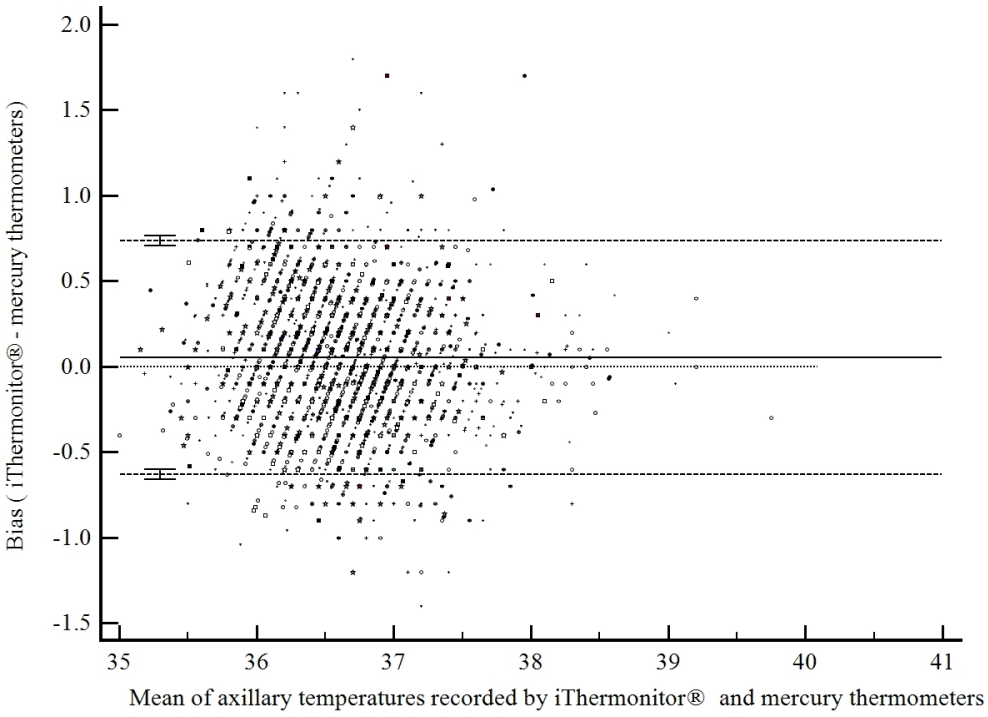
For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Phase I Clinical Trial. *Clin Transl Sci* 2019;12(3):247-56.
17. Downey C, Ng S, Jayne D, et al. Reliability of a wearable wireless patch for continuous remote monitoring of vital signs in patients recovering from major surgery: a clinical validation study from the TRaCINg trial. *BMJ Open* 2019;9(8):e031150.
18. Pei L, Huang Y, Mao G, et al. Axillary Temperature, as Recorded by the iThermonitor WT701, Well Represents Core Temperature in Adults Having Noncardiac Surgery. *Anesth Analg* 2018;126(3):833-38.
19. Lu SH, Dai YT, Yen CJ. The effects of measurement site and ambient temperature on body temperature values in healthy older adults: a cross-sectional comparative study. *Int J Nurs Stud* 2009;46(11):1415-22.
20. Marcy SM, Kohl KS, Dagan R, et al. Fever as an adverse event following immunization: case definition and guidelines of data collection, analysis, and presentation. *Vaccine* 2004;22(5-6):551-56.
21. 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(7):872-80.
22. Lu SH, Dai YT. Normal body temperature and the effects of age, sex, ambient temperature and body mass index on normal oral temperature: a prospective, comparative study. *Int J Nurs Stud* 2009;46(5):661-8.
23. Kakarmath SS, de Redon E, Centi AJ, et al. Assessing the Usability of an Automated Continuous Temperature Monitoring Device (iThermonitor) in Pediatric Patients: Non-Randomized Pilot Study. *JMIR Pediatr Parent* 2018;1(2):e10804.
24. Abbasi J. Wearable Digital Thermometer Improves Fever Detection. *JAMA* 2017;318(6):510.
25. Wright WF, Mackowiak PA. Origin, Evolution and Clinical Application of the Thermometer. *Am J Med Sci* 2016;351(5):526-34.
26. WHO. WHO calls for the phase out of mercury fever thermometers and blood pressure measuring devices by 2020. Available at: https://www.who.int/news-room/detail/11-10-2013-who-calls-for-the-phase-out-of-mercury-fever-thermometers-and-blood-pressure-measuring-devices-by-2020 [accessed 11 October 2013.
27. Li P, Yang Y, Xiong W. Impacts of Mercury Pollution Controls on Atmospheric Mercury Concentration and Occupational Mercury Exposure in a Hospital. *Biol Trace Elem Res* 2015;168(2):330-4.
28. Chadli-Debbiche A, Allani R, Jouini R, et al. Mercurial hospital waste: management, becoming and recommendations. *Tunis Med* 2018;96(5):273-80.
29. Apa H, Gozmen S, Bayram N, et al. Clinical accuracy of tympanic thermometer and noncontact infrared skin thermometer in pediatric practice: an alternative for axillary digital thermometer. *Pediatr Emerg Care* 2013;29(9):992-7.
30. Wright WF, Auwaerter PG. Fever and Fever of Unknown Origin: Review, Recent Advances, and Lingering Dogma. *Open Forum Infect Dis* 2020;7(5):ofaa132.

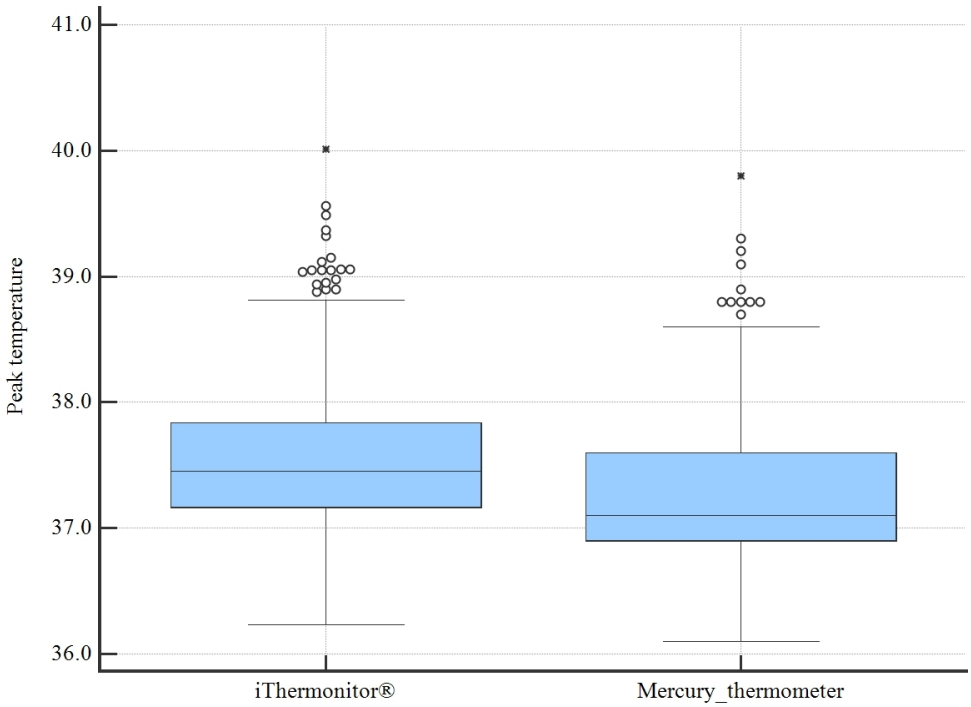
For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



914x1354mm (72 x 72 DPI)

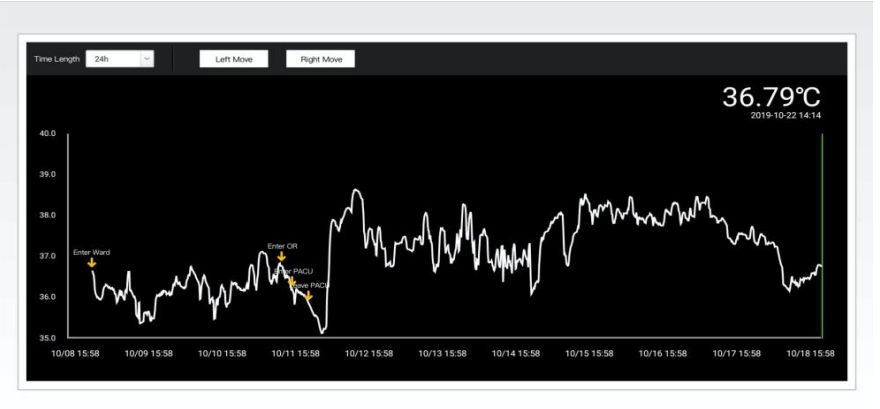


101x76mm (300 x 300 DPI)

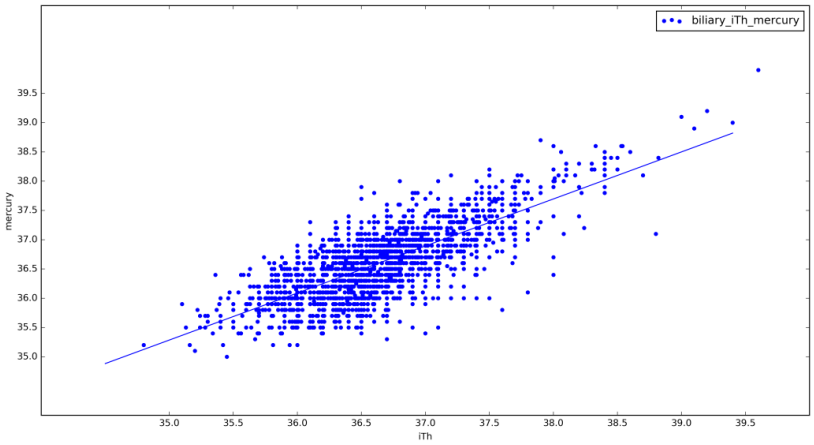


101x76mm (300 x 300 DPI)

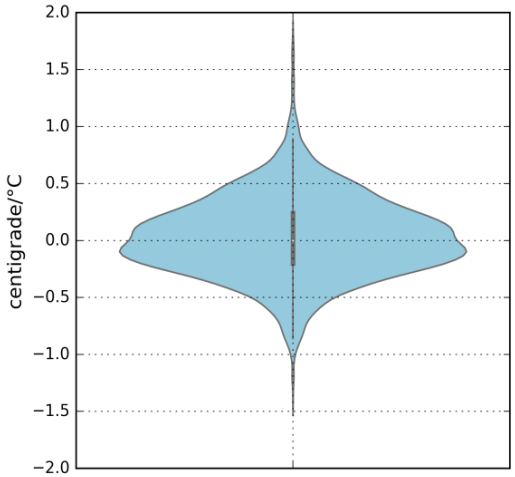
Supplementary Materials



Appendix figure 1 The temperature curve with time stamps across different clinical scenarios.



Appendix figure 2 Scatter plot of iThermonitor® axillary temperatures (x-axis) against mercury thermometertemperatures (y-axis). Pearson`s correlation coefficient r =0.755, P<0.001.



Appendix figure 3 Distribution of the difference of axillary temperatures, calculated by iThermonitor® minus mercury thermometer.

Appendix table 1 The questionnaire to evaluate the feasibility of theiThermonitor® (N=109)

Selections	Score	Patients` Options (%)
I felt very uncomfortable and could not bear to wear it.	1	0 (0)
I felt slight discomfort, but it's tolerable.	2	21 (19.3)
I didn't have noticeable feelings of wearing it.	3	81 (74.3)
I felt comfortable with it.	4	6 (5.5)
I felt very comfortable with it.	5	1 (0.9)

For peer review only

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	Page 2
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	Page 2-3
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	Page 4-5
	4	Study objectives and hypotheses	Page 6
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	Page 6
Participants	**6**	Eligibility criteria	Page 6
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	Page 6
	8	Where and when potentially eligible participants were identified (setting, location and dates)	Page 6
	9	Whether participants formed a consecutive, random or convenience series	Page 6
Test methods	**10a**	Index test, in sufficient detail to allow replication	Page 7-8
	10b	Reference standard, in sufficient detail to allow replication	Page 7-8
	11	Rationale for choosing the reference standard (if alternatives exist)	Page 8
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	Page 10
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	Page 10
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	Page 9
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	Page 9
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	Page 9
	15	How indeterminate index test or reference standard results were handled	Page 10
	16	How missing data on the index test and reference standard were handled	Page 11
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	Page 13
	18	Intended sample size and how it was determined	Page 10
RESULTS			
Participants	**19**	Flow of participants, using a diagram	Page 11
	20	Baseline demographic and clinical characteristics of participants	Page 11-12
	21a	Distribution of severity of disease in those with the target condition	Page 15
	21b	Distribution ofalternative diagnoses in those without the target condition	Page 15
	22	Time interval and any clinical interventions between index test and reference standard	Page 11
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	Page 16
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	Page 16
	25	Any adverse events from performing the index test or the reference standard	Page 17
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	Page 21
	27	Implications for practice, including the intended use and clinical role of the index test	Page 20
OTHER INFORMATION			
	28	Registration number and name of registry	Page 22
	29	Where the full study protocol can be accessed	Page 22
	30	Sources of funding and other support; role of funders	Page 22

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

For peer review only

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the abilityof one or more medical tests to correctly classify study participants as having a**target condition.**This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**.The reference standardisthe best available method for establishing the presence or absence ofthe target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against thoseof the reference standardcan be used to estimate the**sensitivity** of the index test(the proportion of participants *with* the target conditionwho have a positive index test), and its **specificity** (the proportion *without* the target conditionwho have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test.Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiveroperatingcharacteristic(ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage testis used before an existing test; an add-on test is used after an existing test.

Besides diagnosticaccuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis.The STARD list was not explicitly developed for these other outcomes, statistics, and studytypes, although mostSTARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.

For peer review only

BMJ Open

Evaluation of a wearable wireless device with artificial intelligence, iThermonitor® WT705, for continuous temperature monitoring for patients in surgical wards: a prospective comparative study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-039474.R2
Article Type:	Original research
Date Submitted by the Author:	10-Oct-2020
Complete List of Authors:	Liu, Yuwei; Sichuan University West China Hospital, Department of general surgery Liu, Changqing; Sichuan University West China Hospital, Operating room Gao, Min; Sichuan University West China Hospital, Department of biliary surgery Wang, Yan; Sichuan University West China Hospital, Department of biliary surgery Bai, Yangjing; Sichuan University West China Hospital, Department of cardiac and macrovascular surgery Xu, Ruihua; Sichuan University West China Hospital, Department of biliary surgery Gong, Renrong; Sichuan University West China Hospital, West China School of Nursing
Primary Subject Heading:	Nursing
Secondary Subject Heading:	Surgery
Keywords:	Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Adult surgery < SURGERY, Physiology < NATURAL SCIENCE DISCIPLINES, Telemedicine < BIOTECHNOLOGY & BIOINFORMATICS

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

Evaluation of a wearable wireless device with artificial intelligence, iThermonitor® WT705, for continuous temperature monitoring for patients in surgical wards: a prospective comparative study

Yuwei Liu,[1,2] Changqing Liu,[3] Min Gao,[2] Yan Wang,[2] Yangjing Bai,[4] Ruihua Xu,[1, 2] Renrong Gong[1,5]

[1] *West China School of Nursing, Sichuan University, Chengdu, P.R.China*

[2] *Department of Biliary Surgery, West China Hospital of Sichuan University, Chengdu, P.R.China*

[3] *Operating Room of Anesthesia Surgery Center, West China Hospital of Sichuan University, Chengdu, P.R.China*

[4] *Department of Cardiac And Macrovascular Surgery, West China Hospital of Sichuan University, Chengdu, P.R.China*

[5] *Department of Surgery, West China Hospital of Sichuan University, Chengdu, P.R.China*

Key words: Monitoring, Physiologic; Body Temperature; Wearable Electronic Devices

Correspondence to

Renrong Gong; gongrenrong@wchscu.cn

RuihuaXu; xuruihua@wchscu.cn

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

ABSTRACT

Objectives: To evaluate a new-generation non-invasive wireless axillary thermometer with artificial intelligence, iThermonitor® (WT705, Raiing Medical, Beijing, China), and to ascertain its feasibility for perioperative continuous body temperature monitoring in surgical patients.

Setting: Departments of Biliary Surgery and Operating Room and the post-anaesthesia care unit of a university teaching hospital in Chengdu, China.

Participants: A total of 526 adult surgical patients were consecutively enrolled.

Design: This was a prospective observational study. Axillary temperatures were continuously recorded with iThermonitor® throughout the whole perioperative period. The temperatures of the contralateral armpit were measured with mercury thermometers at 8:00, 12:00, 16:00, and 20:00 every day and were used as references.

Outcome measures: The outcomes were the accuracy and precision of the temperatures measured with iThermonitor®, the validity to detect fever, and the feasibility of continuous wear. Pairs of temperatures were evaluated with Student's *t*-test, Pearson's correlation, and repeated-measures Bland-Altman plot.

Results: A total of 3621 pairs of body temperatures were obtained. The temperatures measured with iThermonitor® agreed with those measured with the mercury thermometers overall, with a mean difference of 0.03°C±0.36°C and a moderate correlation (r=0.755, P<0.001). The 95% limits of agreement (95% LoA) ranged from -0.63°C to 0.73°C, with 5.11% of the differences outside the 95% LoA. The intraclass correlation coefficient was 0.753. Continuous temperature monitoring captured more

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

fevers than intermittent observation (117/526 vs. 91/526, P<0.001), detected fever up to 4.35 hours earlier, and captured a higher peak temperature (0.29°C±0.27°C, 95% confidence interval: 0.26-0.31). All subjects felt that wearing iThermonitor® was more or less comfortable and did not affect their daily activities.

Conclusions: iThermonitor® is promising for continuous remote temperature monitoring in surgical patients. However, further developments are still needed to improve the precision of this device, especially for temperature detection in underweight patients and those with lower body temperature.

Strengths and limitations of this study

- A wearable wireless device with artificial intelligence, iThermonitor® WT705, was evaluated for continuous temperature monitoring in surgical patients.
- Axillary temperature was remotely monitored in different clinical scenarios throughout the whole perioperative period.
- More algorithm training and developments are still needed to improve the precision of this device, especially for temperature detection in underweight patients and those with lower body temperature.
- Only axillary temperatures were detected for the evaluation of iThermonitor®.
- The validity of the device was not tested in the intensive care unit (ICU) or in patients with compromised haemodynamics which might change skin perfusion and temperature.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

INTRODUCTION

Body temperature is one of the most foundational vital signs of patients. Surgical patients are typically exposed to cold environments, administrations of unwarmed intravenous fluids, bacterial invasions and anaesthetic drugs that may impair the thermoregulatory system,[1] leading to perturbations in body temperature. Accurately monitoring body temperature is essential for preventing hypothermia and detecting infectious complications in surgical patients.[2 3]

No ideal device has been developed yet[4] to continuously monitor body temperature across different clinical settings with satisfactory accuracy, availability and affordability.[5] Peripheral thermometers measuring temperature from the tympanic membrane, temporal artery, oral cavity, forehead, or other parts are not considered to be stable or accurate enough.[6-8] Inserts of a temperature probe to the oesophageal, pulmonary artery, nasopharynx, rectum, or bladder can precisely and continuously detect the core temperature,[9] but these invasive devices increase the risk of infection and are only used for patients in intensive care units (ICU) and surgical patients under anaesthesia when necessary.[10 11] Currently, there is still an urgent need for thermometers that accurately and continuously monitor body temperature in clinical practice.

Wearable technology is changing the way that body temperatures have been measured and clinical care has been performed.[12] In recent years, several wireless dermal wearable thermometers have increased the feasibility of continuous body temperature monitoring outside of the critical care setting.[13] However, only a small proportion of

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

wearable devices have a Conformité Européenne (CE) marking class IIor above, or approval from the US Food and Drug Administration (FDA) as medical devices.[14] There are still few options with convincing evidence to support their usage in clinical environments.[15] Wearable devices reflecting skin temperatures have been provent to have strong bias and poor correlations with oral temperatures,[16] or tympanic temperatures.[17] Zsuzsanna Balla *et al*.[13] found seven wireless dermal thermometers reflecting core temperatures through internet searching, and tested four (FeverSmart, iThermonitor®WT701, Quest Temp Sitter, and Thermochron iButton) that were commercially available. The results indicated that they were not reliable or accurate enough for most types of clinical studies, although the iThermonitor® WT701 system had the least unsatisfactory correlation to rectal thermometers. Moreover, surgical patients were typically transferred between multiple units of care (i.e., surgical ward, operating room, post-anaesthesia care unit). Because of the challenges during scenario changes, such as tissue perfusion, physical activity, length of wearing time, and different compliances between awake patients and those under anaesthesia, it remains unclear whether these devices are capable of continuous temperature monitoring across different clinical scenarios. In addition, battery life and internal storage space limit their application.

A new generation of the non-invasive wireless axillary thermometer iThermonitor® WT705 was developed with advanced versions of machine learning algorithms for continuous remote monitoring of body temperature (including core temperature) in different clinical settings. Although the previous version of iThermonitor® has been

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

reported to accurately represent core temperature (distal oesophagus temperature) in patients under anaesthesia,[18] the performance of the latest version has not yet been tested in surgical wards. Therefore, we conducted a prospective study to assess the accuracy and feasibility of iThermonitor® WT705 for testing body temperature in awake patients in surgical wards, and its potential for continuous monitoring of body temperature throughout the whole perioperative period at real clinical settings.

METHODS

Study design, subjects, and setting

This was a prospective comparative descriptive study to evaluate iThermonitor® for continuous temperature monitoring by comparing it with mercury thermometers. Patients admitted to the Department of Biliary Surgery at West China Hospital of Sichuan University were consecutively recruited for this study, from August to December 2019. The inclusion criterion was that the patient signed an informed consent form. Patients with any impediment to wearing the iThermonitor® in their axillae were excluded. Finally, 526 patients were enrolled, and all signed informed written consent forms.

Instruments

The study instruments were as follows: (1) Wireless non-invasive dermal thermometer iThermonitor® (model WT705, Raiing Medical Company, Beijing, China), with an accuracy of ±0.1°C (5°C-40°C). This is a battery-operated reusable electronic device with 30 days of battery life. The FDA approved it as a class II

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

medical device. iThermonitor® was securely attached to the axilla (shaved if necessary) of patients with adhesive tape provided by the manufacturer (Figure 1). iThermonitor® WT705 was equipped with a more powerful chip developed from the previous version of WT701, which enables it to store more data with time and clinical scenario stamps. The sensor recorded the axillary temperature once every 4 seconds and then transmitted the raw data wirelessly to its associated signal repeater (cHub, Raiing Medical Company, Beijing, China). The cHub was attached to the bedside, with two versions of patented machine learning algorithms running inside. One version was for patients in the operating room to estimate core temperatures based on axillary temperatures. The other version was for patients in the ward and/or home to test the axillary temperatures, as used in the present study. The algorithms were designed to improve the accuracy of temperature tests by adjusting for possible interference, including anaesthetics, daily activities, body posture changes, loose adhesive tape and/or ambient temperature, which could be preset as needed, together with the data output frequency. The average temperature per minute was transmitted wirelessly via Bluetooth or WiFi to the central computer in the nurse station. The dynamic temperature curves of all patients were visualized on the screen (online supplementary appendix Figure 1). All iThermonitor® sensors were proofread for accuracy according to the manufacturer`s standard before use. (2)Mercury-in-glass thermometers (Riyue Medical Company, Chongqing, China) with an accuracy of ±0.1°C, over the range of 35°C-42°C. Mercury thermometers were calibrated each week by comparison to a high-precision industrial mercury thermometer (HX-290, Chuangji Instruments

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Company, Hebei, China). Mercury thermometers with temperature deviations greater than 0.2°C were not used. iThermonitor® could be tested as a substitute choice for mercury thermometers since efforts have been made by countries to find alternative methods to mercury thermometers.

Data collection

Age, sex, and body mass index (BMI) were extracted from the medical records. The iThermonitor® sensor was attached to the dried and cleaned skin in the armpit region on the first day of admission to continuously record the body temperature throughout the whole perioperative period, except during a computed tomography (CT) scan or a shower. The axillary skin was checked by registered nurses every day for local skin allergies or other adverse reactions. Patients were asked to complete a questionnaire (online supplementary appendix Table 1) to evaluate the tolerability of iThermonitor® on the day of discharge.

Temperature was measured with mercury thermometers at 8:00 ante meridiem (AM), 12:00 AM, 16:00 post meridiem (PM) and 20:00 PM every day, the 8:00 AM and 16:00 PM measurements were selected and paired with the temperatures measured with iThermonitor® because two specially trained registered nurses were assigned to take these measurements during the day shift. One nurse measured the temperature with a mercury thermometer and recorded it together with the time of the measurement. The other nurse read the temperature detected with iThermonitor® of the same patient at the same time from the central monitoring station. Meanwhile, activity, state of consciousness, and armpit sweating were assessed and

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

recorded.

Once the difference between the concurrent temperatures measured with the two devices was over 1°C, the measurement was repeated immediately with a mercury thermometer, and the fit of iThermonitor® was checked. When improper measurements or/and fit were confirmed, the corresponding sets of data were excluded. Only the differences were confirmed to truly exist, and the data were included for further analyses.

Main end points

The following endpoints were evaluated:

- Accuracy. The accuracy was expressed as the mean of the difference (also called bias) and the standard deviation, as calculated by the temperature recorded with iThermonitor® minus that recorded with a mercury thermometer. A priori, an absolute difference of 0.5°C was considered to be clinically acceptable.[18 19]

- Precision. Precision (also called reliability) was tested by the 95% limits of agreement (95% LoA) and intraclass correlation coefficient (ICC). As in previous studies,[17 18] the 95% LoA within ±0.5°C and ICC greater than 0.7 were considered well accepted precision.

- Validity. Taking recordings of mercury thermometers as references, validity of the iThermonitor® was evaluated by the abilityto detect fever. Axillary temperatures of 38°C were considered a cut-off value of fever, according to the Brighton Collaboration Fever Working Group's definition.[20] The peak temperatures recorded by the two methods were also compared. In addition, the first time iThermonitor® indicated body temperatures exceeding 38°C (lasting at least 5 min) was compared

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

with the first time for a mercury thermometer to detect fever.

- Feasibility. The feasibility of continuous temperature monitoring was assessed with a comfort score (online supplementary appendix Table 1) and possible adverse events. Any adverse events, such as skin blisters or ulcers caused by iThermonitor®, or the breaking of a mercury thermometer, were included.

Statistics

Means and standard deviations of temperatures were calculated for iThermonitor® and mercury thermometers. Quantitative data are expressed as the mean±standard deviation. Student's *t*-test for matched pairs, Pearson's correlation analysis and repeated-measures Bland-Altman plot were used to evaluate the relationship between the two sets of temperatures. The calculations of the Bland-Altman plot with multiple measurements per subject were performed as described by Bland *et al.*[21], and the confidence intervals of 95% LoA were estimated as described by Zou.[22] Possible factors associated with the accuracy were estimated by multiple linear regression analysis, in which the differences between iThermonitor® and mercury thermometer readings were considered the dependent variable and all the factors considered were the explanatory variables. $P< 0.05$ was considered statistically significant. Data were analysed with Python (Version 3.5.1), and MedCalc (Version 19.1.3) software for repeated-measures Bland-Altman plot.

Sample size considerations

The sample size module for the Bland-Altman plot of MedCalc software was used to estimate the sample size. An expected mean of differences was set as 0.03°C, with a

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

standard deviation of 0.23°C, according to our previous pilot study. This calculation set a maximally allowed difference of 0.5°C (usually recognized as clinically significant),[17 18] a type I error rate (Alpha) of 0.05, and a power of 80%. Finally, 3292 pairs of data were deemed sufficient to detect a difference between iThermonitor® and the mercury thermometers.

Patient and public involvement

No patients or members of the public were involved in the design of this study, the implementation of the study, or the dissemination of the results.

RESULTS

Characteristics of the patients

A total of 526 patients were enrolled. All of them completed the study and were included in the final analysis. No missing values needed to be processed. Each patient wore iThermonitor® temperature sensor from admission to discharge, with an average of 3.37±2.95 days. The cumulative monitoring duration was 1768 days. Temperature curves with time stamps across different clinical scenarios were recorded for each patient (online supplementary appendix Figure 1). Among the 526 patients, there were 197 (37.5%) males and 329 (62.5%) females, with an average age of 53.47±14.46 years (over the range of 15-86 years). Patients were allowed to ambulate inside the hospital although most patients remained at their bedsides for daily activities, due to their surgeries. The room temperature was 24.2°C±1.3°C (22°C-26°C), maintained by the central air conditioning system in the hospital. Demographic characteristics are

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

shown in Table 1.

Table 1 Description of the patients

Variable	Summary
Patient Characteristics (N_1=526)	
Age (years)	53.47±14.46 (15-86)
Sex	
Male (%)	197 (37.5%)
Female (%)	329 (62.5%)
BMI (kg/m^2)	23.21±3.09 (15.21-37.78)
Average monitor duration (days)	3.37±2.95 (1-22)
Cumulative monitor duration (days)	1768
Patient status when measuring temperature (N_2=3621)	
Mobility	
Bedridden	3490 (96.38%)
Off-bed activities	131 (3.62%)
Consciousness	
Awake	3572 (98.65%)
Sleep	49 (1.35%)
Sweating in the axilla	26 (0.72%)

Data are presented as n (%) for categorical variables or the mean±SD for continuous variables. Abbreviations: BMI: body mass index, N_1: number of patients, N_2: number of paired temperature data sets.

Accuracy and precision of iThermonitor®

A total of 3621 pairs of body temperatures were obtained. The mean temperature measured with iThermonitor® was 36.61°C±0.49°C, ranging from 34.8°C to 39.6°C, while the mean temperature measured with the mercury thermometers was 36.58°C±0.52°C, ranging from 35.0°C to 39.9°C. The mean of difference (bias) between the two methods was 0.03±0.35°C, ranging from -1.40°C to 1.80°C. The

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

biases within ±0.5°C, the clinically acceptable standard defined by a priori, accounted for 87.68%, while 99.17% were within ±1.0°C. A scatter plot shows a relatively strong linear correlation (r=0.755, P<0.001) between the two groups of temperatures (online supplementary appendix Figure 2). The Bland-Altman plot shows that the 95% LoA were broader than the predefined range, with the upper limit at 0.73°C, and the lower limit at -0.63°C. Meanwhile, 5.11% (185/3621) of the points were outside the 95% LoA (Figure 2). The ICC was 0.753, indicating that the temperatures measured by the two methods moderately agreed overall.

Factors associated with the difference between the two methods

Possible factors associated with the difference in readings between iThermonitor® and the mercury thermometers were evaluated by the multiple linear regression model (Table 2). Readings of mercury thermometers, BMI, and male sex were negatively correlated with the difference. Age, mobility, consciousness, length of wearing, and sweating in the axilla were not significantly associated with the difference.

Table 2 Relationships of considered factors with the differences between iThermonitor® temperatures and mercury thermometer temperatures *

Factor considered	*B*	95% CI	*Bs*	*t*	*P* value
Reading of mercury thermometers	-0.295	-0.315 : -0.275	-0.433	-28.737	0.000
Sex (female vs. male)	-0.090	-0.111 : -0.069	-0.126	-8.453	0.000
BMI (kg/m^2)	-0.013	-0.016 : -0.009	-0.111	-7.399	0.000
Age (year)	0.001	0.000 : 0.001	0.028	1.870	0.062
Length of wearing (day)	0.001	-0.003 : 0.003	-0.002	-0.130	0.897
Mobility (bedridden vs. off-bed activities)	-0.020	-0.075 : 0.034	-0.011	-0.728	0.467

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Sweating in the axilla	0.117	-0.005 : 0.238	0.028	1.884	0.060
Consciousness (sleep vs. awake)	-0.021	-0.110 : 0.067	-0.007	-0.466	0.641

Estimated by multiple linear regression analysis with the dependent variable as the difference of iThermonitor® temperatures minus mercury thermometer temperatures, with all the factors entered into the regression. *B*, linear regression coefficient. 95% CI, 95% confidence interval for linear regression coefficient. *Bs*, standardized coefficient. *Squared variation coefficient R^2=0.215, F=123.474, P<0.001.

The effects of the associated factors were further evaluated for the differences between iThermonitor® temperatures and mercury thermometer temperatures (Table 3). The differences were significant when the readings of the mercury thermometers were below 36.0°C or ≥38.0°C, but not significant between 36.0°C and 37.9°C. Regarding the effects of sex, the iThermonitor® readings compared to the mercury thermometer readings were significantly lower in male subjects but higher in female subjects. In addition, there were significant differences in patients with BMI<18.5 kg/m^2 or ranging from 18.5 kg/m^2 to 23.9 kg/m^2, but not in those with BMI ranging from 24.0 kg/m^2 to 27.9 kg/m^2 or ≥28.0 kg/m^2.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 3 The effects of associated factors on the temperature differences calculated with iThermonitor® readings minus mercury thermometer readings.

Factors	*N**	iThermonitor® (°C)	Mercury Thermometer(°C)	Bias (°C)	95% CI of Bias	*t*	*P*†
Readings of mercury thermometers							
Hypothermia (<36.0°C)	275	36.09±0.40	35.71±0.18	0.38±0.40	0.33 : 0.43	15.839	0.000
Normal (36.0-37.9°C)	3285	36.63±0.42	36.62±0.42	0.01±0.33	-0.01 : 0.02	0.979	0.327
Fever (≥38.0°C)	61	38.11±0.54	38.27±0.36	-0.17±0.35	-0.26 : -0.08	-3.682	0.000
Sex							
Male	1480	36.57±0.45	36.60±0.49	-0.03±0.35	-0.05 : -0.02	-3.610	0.000
Female	2141	36.64±0.51	36.57±0.54	0.08±0.35	0.06 : 0.09	10.003	0.000
BMI(kg/m^2)							
Low (<18.5)	187	36.74±0.42	36.60±0.47	0.14±0.34	0.09 : 0.19	5.368	0.000
Normal (18.5-23.9)	2156	36.66±0.50	36.63±0.53	0.04±0.34	0.02: 0.05	4.869	0.000
Overweight (24.0-27.9)	1018	36.51±0.46	36.50±0.50	0.01±0.36	-0.02 : 0.03	0.653	0.514
Obese (≥28.0)	260	36.51±0.45	36.50±0.46	0.01±0.43	-0.04 : 0.07	0.529	0.598
Total	3621	36.61±0.49	36.58±0.52	0.03±0.35	0.02: 0.04	5.326	0.000

* Number of paired temperature data sets. †Estimated by Student's *t* test for matched pairs.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Comparisons of the two methods in detecting fever

Among the 526 patients, 117 were detected to have fever by continuous recordings with iThermonitor®, while only 91 patients were detected by intermittent readings with mercury thermometers (chi-square test for paired sample, $P<0.001$). A total of 124 patients were observed to have fevers with iThermonitor® or the mercury thermometers, and 84 subjects were observed by both methods.

Comparisons of average peak temperatures

The average peak temperature captured by intermittent measurements with mercury thermometers was 37.26°C±0.56°C, whereas continuous monitoring with iThermonitor® detected a higher peak temperature of 37.55°C±0.59°C. A mild but statistically significant difference of 0.29°C±0.27°C (range from-0.45°C to 1.26°C, 95% CI: 0.26°C-0.31°C) was noted between the peak temperatures recorded by the two methods (Figure 3).

Comparison of the earliest time to detect fevers

Continuous monitoring with iThermonitor® detected fevers earlier than intermittent measurement with mercury thermometers. A mean time interval of 4.35 hours was observed, with a minimum difference of -0.92 hours and a maximum difference of 25.34 hours.

Feasibility of continuous temperature monitoring

During 1768 monitoring days, 4.37% (23/526) of the patients complained of slight

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

itching during the wearing period. Local skin redness was observed in two patients but resolved after iThermonitor® was moved to the contralateral axilla. A comfort score from 1 to 5 was used to evaluate the feasibility of wearing iThermonitor®, with 1 meaning most uncomfortable and intolerable and 5 meaning very comfortable. Of the 109 patients enrolled in August and September 2019, 21 selected a score of 2, 81 selected a score of 3, six selected a score of 4, and one selected a score of 5. All subjects felt that wearing iThermonitor® was more or less comfortable and did not affect their daily activities, indicating acceptable compliance of iThermonitor®.

DISCUSSION

This was a prospective study to evaluate the performance of iThermonitor® WT705, a wireless dermal temperature sensor, for continuous temperature monitoring in surgical patients. iThermonitor® was selected because its performance had not been tested before for continuous temperature monitoring in surgical patients, although it has been one of the best performing devices among the available wireless dermal thermometers due to its relatively high sensitivity and robustness in reflecting core temperature.[13]

Principal findings in accuracy and precision

iThermonitor® was compared with mercury thermometers. An overall small bias (0.03°C±0.35°C, Table 3) and moderate correlation (Figure 2) were observed between the two devices. Nevertheless, the analysis of 95% LoA indicated that the comparisons with a bias less than ±0.5°C only accounted for 87.68%. This

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

discrepancy was shown to be associated with the readings of the mercury thermometers, sex and BMI. The largest bias (0.38°C±0.40°C) existed when the readings of the mercury thermometers were below 36°C, with the readings of iThermonitor® higher than those of the mercury thermometers (Table 3). Errors might exist in manual measurements with mercury thermometers if the patient did not maintain the proper measurement posture for enough time, leading to false body temperature readings below 36°C. Stricter supervision and repeated measurements may improve the results when the readings of mercury thermometers are below 36°C. On the other hand, the temperature output of iThermonitor® may also be inaccurate, but it is difficult to distinguish which device, or both, is the main cause of the deviation. Moreover, not every reading of the mercury thermometers was double-checked. This might also lead to biases. In addition, a significantly larger bias was also noticed in underweight patients (BMI<18.5 kg/m^2), but no significant differences were observed in overweight or obese patients (Table 3). This result supports the earlier findings of Rubia-Rubia[23], who tested four axillary thermometers and found that the difference in readings between axillary thermometers and pulmonary artery temperatures increased with decreasing weight. A possible explanation is that underweight patients might have difficulty having thermometers properly attached to the axilla, leading to errors in the readings of axillary thermometers. Moreover, larger biases were observed in female patients (Table 3). Evidence has indicated that there are sex differences in thermoregulatory mechanisms. Lu *et al*[24] reported higher oral temperatures in older women than in men. Rubia-Rubia

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

et al[23] found that male sex increased the bias between temperatures measured by digital axillary thermometer and core temperatures, although they were unable to find a satisfactory explanation.

Compared to similar studies

Different reference standards and algorithm versions have resulted in inconsistent accuracy of iThermonitor® in previous reports. Pei *et al*.[18] tested the intraoperative version of iThermonitor® WT701, and found that axillary temperature recordings well represented core temperatures in adults under anaesthesia with a mean difference of only 0.14°C±0.26°C (oesophageal minus axillary) and 95% LoA from -0.38°C to 0.66°C. It was also reported that iThermonitor® WT701 rendered lower average temperatures than rectal temperatures (bias -0.77°C±0.53°C).[13] In addition, iThermonitor® was highly easy to use to continuously monitor children's body temperatures at home.[25] In the present study, the novel version of iThermonitor® WT705 was selected and the algorithms were preset for ward/home use. The results demonstrated that iThermonitor® could be used to continuously monitor temperatures in surgical patients although improvements are still needed. Moreover, all subjects in the study felt that wearing iThermonitor® was more or less comfortable and did not affect daily activities. This work adds evidence to support the applicability of iThermonitor® in surgical wards as a continuation or extension of intraoperative temperature monitoring.

Clinical implications

The advantages and feasibility of continuous vital sign monitoring in general wards

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

are attracting increasing attention.[10 26] Regarding body temperature monitoring, detecting the time to have fevers and the peak of fever are important for diagnoses and clinical decision-making. Dakappa *et al*.[11] noted a higher peak temperature with a significant difference of 1.52°C using a continuous tympanic temperature recording device (TherCom®) than using mercury thermometers three times a day. Another study reported that a wearable digital thermometer (TremTraq, with a battery life of 72 hours) detected increases in body temperature 180 min earlier than the standard monitoring strategy.[26] As claimed in a previous study, iThermonitor® was marginally superior in following the individual trends than in assessing absolute temperatures.[13] The present study demonstrated that continuous temperature monitoring with iThermonitor® was better at capturing the peak of fever, and could detect fever 4.35 hours earlier than intermittent temperature monitoring. These findings may have important therapeutic implications. Furthermore, maintaining the continuity of body temperature monitoring across different clinical scenarios helps to install a real-world database of a patient's perioperative body temperature, which would provide more information in exploring the regularity of perioperative temperature fluctuation.

Mercury-containing thermometers have been widely used for hundreds of years because of their stable performance in reflecting temperature.[27] However, these glass-based thermometers are fragile, and mercury, which is one of the top ten chemicals of major public health concern, can leak.[28] Although the World Health Organization has called for the phase out of mercury fever thermometers by 2020,[28] they are still widely used in many countries.[5 29 30] Gaps are evident in practices on

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

promoting mercury-free thermometers. The lack of an ideal alternative device for temperature measurement is an important reason. Using iThermonitor® instead of mercury thermometers to reduce medical mercury emissions is beneficial to patients, health personnel and public health.[29 30]

Limitations

There were several limitations in the present study. First, only axillary temperatures were detected using iThermonitor® and mercury thermometers. Adding a set of core body temperatures would help better understand the validity of the device for monitoring body temperatures. Second, the validity of the device was not tested in the ICU or in patients with compromised haemodynamic which might change skin perfusion and temperature. Third, armpits were randomly selected for temperature measurements, and the difference in armpit temperature between the dominant and non-dominant arms, which might not exist,[23] was not taken into account. In addition, an axillary temperature of 38°C was set as a fever reference in the present study. However, the most convincing cut-off value remains unclear, due to varying definitions of fever.[7 31 32]

CONCLUSIONS

iThermonitor® is a promising device for the continuous monitoring of temperature in surgical patients. This device can improve fever detection by dynamically reflecting the individual trends in body temperature throughout the whole perioperative period. However, more algorithm training is still needed to improve the accuracy of this device, especially when it is used in hypothermia or fever patients,

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

female patients, or underweight or even normal-weight patients.

Contributors: Yuwei Liu participated in the conception, statistical analysis, and manuscript draft preparation. Changqing Liu was in charge of body temperature management in the OR and PACU. Min Gao and Yan Wang performed the temperature measurements. Yangjing Bai participated in the design of the study. Ruihua Xu administered the project. Renrong Gong conceived, designed and managed the running of the study, interpreted the data, and revised and finally approved the manuscript. All authors read and approved the manuscript.

Funding: This study was funded by the Sichuan University West China Nursing Discipline Development Special Fund Project (No.HXHL19031).

Competing interest: None declared.

Ethics approval: Ethical approval was obtained from the Biomedical Research Ethics Committee of West China Hospital of Sichuan University on July 4, 2019 (No. 2019-447).

Clinical trial registration: The study was registered at the Chinese Clinical Trail Registry (ChiCTR1900024549) on July 5, 2019. The full study protocol can be accessed at*www.chictr.org.cn*.

Data sharing statement: Data are available upon reasonable request.

Open access: This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC4.0) 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, appropriate credit is given, any changes made indicated, and the useis non-commercial. See: http:// creativecommons. org/ licenses/ by- nc/ 4. 0/

REFERENCES

1. Sessler DI. Opioids and postoperative shivering. *J Clin Anesth* 2016;31:42-43.
2. Middleton S, McElduff P, Drury P, et al. Vital sign monitoring following stroke associated with 90-day independence: A secondary analysis of the QASC cluster randomized trial. *Int J Nurs Stud* 2019;89:72-79.
3. Larach MG, Brandom BW, Allen GC, et al. Malignant Hyperthermia Deaths Related to Inadequate Temperature Monitoring, 2007–2012. *Anesthesia & Analgesia* 2014;119(6):1359-66.
4. Sollai S, Dani C, Berti E, et al. Performance of a non-contact infrared thermometer in

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

healthy newborns. *BMJ Open* 2016;6(3):e008695.
5. Halder N, Peshin SS, Pandey RM, et al. Awareness assessment of harmful effects of mercury in a health care set-up in India: A survey-based study. *Toxicol Ind Health* 2015;31(12):1144-51.
6. Opersteny E, Anderson H, Bates J, et al. Precision, Sensitivity and Patient Preference of Non-Invasive Thermometers in a Pediatric Surgical Acute Care Setting. *J Pediatr Nurs* 2017;35:36-41.
7. Niven DJ, Gaudet JE, Laupland KB, et al. Accuracy of Peripheral Thermometers for Estimating Temperature. *Annals of Internal Medicine* 2015;163(10)
8. Geijer H, Udumyan R, Lohse G, et al. Temperature measurements with a temporal scanner: systematic review and meta-analysis. *BMJ Open* 2016;6(3):e009509.
9. 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. *Journal of PeriAnesthesia Nursing* 2010;25(2):71-78.
10. Downey CL, Chapman S, Randell R, et al. The impact of continuous versus intermittent vital signs monitoring in hospitals: A systematic review and narrative synthesis. *Int J Nurs Stud* 2018;84:19-27.
11. Dakappa PH, Bhat GK, Bolumbu G, et al. Comparison of Conventional Mercury Thermometer and Continuous TherCom((R)) Temperature Recording in Hospitalized Patients. *J Clin Diagn Res* 2016;10(9):OC43-OC46.
12. Liao Y, Thompson C, Peterson S, et al. The Future of Wearable Technologies and Remote Monitoring in Health Care. *Am Soc Clin Oncol Educ Book* 2019;39:115-21.
13. Zsuzsanna Balla H, Theodorsson E, Strom JO. Evaluation of commercial, wireless dermal thermometers for surrogate measurements of core temperature. *Scand J Clin Lab Invest* 2019;79(1-2):1-6.
14. Muzny M, Henriksen A, Giordanengo A, et al. Wearable sensors with possibilities for data exchange: Analyzing status and needs of different actors in mobile health monitoring systems. *Int J Med Inform* 2020;133:104017.
15. Breteler MJMM, Huizinga E, van Loon K, et al. Reliability of wireless monitoring using a wearable patch sensor in high-risk surgical patients at a step-down unit in the Netherlands: a clinical validation study. *BMJ Open* 2018;8(2):e020162.
16. Izmailova ES, McLean IL, Bhatia G, et al. Evaluation of Wearable Digital Devices in a Phase I Clinical Trial. *Clin Transl Sci* 2019;12(3):247-56.
17. Downey C, Ng S, Jayne D, et al. Reliability of a wearable wireless patch for continuous remote monitoring of vital signs in patients recovering from major surgery: a clinical validation study from the TRaCINg trial. *BMJ Open* 2019;9(8):e031150.
18. Pei L, Huang Y, Mao G, et al. Axillary Temperature, as Recorded by the iThermonitor WT701, Well Represents Core Temperature in Adults Having Noncardiac Surgery. *Anesth Analg* 2018;126(3):833-38.
19. Lu SH, Dai YT, Yen CJ. The effects of measurement site and ambient temperature on body temperature values in healthy older adults: a cross-sectional comparative study. *Int J Nurs Stud* 2009;46(11):1415-22.
20. Marcy SM, Kohl KS, Dagan R, et al. Fever as an adverse event following immunization:

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

case definition and guidelines of data collection, analysis, and presentation. *Vaccine* 2004;22(5-6):551-56.

21. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *Journal of biopharmaceutical statistics* 2007;17(4):571-82.
22. Zou GY. Confidence interval estimation for the Bland-Altman limits of agreement with multiple observations per individual. *Statistical methods in medical research* 2013;22(6):630-42.
23. 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(7):872-80.
24. Lu SH, Dai YT. Normal body temperature and the effects of age, sex, ambient temperature and body mass index on normal oral temperature: a prospective, comparative study. *Int J Nurs Stud* 2009;46(5):661-8.
25. Kakarmath SS, de Redon E, Centi AJ, et al. Assessing the Usability of an Automated Continuous Temperature Monitoring Device (iThermonitor) in Pediatric Patients: Non-Randomized Pilot Study. *JMIR Pediatr Parent* 2018;1(2):e10804.
26. Abbasi J. Wearable Digital Thermometer Improves Fever Detection. *JAMA* 2017;318(6):510.
27. Wright WF, Mackowiak PA. Origin, Evolution and Clinical Application of the Thermometer. *Am J Med Sci* 2016;351(5):526-34.
28. WHO. WHO calls for the phase out of mercury fever thermometers and blood pressure measuring devices by 2020. Available at: https://www.who.int/news-room/detail/11-10-2013-who-calls-for-the-phase-out-of-mercury-fever-thermometers-and-blood-pressure-measuring-devices-by-2020 [accessed 11 October 2013].
29. Li P, Yang Y, Xiong W. Impacts of Mercury Pollution Controls on Atmospheric Mercury Concentration and Occupational Mercury Exposure in a Hospital. *Biol Trace Elem Res* 2015;168(2):330-4.
30. Chadli-Debbiche A, Allani R, Jouini R, et al. Mercurial hospital waste: management, becoming and recommendations. *Tunis Med* 2018;96(5):273-80.
31. Apa H, Gozmen S, Bayram N, et al. Clinical accuracy of tympanic thermometer and noncontact infrared skin thermometer in pediatric practice: an alternative for axillary digital thermometer. *Pediatr Emerg Care* 2013;29(9):992-7.
32. Wright WF, Auwaerter PG. Fever and Fever of Unknown Origin: Review, Recent Advances, and Lingering Dogma. *Open Forum Infect Dis* 2020;7(5):ofaa132.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

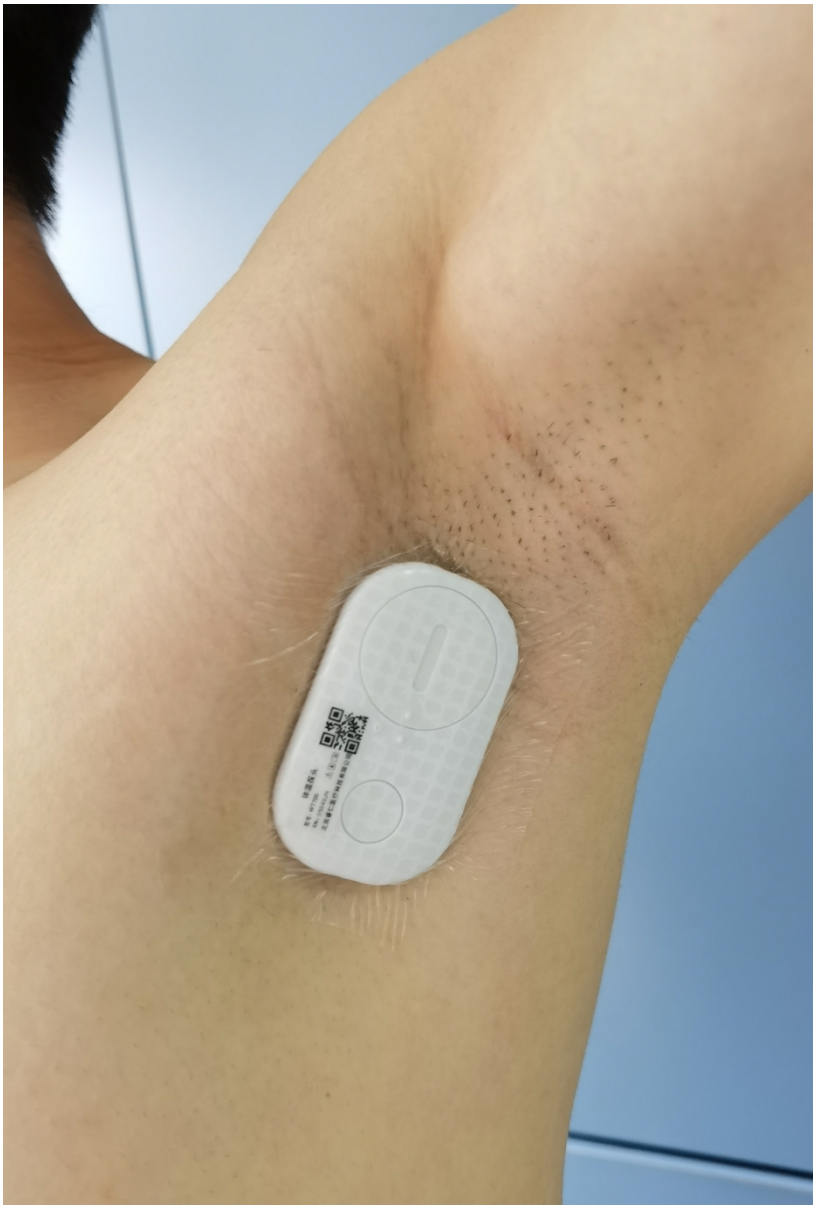
Figure 1 The position of the iThermonitor® sensor in the axilla.

Figure 2 Repeated-measures Bland-Altman plot of iThermonitor® axillary temperatures against mercury thermometer temperatures. ICC=0.753, 95% LoA were from -0.63°C to 0.73°C. A total of 5.11% of the points were outside the 95% LoA.

Figure 3 Peak temperatures recorded with iThermonitor® and the mercury thermometers.

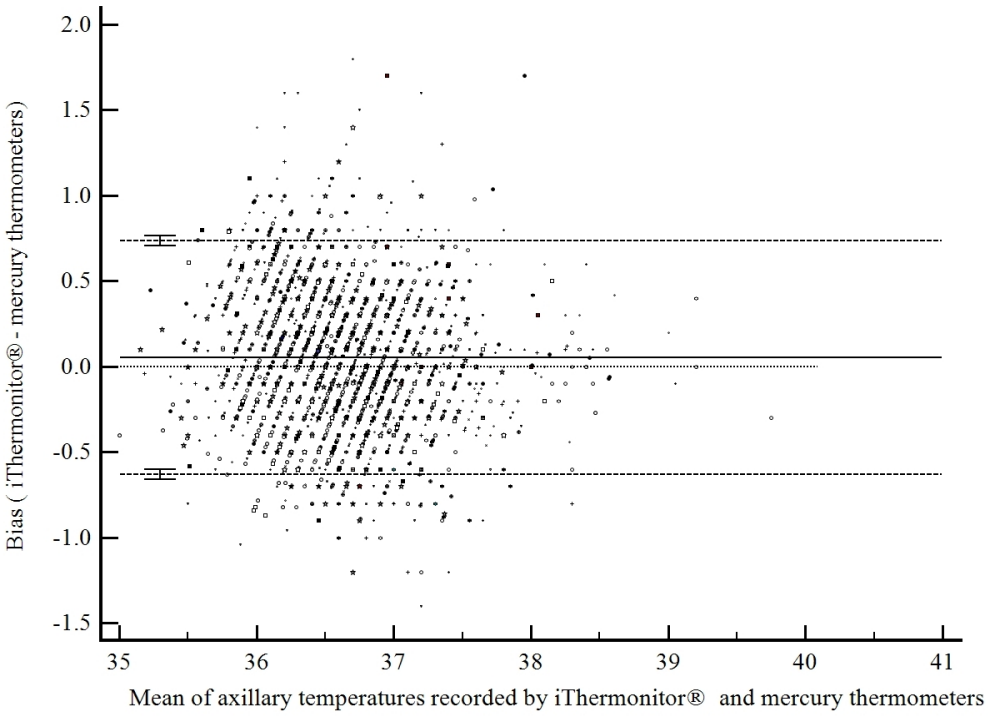
For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



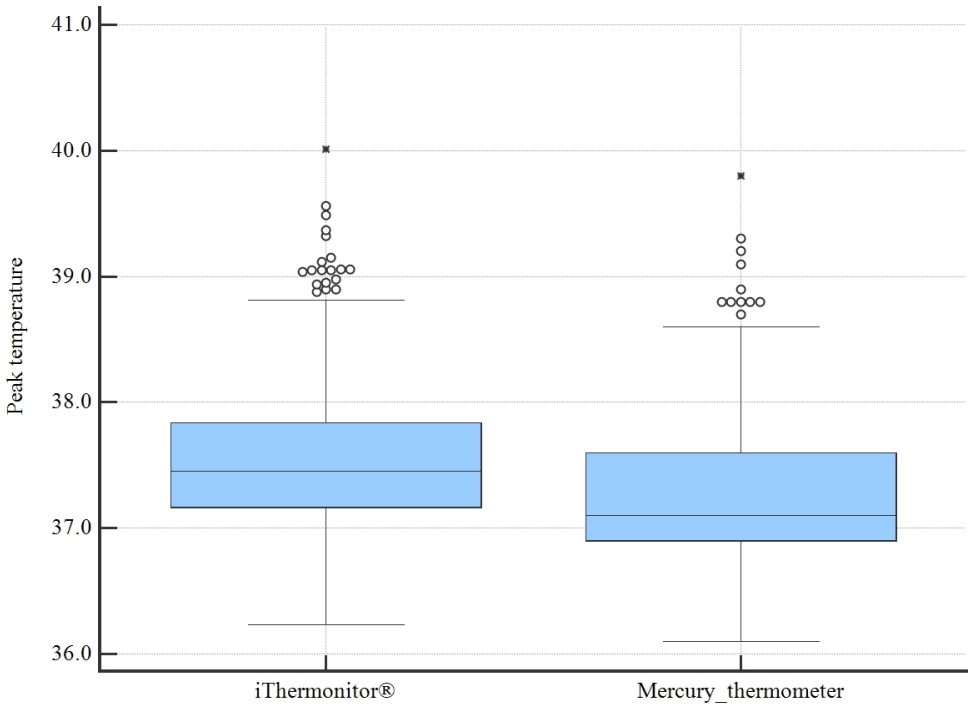
The position of the iThermonitor® sensor in the axilla.

914x1354mm (72 x 72 DPI)



Repeated-measures Bland-Altman plot of iThermonitor® axillary temperatures against mercury thermometer temperatures. ICC=0.753, 95% LoA were from -0.63°C to 0.73°C. A total of 5.11% of the points were outside the 95% LoA.

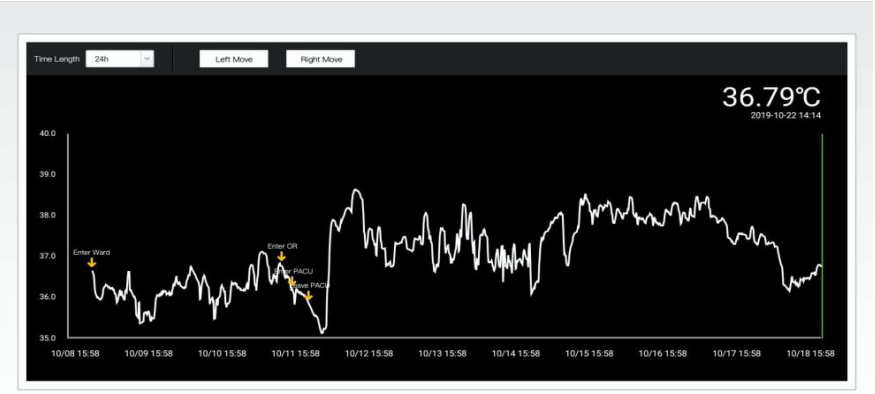
101x76mm (300 x 300 DPI)



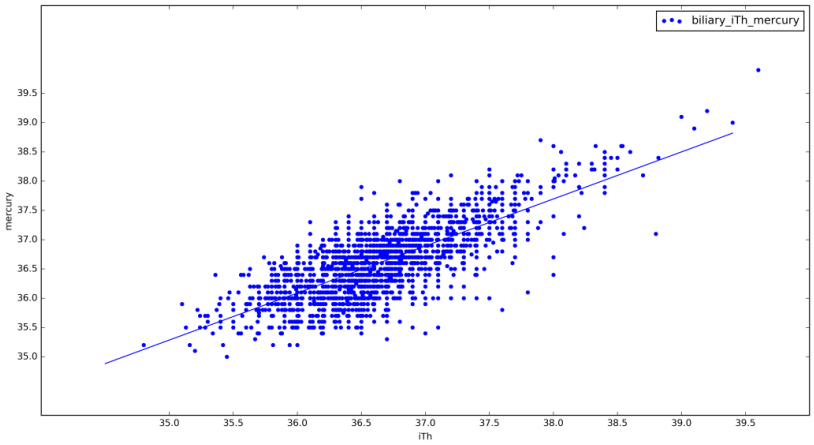
Peak temperatures recorded with iThermonitor® and the mercury thermometers.

101x76mm (300 x 300 DPI)

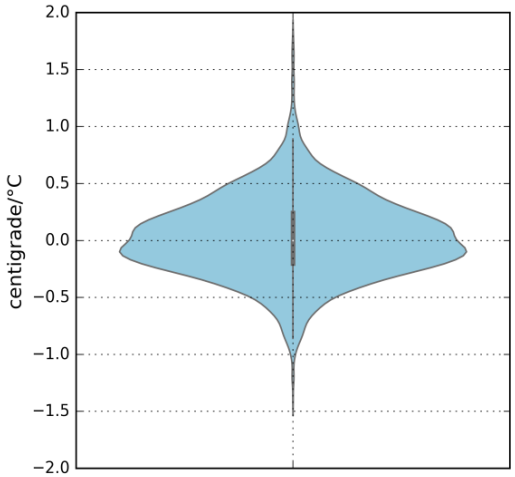
Supplementary Materials



Appendix figure 1 The temperature curve with time stamps across different clinical scenarios.



Appendix figure 2 Scatter plot of iThermonitor® axillary temperatures (x-axis) against mercury thermometertemperatures (y-axis). Pearson`s correlation coefficient r =0.755, P<0.001.



Appendix figure 3 Distribution of the difference of axillary temperatures, calculated by iThermonitor® minus mercury thermometer.

Appendix table 1 The questionnaire to evaluate the feasibility of theiThermonitor® (N=109)

Selections	Score	Patients` Options (%)
I felt very uncomfortable and could not bear to wear it.	1	0 (0)
I felt slight discomfort, but it's tolerable.	2	21 (19.3)
I didn't have noticeable feelings of wearing it.	3	81 (74.3)
I felt comfortable with it.	4	6 (5.5)
I felt very comfortable with it.	5	1 (0.9)

For peer review only

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	Page 2
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	Page 2-3
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	Page 4-5
	4	Study objectives and hypotheses	Page 6
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	Page 6
Participants	**6**	Eligibility criteria	Page 6
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	Page 6
	8	Where and when potentially eligible participants were identified (setting, location and dates)	Page 6
	9	Whether participants formed a consecutive, random or convenience series	Page 6
Test methods	**10a**	Index test, in sufficient detail to allow replication	Page 7-8
	10b	Reference standard, in sufficient detail to allow replication	Page 7-8
	11	Rationale for choosing the reference standard (if alternatives exist)	Page 8
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	Page 9
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	Page 9
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	Page 8
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	Page 8
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	Page 10
	15	How indeterminate index test or reference standard results were handled	Page 10
	16	How missing data on the index test and reference standard were handled	Page 11
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	Page 15
	18	Intended sample size and how it was determined	Page 10
RESULTS			
Participants	**19**	Flow of participants, using a diagram	Page 11
	20	Baseline demographic and clinical characteristics of participants	Page 11-12
	21a	Distribution of severity of disease in those with the target condition	Page 15
	21b	Distribution ofalternative diagnoses in those without the target condition	Page 15
	22	Time interval and any clinical interventions between index test and reference standard	Page 11
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	Page 16
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	Page 16
	25	Any adverse events from performing the index test or the reference standard	Page 17
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	Page 21
	27	Implications for practice, including the intended use and clinical role of the index test	Page 19
OTHER INFORMATION			
	28	Registration number and name of registry	Page 22
	29	Where the full study protocol can be accessed	Page 22
	30	Sources of funding and other support; role of funders	Page 22

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the abilityof one or more medical tests to correctly classify study participants as having a**target condition.**This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**.The reference standardisthe best available method for establishing the presence or absence ofthe target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against thoseof the reference standardcan be used to estimate the**sensitivity** of the index test(the proportion of participants *with* the target conditionwho have a positive index test), and its **specificity** (the proportion *without* the target conditionwho have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test.Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiveroperatingcharacteristic(ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage testis used before an existing test; an add-on test is used after an existing test.

Besides diagnosticaccuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis.The STARD list was not explicitly developed for these other outcomes, statistics, and studytypes, although mostSTARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.

For peer review only