

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Evaluation of a wearable wireless device with artificial intelligence, iThermonitor® WT705, for continuous temperature monitoring for patients in surgical wards: a prospective comparative study
AUTHORS	Liu, Yuwei; Liu, Changqing; Gao, Min; Wang, Yan; Bai, Yangjing; Xu, Ruihua; Gong, Renrong

VERSION 1 – REVIEW

REVIEWER	Goy Jean-Jacques Cardiology Department, Clinique Cecil, Lausanne, Switzerland
REVIEW RETURNED	01-May-2020

GENERAL COMMENTS	<p>This manuscript describes the evaluation of a new wireless thermometer. The results are quite convincing that this new device could be helpful in the daily practice. The english could be improved. The device has been tested in patients after surgery. The authors must add another limitations for example: the device has not been tested in the ICU and / or in patients with compromised hemodynamic which may changed the skin perfusion and temperature. The authors should also discussed the use of such devices in obese patients with poor skin contacts and sweating.</p> <p>Minor comments:</p> <p>A picture of the device with its position would be useful for the reader. Please add one. Which thermometer was use as a gold standard ? Characteristics of the patients should be added in the demography table. Table 3 is difficult to understand. Please improve the understanding. How do the authors explain the small but significant difference of peak temperature?</p>
-------------------------	--

REVIEWER	David Wong University of Manchester, UK
REVIEW RETURNED	15-May-2020

GENERAL COMMENTS	<p>Thank you for the opportunity to read your paper on the reliability of the iThermonitor for measuring body temperature. As a whole, the paper is well written and clear. I do think that the English may be improved, but this can be addressed at the copy-editing stage.</p> <p>In terms of the scientific content of the manuscript, I feel 4 major aspects require extra detail:</p>
-------------------------	---

	<p>1.) There is not ethics statement. In the UK and Europe, one would be expected to have research approval from an institutional ethics board prior to doing research on human participants. I would like confirmation ethics approval was not required (if that is indeed the case)</p> <p>2.) Your data set has repeated measurements from the same patients. The bland-altman analysis should be adjusted to account for this. One way to do this is described in our prior work (https://bmjopen.bmj.com/content/9/8/e031150.abstract). If you have already done this, then it should be made explicit in your Methods section. I note that I am not a medical statistician, but I have recommended (to the editors) that this is reviewed by a statistician to confirm my recommendation.</p> <p>3.) Additional detail on the patient setting would be very useful in the Methods (or limitations if the information is unknown). In particular, information on the mobility of the patients (which may affect temperature via exercise, for instance), and clothing (which may affect thermometer by providing insulation, either to the axilla thermometer, or the 'variable ambient temperature' (I assume that there is more than one thermometer on the device).</p> <p>4.) There was no a priori agreement on acceptable limits of agreement. In many ways, the LoA is more important than the bias (which can be adjusted for if it is systematic). Given that there was no a priori LoA, the discussion should be more careful in its recommendations - a difference of over 0.5C in 1/20 measurements is very concerning, clinically, as 0.5C represents quite a large percentage of the normal range of temperature.</p> <p>Minor issues:</p> <ul style="list-style-type: none"> - Regarding point 3, it would be helpful to have a schematic/picture of the thermometer and how it attaches to a participant. - P8, L20 - not clear what the 'calculation' is doing - an average over 1 minute? - P8, L22 - should state explicitly that the WT705 is an update to the WT701(?) - P8 L55 - 58 - should state the manufacturer/model of the mercury thermometer and high-precision thermometer - P9 L40 - should state the *exact* question that patients answered, as the question can introduce bias. I appreciate that there may be language barrier issues here. I suggest providing the original question in an appendix, and a translation in the text. - P10 L12 - this needs to be carefully reworded. I think that you mean to say that you accepted the first measurement if it was consistent with a second measurement. You note that this procedure is a potential source of bias in the limitations. However, you should also add that this process may lead to bias, as you are not double checking every temperature reading. This means you may miss cases in which the ithermonitor is *falsely* giving the same reading as the mercury thermometer. - P16, line 18 - the sentence starting 'Sum to 21...' is unclear to me - I could not understand it, and I think it should be rephrased.
--	---

REVIEWER	Sara Khalid University of Oxford, United Kingdom
REVIEW RETURNED	27-Jul-2020

GENERAL COMMENTS	<p>A well-presented manuscript.</p> <p>Points to be addressed by authors:</p> <ol style="list-style-type: none"> 1) How long was each patient followed up for? How long did they wear the device. 2) How were the 3621 measurement pairs arrived at? 3) Baseline characteristics for both test and control groups should be presented. This is important to checking for confounding. 4) sample size calculation: 0.5 degrees seems to be clinically significant difference. This should either be addressed, or explained in study limitations. 5) Table 2 row: Within this table, this temperature range has largest number of samples (3285) and smallest bias (0.006) yet the P value is non-significant. Result looks counter intuitive and needs to be re-checked by authors. 6) Typos and grammar needs to be checked throughout.
-------------------------	--

VERSION 1 – AUTHOR RESPONSE

Responses to the comments of reviewer 1

Reviewer Name: Goy Jean-Jacques

Institution and Country: Cardiology Department, Clinique Cecil, Lausanne, Switzerland

1. This manuscript describes the evaluation of a new wireless thermometer. The results are quite convincing that this new device could be helpful in the daily practice. The english could be improved. The device has been tested in patients after surgery. The authors must add another limitations for example: the device has not been tested in the ICU and / or in patients with compromised hemodynamic which may changed the skin perfusion and temperature.

Responses: We are grateful for the comment. We have re-edited the manuscript intensively to improve the language. We also followed the suggestion to add a bullet point under the “Strengths and limitations of this study” section (after the Abstract) and the second limitation under the "Limitations" section of the manuscript (Page 22, Line 14).

2. The authors should also discussed the use of such devices in obese patients with poor skin contacts and sweating.

Responses: This is a good comment. We followed the suggestion and added a multiple linear regression to explore possible factors associated with the accuracy of temperature measurements by the iThermonitor®. The results show that BMI, but not sweating in the axilla, were associated with the differences between the iThermonitor® temperatures and mercury thermometer temperatures (Table 2, Page 15). We further explored the effects of BMI on the difference and found that the difference was significant only in skinny and normal BMI patients, but not in overweight or obesity patients (Table 3, Page 16). We also discussed this finding under the DISCUSSION section of the manuscript (Page 19, Line 17).

3. Minor comments: A picture of the device with its position would be useful for the reader. Please add one.

Responses: The picture as follows has been added (Figure 1, Page 9).



4. Which thermometer was use as a gold standard?

Responses: Readings of axillary mercury thermometers were set as the reference in the study. We evaluated the iThermonitor® using mercury thermometers as references, but did not use a third thermometer as a gold standard. Corresponding changes have been made accordingly in the manuscript (Page 22, Line 11).

5. Characteristics of the patients should be added in the demography table.

Responses: Thank you for your suggestion. We added "mobility", "consciousness", and "sweating in axilla" as the variable of the characteristics of patients in Table 1, page 13.

6. Table 3 is difficult to understand. Please improve the understanding.

Responses: Thanks for the comment. We have re-organized the tables in the manuscript (Page 17, Line 2).

7. How do the authors explain the small but significant difference of peak temperature?

Responses: This is a good question. We did find that continuous monitoring captured higher peak temperatures than intermittent measuring, as previously reported by others (Dakappa et al. Page 21, Line 7). Our explanation for the difference is that continuous monitoring can capture the real peak, while intermittent measuring cannot always capture the real peak, but only captures apparent peaks.

Responses to the comments of reviewer 2

Reviewer Name: David Wong

Institution and Country: University of Manchester, UK

1. There is not ethics statement. In the UK and Europe, one would be expected to have research approval from an institutional ethics board prior to doing research on human participants. I would like confirmation ethics approval was not required (if that is indeed the case)

Responses: We are very sorry that we did not state this important issues in our previous manuscript, which we have added in the current revised version (page 23, line 16).

2. Your data set has repeated measurements from the same patients. The bland-altman analysis should be adjusted to account for this. One way to do this is described in our prior work (<https://bmjopen.bmj.com/content/9/8/e031150.abstract>). If you have already done this, then it should be made explicit in your Methods section. I note that I am not a medical statistician, but I have recommended (to the editors) that this is reviewed by a statistician to confirm my recommendation.

Responses: This is a good comment. We followed the suggestion to add Figure 2 (Page 14).

Corresponding changes have been made in the text including in the Methods section (Page 11, Line 18).

3. Additional detail on the patient setting would be very useful in the Methods (or limitations if the information is unknown). In particular, information on the mobility of the patients (which may affect temperature via exercise, for instance), and clothing (which may affect thermometer by providing insulation, either to the axilla thermometer, or the 'variable ambient temperature' (I assume that there is more than one thermometer on the device)).

Responses: We are grateful for this suggestion. We added "mobility", "consciousness", and "sweating in the axilla" as variables in the table of characteristics of patients (Table 1, page 13). We further explored possible factors associated with the accuracy of the iThermonitor®, including reading of mercury thermometers, gender, age, BMI, length of wearing, mobility, consciousness and sweating in the axilla, through multiple linear regression (Table 2, page 15).

4. There was no a priori agreement on acceptable limits of agreement. In many ways, the LoA is more important than the bias (which can be adjusted for if it is systematic). Given that there was no a priori LoA, the discussion should be more careful in its s - a difference of over 0.5C in 1/20 measurements is very concerning, clinically, as 0.5C represents quite a large percentage of the normal range of temperature.

Responses: This is a good comment, which we agree. Therefore, we were very careful to revise the manuscript, especially the recommendation of the device. The principal idea is as followed: The iThermonitor® is promising for continuous remote temperature monitoring in surgical patients. It can improve fever detection by dynamically reflecting the individual trends of body temperature throughout the whole perioperative period. 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. Corresponding changes have been made throughout the manuscript (Page 23, line 1).

5. Minor issues:

- Regarding point 3, it would be helpful to have a schematic/picture of the thermometer and how it attaches to a participant.

Responses: The picture as follows has been added (Figure 1, Page 9).



- P8, L20 - not clear what the 'calculation' is doing - an average over 1 minute?

Responses: We are very sorry that we did not express clearly. The iThermonitor® sensor would record the axillary temperature once every 4 seconds, then output the average temperature per minute. Corresponding revisions have been made in the manuscript (Page 8, line 16).

- P8, L22 - should state explicitly that the WT705 is an update to the WT701?

Responses: Detail information has been provided under the "Instruments" section of the manuscript (Page 8, line 4).

- P8 L55 - 58 - should state the manufacturer/model of the mercury thermometer and high-precision thermometer

Responses: Thank you for this suggestion. We have added the manufacturer/model of the mercury thermometer and high-precision thermometer under the "Instruments" section. (Page 8, line 20 and Page 9, line 1).

- P9 L40 - should state the *exact* question that patients answered, as the question can introduce bias. I appreciate that there may be language barrier issues here. I suggest providing the original question in an appendix, and a translation in the text.

Responses: This is a good suggestion. We have translated our questions from Chinese into English and provided them as an online supplementary appendix Table 1 as follows.

Selections	Score	Your Option
I felt very uncomfortable and could not bear to wear it.	1	
I felt slight discomfort, but it's tolerable.	2	
I didn't have noticeable feelings of wearing it.	3	
I felt comfortable with it.	4	
I felt very comfortable with it.	5	

- P10 L12 - this needs to be carefully reworded. I think that you mean to say that you accepted the first measurement if it was consistent with a second measurement. You note that this procedure is a potential source of bias in the limitations. However, you should also add that this process may lead to bias, as you are not double checking every temperature reading. This means you may miss cases in which the iThermometer is *falsely* giving the same reading as the mercury thermometer.

Responses: We agree with the comment. Therefore, we added the following discussion under the sub-headline of the "Discussion" section: 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 (Page 19, line 11).

- P16, line 18 - the sentence starting 'Sum to 21...' is unclear to me - I could not understand it, and I think it should be rephrased.

Responses: We are very sorry for our confusing sentence, which we have rephrased in the revised manuscript (Page 18, line 10).

Responses to the comments of reviewer 3

Reviewer Name: Sara Khalid

Institution and Country: University of Oxford, United Kingdom

1. How long was each patient followed up for? How long did they wear the device.

Responses: Thank you for your comment. Each patient wore the iThermometer® temperature sensor from admission to discharge, with an average of 3.37 ± 2.95 (1-22) days. The follow-up lasted from admission to discharge. This information is presented in "Characteristics of the patients" section (Page 12, Line 18).

2. How were the 3621 measurement pairs arrived at?

Responses: We are very sorry for not describing the details, which we have added under the Data collection section of the revised manuscript as follows: " 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 were selected and paired with the temperatures tested by the iThermometer®, 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." (Page 9, Line 15).

3. Baseline characteristics for both test and control groups should be presented. This is important to checking for confounding.

Responses: Thank you for this comment. Actually, the present study was to compare the temperatures measured by the iThermonitor® and mercury thermometer of the same patients, but not the temperatures of test and control groups. Every variable including temperature was paired by one another, even at the very beginning. Baseline characteristics are shown in Table 1 (page 13).

4. Sample size calculation: 0.5 degrees seems to be clinically significant difference. This should either be addressed, or explained in study limitations.

Responses: We added "(usually recognized as clinically significant)" right after 0.5°C, on page 12, line 4.

5. Table 2 row: Within this table, this temperature range has largest number of samples (3285) and smallest bias (0.006) yet the P value is non-significant. Result looks counter intuitive and needs to be re-checked by authors.

Responses: Thank you for the comment. We re-checked our data and analyses. The results were confirmed.

6. Typos and grammar needs to be checked throughout.

Responses: Thank you for this comment. We checked the manuscript intensively to correct the typos and grammar mistakes.

7. This should be discussed in limitations, along with a justification of the respective cut-offs used.

Responses: Thank you for your suggestion. We defined the axillary temperatures of 38°C as a cut-off value of fever, according to the Brighton Collaboration Fever Working Group's definition (page 11, line 3). Moreover, we added some discussions at the end of the Limitations section as follows: "Besides, 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." (page 22, line 18).

VERSION 2 – REVIEW

REVIEWER	Jean Jacques Goy Clinique Cecil Lausanne I read the abstract and in the present form the manuscript could be accepted for publication. With kind regards. Dr JJ Goy
REVIEW RETURNED	08-Sep-2020

GENERAL COMMENTS	Following the modifications done by the authors I think that in the present form I recommend to accept the manuscript for publication.
-------------------------	--

REVIEWER	David Wong
-----------------	------------

	University of Manchester
REVIEW RETURNED	20-Sep-2020

GENERAL COMMENTS	<p>The authors have mostly addressed the points that I raised previously, and the discussion and conclusion is much more considered in this revision. I have limited this review to the points I raised in the first review.</p> <p>For the repeated measures bland-altman analysis, the authors should provide a reference (as there are a few different ways to deal with repeated measures), and also a reference to any programming library used, if any.</p> <p>The authors have included information on the patient setting and included more detail that I had requested by completing a multiple linear regression. The proposed analysis needs to be described in a little more detail in the Method section. In particular, you should explicitly state that the output variable is the difference between the mercury and ithermonitor readings. For Table 2, you should also state the meaning of each column – the headings that you have provided are slightly different to the standard APA-style.</p> <p>In the discussion, you interpret this finding by suggesting that there may be errors in the mercury thermometer recordings, which may indeed be true. However, the alternative interpretation, that the ithermonitor is incorrect, should be also be stated clearly.</p> <p>There are some remaining issues with spelling and grammar, but this should be dealt with in copy-editing and are out of scope of this review.</p>
-------------------------	--

REVIEWER	Sara Khalid University of Oxford, United Kingdom
REVIEW RETURNED	15-Sep-2020

GENERAL COMMENTS	Author responses to the reviewer's comments should be incorporated into the manuscript for completeness of information.
-------------------------	---

VERSION 2 – AUTHOR RESPONSE

Responses to the comments of reviewer 2

1. For the repeated measures bland-altman analysis, the authors should provide a reference (as there are a few different ways to deal with repeated measures), and also a reference to any programming library used, if any.

Responses: Thank you for this suggestion. We performed the Bland-Altman plot with multiple measurements per subject with MedCalc (Version 19.1.3) software. According to the official instructions of the software (Available from: <https://www.medcalc.org/manual/blandaltmanmultiple.php>),

the calculations running inside the software are programmed as described by Bland JM and Zou GY, which we have added into the manuscript as references (Page 10, line 13).

2. The authors have included information on the patient setting and included more detail that I had requested by completing a multiple linear regression. The proposed analysis needs to be described in a little more detail in the Method section. In particular, you should explicitly state that the output variable is the difference between the mercury and iThermonitor readings. For Table 2, you should also state the meaning of each column – the headings that you have provided are slightly different to the standard APA-style.

Responses: We are very sorry that we did not express clearly. We followed the suggestion to add detail descriptions in the “statistics” section (Page 10, Line 15). Corresponding changes have been made in Table 2 (Page 13), including explanation for the coefficients and modification of the format of the table.

3. In the discussion, you interpret this finding by suggesting that there may be errors in the mercury thermometer recordings, which may indeed be true. However, the alternative interpretation, that the iThermonitor is incorrect, should be also be stated clearly.

Responses: We are grateful for this suggestion. Therefore, we added the following discussion under the sub-headline of the "Discussion" section: “ 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.” (page 18, line 9).

4. There are some remaining issues with spelling and grammar, but this should be dealt with in copy-editing and are out of scope of this review.

Responses: We are sorry for the mistakes in spelling and grammar. We have rephrased the manuscript to improve the language.

Responses to the comments of reviewer 3

1. Author responses to the reviewer's comments should be incorporated into the manuscript for completeness of information.

Responses: Thank you for your comment. We have re-checked our manuscript to confirm that each of our point-by-point responses to your comments has been incorporated into the manuscript.

VERSION 3 – REVIEW

REVIEWER	David Wong University of Manchester, UK
REVIEW RETURNED	13-Oct-2020

GENERAL COMMENTS	Dear authors, thank you for revising your manuscript. You have addressed all of my previous comments. All the best!
-------------------------	---

REVIEWER	Sara Khalid University of Oxford, UK
REVIEW RETURNED	14-Oct-2020

GENERAL COMMENTS	No further comments
-------------------------	---------------------