Author’s response to reviews

Title: Nocturnal finger skin temperature in menstrual cycle tracking: Ambulatory pilot study using a wearable Oura ring

Authors:

Anna Maijala (anna.maijala@oulu.fi)
Hannu Kinnunen (hannu.kinnunen@ouraring.com)
Heli Koskimäki (heli.koskimaki@ouraring.com)
Timo Jämsä (timo.jamsa@oulu.fi)
Maarit Kangas (maarit.kangas@oulu.fi)

Version: 1 Date: 28 Jun 2019

Author’s response to reviews:

Nocturnal finger skin temperature in menstrual cycle tracking: Ambulatory pilot study using a wearable Oura ring

Anna Maijala, Hannu Kinnunen, Heli Koskimäki, Timo Jämsä, Maarit Kangas

Dear Editor-in-Chief of BMC Women’s Health,

Thank you for the feedback and the opportunity for revision. Please find the revised manuscript entitled ‘Nocturnal finger skin temperature in menstrual cycle tracking: Ambulatory pilot study using a wearable Oura ring’, and revised figures 4, 5 and 6. Please note that we have slightly revised the title of the manuscript.

Major revision has been performed according to the recommendations, see point-by-point response for the reviews below. Track changes functionality has been used to show the changes. For practical reasons, changes done in references have been accepted.

Oulu, June 28, 2019
Yours sincerely,
Anna Maijala
Hannu Kinnunen
Heli Koskimäki
Timo Jämsä
Maarit Kangas

Response to reviewers

Thank you for the reviewers’ valuable comments on this manuscript. Track changes functionality has been used to show the changes. In order to lines described in replies to correspond to correct lines in the manuscript and the deleted text to show next to the text, please use following settings in Word in Track Changes: All Markup, and Show Markup, Balloons, Show Revisions in Balloons. For practical reasons, changes done in references have been accepted.

Reviewer reports:

Reviewer 2 (Reviewer 2): PEER REVIEWER ASSESSMENTS:

GENERAL COMMENTS: The information provided is novel and could be interesting for the researchers and stimulate further investigation. Nevertheless, there could be severe methodological biases, which prevent to draw a firm conclusion about the main questions of the study.

In particular, authors should highlight how they managed the sample size of the study, since the number of enrolled women is extremely low. In addition, it is not clear if the authors excluded possible confounding factors (such as extremely low/high BMI, age and other variables) that may have influenced the main outcomes of the study.

Authors did not declare whether the study followed the "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement" (PMID: 18064739), available through the Enhancing the QUAlity and Transparency Of health Research (EQUATOR) network.
Finally, many references are really too old: there are even two references [28] and [29] of articles published in 1990.

REQUESTED REVISIONS:

Comment: Authors appropriately stated that skin temperature patterns could be influenced by several elements. Did they normalize this parameter for all the other variables that are known to influence the main outcomes of the study?

Reply: Parameters were not normalized for confounder variables. We have added some more detailed information on the characteristics of the participants (Lines 170-173, 342-345) in order to give somewhat deeper insight into the matter. Additionally, we have added discussion on this topic (Lines 480-482, 492-495).

Comment: Did authors perform any sample size before to start the enrollment? The number of enrolled women (n=23) is extremely low to draw firm conclusions. In addition, authors had 3 drop-outs and 6 women with incomplete data, so the number of final analyzed women (14) is even lower!

Reply: We added clarification considering the use of data for the drop-outs (Lines 153-158). Additionally, we rephrased the part explaining missing data (Lines 153-158), and added number of participants also to Results for clarity (Line 322). Altogether, we used data from all sub-study participants (n=23) in the analyses though for some participants the study length was under the originally planned 120 days.

This paper describes a real-life pilot test of the methodology. The sample size was designed in accordance with the available resources. We have modified our text to emphasize that the sample size was low and declared that further larger study is needed to validate the results (Lines 2, 27, 128, 482-484, 513-514). In addition, we have modified the text (Lines 512-513) to emphasize that the method is applicable and shows potential in menstrual cycle phase monitoring. The results of this study can be used for sample size calculations in future studies.

Comment: Did authors exclude patients affected by significant comorbidities? If not, this may represent a significant bias.
Reply: We have provided more information on the underlying diseases (5 diseases, 4 participants) and continuous medications (6 medications, 5 participants) reported by the participants. We have specified whether these could potentially affect temperature or menstrual cycle, and if potential effect found, it has been reported (Lines 170-173). We have also added text about potential effects on the results (Lines 480-482, 494-495).

Comment: Did enrolled women get any reward/payment to enter/continue the study? It would be mandatory to declare about this point.

Reply: We have added clarification concerning this matter in the text (Lines 145-148): “The enrolled women did not get any payment to enter the study but those who completed the measurements were offered the possibility to continue the use of the Oura ring after the study. As a commercial product the Oura ring gave the participants health-related information such as summary of their sleep and physical activity.”

Comment: Did the study conform the "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement" (PMID: 18064739), available through the Enhancing the QUAlity and Transparency Of health Research (EQUATOR) network? It would be mandatory to design this type of study following the abovementioned international standards.

Reply: Thank you for this valuable comment. After your comment, we went through the list of recommendations in Strobe statement, and checked our manuscript regarding to that when applicable. For clarity we have added the years of data collection and country where the study took place (Lines 177-178).

Comment: The age of enrolled women seems extremely variable (21-49): these extreme ages may have severely influenced reported outcomes.

Reply: Thank you for this comment. We have designed the study protocol in order to evaluate the applicability of new method in real-life conditions and thus we did not have any restrictions for age of the participants, as long as they were not in menopause. We had fairly even number of participants from each age group: 20-29 years (n=7), 30-39 years (n=8), 40-49 (n=8).
It is true that menstrual cycle can vary according to age, for example affecting the cycle length, and the ages close to 50 years are close to menopause. However, the two eldest participants in this study declared regular menstruation in the preliminary survey. Two of the youngest participants declared irregular menstruation with slightly prolonged cycles in the preliminary survey. However, typically this age window is not itself anymore an exposing factor for irregular cycles as it can be for example for adolescents.

It is also notable that distal blood circulation decreases and therefore may affect the distal temperature as getting older. However, this does not yet seem to affect the highest distal skin temperature or its timing in middle-aged compared to young people [Batinga H, Martinez-Nicolas A, Zornoza-Moreno M, Sánchez-Solis M, Larqué E, Mondéjar MT, et al. Ontogeny and aging of the distal skin temperature rhythm in humans. Age (Dordr). 2015;37(2):29.].

Comment: The BMI of enrolled women seems extremely variable (20.3-37.2): these extreme BMIs may have severely influenced reported outcomes.

Reply: It is true that obese people have generally higher finger skin temperature, and obesity increases the risk for menstrual disorders. However, as stated in the previous reply, the study was designed to evaluate the applicability of new method in real-life conditions and thus we did not have any restrictions for BMI. For clarity, we have added information about the amount of participants who were classified as obese according to their BMI calculated on the basis of weight and height asked in the preliminary survey (Line 170). Additionally, we have added text considering obesity as potential confounder to temperature and menstrual disorders and stated its possible effect on the results of this study (Lines 480-482, 492-494).

Comment: Most of the references are too old: 2001 [8], 2002 [9], 1996 [10], 1996 [11], 2001 [12], 2005 [13], 1998 [20], 1994 [22], 1997 [28], 1990 [29], 1990 [30]. This is completely not acceptable: I recommend to avoid references published more than 5 years ago.

Reply: Thank you for your important notice. We have taken this into account when modifying the text (Background). Below we have specified the modifications concerning above mentioned old references. We have also added recently published papers about using wearable devices in fertility related phase detection and prediction in Background (references in the new numbering [23], [24], and [25]).

References [8-12]: We declare that these pieces of information stated are widely known, thus the references are seen unnecessary in these cases and left out.

Reference [13]: We declare that this piece of information is not essential for the manuscript, thus it is left out.
Reference [20]: References [19] and [21] already state the use context of oral temperature used in BBT measurements regarding cycle computers, thus reference [20] is left out as unnecessary addition.

Reference [22]: References [23] and [24] already state the issue, thus reference [22] is left out as unnecessary addition.

Reference [28]: Reference [25] states the issue, thus reference [28] is left out as unnecessary addition.

References [29-30]: References [31-32] already act as relevant information sources, thus the two oldest references [29-30] are left out.

Helena Kopp Kallner (Reviewer 3): Review

Comment: This paper is rather well written and describes the use of a ring to measure basal body temperature throughout the menstrual cycle. Exactly what this will be used for is not entirely obvious but it seems the authors are more interested in health applications that contraception and fertility. The English need to be checked but not to a large extent.

Reply: Thank you for the comment. It is true that in this paper we wanted to bring out the versatility of use cases possible when monitoring menstrual cycle phases. However, it is true that the most obvious and really important use case is fertile window tracking whether the aim was to get pregnant or avoid pregnancy. We have expanded the text considering this use case in the beginning of Background (Lines 67-68) and notified it also in Conclusions (Lines 519-520). On the basis of other comments we have also modified the windows used in ovulation detection (Lines 289-293) to also acknowledge this most obvious use case and the results and text throughout.

General comments

Comment: I miss a discussion on improvements. How could the method be improved?

Reply: We have clarified the text considering the future improvement already stated (Line 501), and included ideas for further improvements in the text (Lines 501-502, 517).
Comment: Is it really necessary to open the app and download the temperature? Why is this not done automatically?

Reply: The Oura ring is a commercial product that gathers minute-by-minute temperature data in its short memory, in which the minute-by-minute data will last from day to couple of days. In order to get the accurate data for the study, the ring had to be connected to the application within this time window. In the commercial product itself, some calculation can be done already in the ring. Additionally, in another simultaneous sub-study, the participants were asked to view their data in the application daily, which also required connecting the ring to the application each morning. No changes has been made to the text.

Comment: With longer registration- would it not be better with a "learning algorithm" which improves with longer registration?

Reply: Thank you for this observation. We have included this in the text (Lines 501-502, 517).

Comment: I also miss a discussion on why women participated? Were they offered anything to participate or did they do it "just for fun".

Reply: We have added clarification concerning this matter in the text (Lines 145-148): “The enrolled women did not get any payment to enter the study but those who completed the measurements were offered the possibility to continue the use of the Oura ring after the study. As a commercial product the Oura ring gave the participants health-related information such as summary of their sleep and physical activity.”

Comment: And finally I miss a discussion on the performance of prediction in relation to existing published apps such as the Dot app and Natural Cycles. Even though they are menstruations trackers and contraceptive devices- it is still interesting with a discussion why anyone would choose to use this model with the ring instead of the existing ones Lastly this is quite a mathematical paper and I would let someone familiar with correlation statistics and algorithms review the paper.

Reply: We have added discussion on why would customers choose a wearable over a contraception application (Lines 449-453) and comparison to Natural Cycles (Lines 446-448). However, we decided to mostly concentrate the performance comparison between the wearables since studies concerning contraception applications such as Dot app and Natural Cycles utilize different performance outcomes such as perfect- and typical-use indexes for becoming pregnant when used as contraceptive.
Specific comments

Comment: Line 58 would state the most common premenstrual symptoms here. Such as irritability, anger or depressed mood.

Reply: We have added a list of most common premenstrual symptoms in the text (Lines 72-73).

Comment: Line 61 vaginal flow is awkward. Vaginal secretion?

Reply: Rephrased as suggested (Line 75).

Comment: Line 63 "follow ovulation for". Rewrite- eg The rise in BBT occurs at 1-3 days post ovulation.

Reply: Rephrased as suggested (Lines 79-80).

Comment: Line 74 strange sentence

Reply: Reworded for clarity (Lines 94-96).

Comment: Line 78-83 just to be sure I understand. The distal skin temperature rises as CBT decreases? Reply: Reworded for clarity (Lines 100-103).

Comment: line 231 Please justify using the window lengths used. They seem very wide. Would anyone accept +/- 4 days? Especially considering the survival of sperm of 6 days. If it were to be used for contraception eg or for identifying the fertile window.

Reply: Thank you for the valuable comment. We agree that window length +/- 4 days is wide for ovulation detection, especially when considering the common use case of identifying the fertile window. We have edited the window lengths used. We have deleted the window lengths +/- 4 days and +/- 3 days. We have created two new windows on the basis of fertility point of view: windows -4 to +1 days and -3 to +2 days around the ovulation day as depending on the reference ovum lives for 1-2 days after ovulation and the sperm survives for 6 days inside woman’s body. Relevant corrections have been updated to text (Lines 289-293, 381-388, 438-440) and Figures 5 and 6.
Comment: Line 298 generally good is a bit too subjective. What exactly is meant by good?

Reply: We have rephrased the text and as an example of the good performance added sensitivity for fertile window detection for the best-performing ovulation algorithm (Lines 396-397).

Comment: Line 313-321 would leave out the algorithms that did not perform well.

Reply: We have modified the text by leaving out other algorithms than the best-performing one and concentrated more on discussing the performance of our best-performing algorithm and method itself compared to other existing technological temperature-based methods (Lines 438-448).

Comment: Line 343 don't understand. Is this not what has been done. But retrospectively with the dates already known beforehand. Why would this differ from what has been done?

Reply: Rephrased for clarity (Line 501). In our study, we have used the data from the whole measurement period in preprocessing the data e.g. in finding the menstrual cycle component and in algorithm definition. In the future it would be ideal only to use the data collected until the date of start of menstruation or ovulation to be predicted.

Comment: Line 351 it seems that these areas for use are not the most obvious ones and they are hardly mentioned in the manuscript at all.

Reply: We have added in the list the mention of the most obvious use case (Lines 519-520).

Comment: Line 328. Interesting that you call this limitation- I would call it a strength that it performed so well under these more "typical use" conditions.

Reply: We have added this consideration of the results as another way of interpreting this matter in the text (Lines 483-484): “However, the results demonstrate potential of the method in a real-life use case.”
Victoria H. Jennings, Ph.D. (Reviewer 4):

Comment: This manuscript will require extensive editing for an English-speaking audience.

Reply: Thank you for your comment. The language of the first submitted manuscript has been checked by professional proofreading service. For this version, we have now modified some parts to make the text clearer. For example in order to enhance the readability, we named algorithm tracking the start of menstruation as MENSES.

Comment: I have several comments that should be addressed. Throughout the authors refer to "detecting" [the first day of] menses. From the context, it appears they mean "predicting" rather than "detecting". Please clarify.

Reply: Thank you for this careful notice. We have modified the text throughout when referring to algorithm: verb detect has been replaced with predict.

Comment: This is a very small sample size with missing data (missing temperature measurements) and early exit. This is mentioned in the discussion under shortcomings of the study, but it should also be acknowledged earlier in the study -- perhaps in the section explaining the study design. Why would we think that a data set this small would do anything other than give us an idea about whether the authors are on the right track and should design a larger study?

Reply: The reviewer is correct that the present study presents preliminary findings and potential of the ring-based skin temperature method. We have added clarification considering the use of data for the drop-outs, and rephrased the part explaining missing data for clarity (Lines 153-158). We have also added low number of participants in the limitations and declared that further larger study is needed to validate the results (Lines 482-483, 513-514). In addition, we have modified the text (Lines 512-513) to emphasize that instead of proving the possible use the method is applicable and shows potential in menstrual cycle phase monitoring.

Comment: Please provide more information on One Step Ovulation Testing Midstream. Note also that it measures urinary LH and apparently uses its rise as a proxy for ovulation. This needs to be justified.
Reply: Thank you for this important notice. We have added manufacturer’s information on the threshold and accuracy of the ovulation tests (Lines 206-208). On the basis of this further consideration of the relationship between urinary LH test result and ovulation day, we have decided to change the definition of the verified ovulation day used in this study (Lines 210-211): “The day following the first positive ovulation test result was used as a verified ovulation day (day 0) based on the interpretation guidance of the test instructions and literature [Direito A, Bailly S, Mariani A and Ecochard R. Relationships between the luteinizing hormone surge and other characteristics of the menstrual cycle in normally ovulating women. Fertility and Sterility. 2013;99(1):285.e3.].” Relevant corrections have been updated to all analyses and text (Abstract, Methods, Results, Discussion) and Figures 5 and 6.

Comment: The definition of the Calendar approach is faulty. I am unaware of any similar definition in the literature. See instead: Hatcher et al, Contraceptive Technology, 21st edition: Chapter 12, Fertility Awareness-Based Methods, p. 408. In addition, given the sample size and lack of controls, I recommend that the authors eliminate any comparison with a Calendar approach. It does not strengthen the paper.

Reply: We apologize for misleadingly using name calendar method for our own biological rhythm-based approach. We have modified the definition and the name of the method. As comparison to this biological rhythm-based method does not seem to strengthen the paper, we have considerably cut down its role in the manuscript by removing it from the figures (Figure 4, 5 and 6) and text apart from one notice that underlines the overall strength of our method compared to the biological rhythm-based method (Lines 388-389). Otherwise, we have concentrated more on discussing the performance of our method compared to other existing technological temperature-based methods (Lines 439-454).

Comment: Also note another recent paper: Handel and Wahlstrom, Digital contraceptives based on basal body temperature measurements. Biomedical Signal Processing and Control, Volume 52,, July 2019, pages 141-151. It is relevant to this paper.

Reply: Thank you for this comment. We have added this recent paper to Background (Lines 125-126).