The performance of a fertility tracking device

Niels van de Roemer, Liya Haile & Martin C. Koch

Published online: 08 Feb 2021.

Submit your article to this journal ⬀

Article views: 12354

View related articles ⬀

View Crossmark data ⬀

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

🔓 OPEN ACCESS   ⊙ Check for updates

The performance of a fertility tracking device

Niels van de Roemer[a] 🔸, Liya Haile[b] 🔸 and Martin C. Koch[c]

[a]Department of Medical Affairs, Valley Electronics AG, Zurich, Switzerland; [b]Institute for Reproductive Health, Georgetown University, Washington, DC, USA; [c]Gynaecological Clinic, Ansbach Hospital, Ansbach, Germany

ABSTRACT
Objective: Fertility tracking devices offer women direct-to-user information about their fertility. The objective of this study is to understand how a fertility tracking device algorithm adjusts to changes of the individual menstrual cycle and under different conditions.
Methods: A retrospective analysis was conducted on a cohort of women who were using the device between January 2004 and November 2014. Available temperature and menstruation inputs were processed through the Daysy 1.0.7 firmware to determine fertility outputs. Sensitivity analyses on temperature noise, skipped measurements, and various characteristics were conducted.
Results: A cohort of 5328 women from Germany and Switzerland contributed 107,020 cycles. Mean age of the sample was 30.77 [SD 5.1] years, with a BMI of 22.07 kg/m∘2 [SD 2.4]. The mean cycle length reported was 29.54 [SD 3.0] days. The majority of women were using the device 80-100% of the time during the cycle (53.1%). For this subset of women, the fertility device identified on average 41.4% [SD 6.4] possibly fertile (red) days, 42.4% [SD 8.7] infertile (green) days and 15.9% [SD 7.3] yellow days. The number of infertile (green) days decreases proportionally to the number of measured days, whereas the number of undefined (yellow) days increases.
Conclusion: Overall, these results showed that the fertility tracker algorithm was able to distinguish biphasic cycles and provide personalised fertility statuses for users based on daily basal body temperature readings and menstruation data. We identified a direct linear relationship between the number of measurements and output of the fertility tracker.

Introduction

Fertility Awareness-Based Methods (FABM) are a set of family planning methods based on a woman's periodic fertility [1–3]. FABMs can be used for avoiding or achieving pregnancy, and as a way to monitor gynecological health by observing one or more of the three primary fertility signs (basal body temperature (BBT), cervical mucus, and cervical position) [4]. Research has found that 20% of all women in the United States have tried a FABM at some point in their lives [5]. Recent interest in hormone and side-effect free family planning methods has contributed to the growing demand of FABMs. A population-based survey of women in the United States showed that 41% of women reported that they did not use hormonal contraceptives due to fear of side-effects [5].

Typically, use of a FABM requires an understanding of one's own biology, discipline and the proper educational training [6,7]. Furthermore, reading and interpreting data is open to human error, potentially reducing the method's efficacy [8,9]. In the last two decades, the observation and calculation of fertility with pen and paper has been increasingly replaced by new, sophisticated devices capable of measuring, storing and evaluating the direct or indirect signs of fertility [10,11]. With a rapidly evolving market for digital fertility tracking, there is the need for verification of a mobile application's safety to correctly provide fertility information. Most fertility tracking application (FTA) algorithms are not designed based on evidence-based methods or research nor have they been evaluated in peer-reviewed literature [12,13]. Women around the world are in need of accurate, timely, and easily accessible information about their fertility. The probability of pregnancy changes throughout the menstrual cycle. With digital fertility trackers, women can receive information about their fertility and can make informed decisions based on their reproductive intention [14,15].

The Fertility Tracker (Daysy) is a fertility awareness-based device based on LadyComp, BabyComp and Pearly. Users record daily BBT measurements once a day in the morning immediately after waking up and also confirm their menstruation. The device displays the user's fertility status through LED lights, in which green indicates 'infertile', red indicates 'possibly fertile' and yellow indicates 'undefined' (Figure 1). The colours are indicative of a woman's fecundability determined by the amount of data provided by the users. Consistently measuring BBT lowers the amount of 'red' of 'possibly fertile' days as the device caters to the user's individual menstrual cycle. In Europe, the Fertility Tracker (Daysy) is considered to be an invasive active medical device. DaysyDay, a free mobile app is an optional supplement to the Fertility Tracker. Available on iOS and Android platforms, the DaysyDay app offers a graphical display of the user's fertility status each day of her cycle, temperature curves, and numerous statistics (Figure 1).

Figure 1. Visual of Daysy device and DaysyDay App.

There are several apps and devices on the market based on some variation of the symptothermal method, which use two different symptoms, observation and precise interpretation of cervical mucus and/or daily measurement and evaluation of basal temperature to determine the beginning and end of the fertile window. By means of different rules, the fertile window can be relatively predicted. Few studies have looked at the applicability of these apps and devices to adapt to changes in the menstrual cycle. There are also no studies that investigates how digitally programmed criteria of a BBT shift detection algorithm correlate with individual cycle characteristics on a daily basis. In this study, we aim to assess how the fertility tracking device algorithm adjusts to changes of the individual menstrual cycle.

The analysis aims to better understand and identify any errors or discrepancies in Fertility Tracker's outputs and how physiological factors directly influence the outcome of the algorithm (i.e., age, BMI, cycle length, measurement skipping, high vs. low average temperature, temperature steps).

Methods

Fertility tracker device

The Fertility Tracker was developed in what is referred to as the calculothermal method, which assumes that all non-fertile days are 'possibly' fertile due to the variability of the individual cycle. By measuring BBT, the time-point of ovulation and thus the infertile days after ovulation can be precisely defined [16]. By statistically estimating when ovulation occurred individually in previous cycles, it is also possible to determine the infertile days after menstruation, i.e., before the beginning of the fertile window [17]. Since the calculothermal algorithm takes into account each new cycle individually in the calculation of infertile days, it represents a flexible compromise between variable and constant cycles among the FABMs.

The Fertility Tracker takes into account daily measured BBT as well as the start and end of menstruation each cycle to provide real-time fertility information, but it does not predict the end of the fertile window in advance. The Fertility Tracker only shows the fertility status of the current day, as it is not possible to show status for future days in the cycle. This approach allows the Fertility Tracker to adapt to the natural variation of a woman's menstrual cycle. By combining the acquisition and learning of new data (the daily orally measured BBT, start and end of menstruation, accumulated past cycle data) and statistical methods (e.g., the temperature rises after ovulation), the Fertility Tracker is able to support women with their family planning.

As users continue to use the Fertility Tracker, the algorithm uses statistical methods and previous cycle history to better determine a user's fertility status after menstruation but prior to ovulation. A sustained increase in the rolling average BBT of at least 0.2-0.3 °C for a minimum of two to three calendar days after the expected time of ovulation is necessary to determine a temperature shift by the algorithm. The algorithm then compares the predicted date to the calculated ovulation date at the end of the cycle to update the model accordingly, as shown in Figure 2.
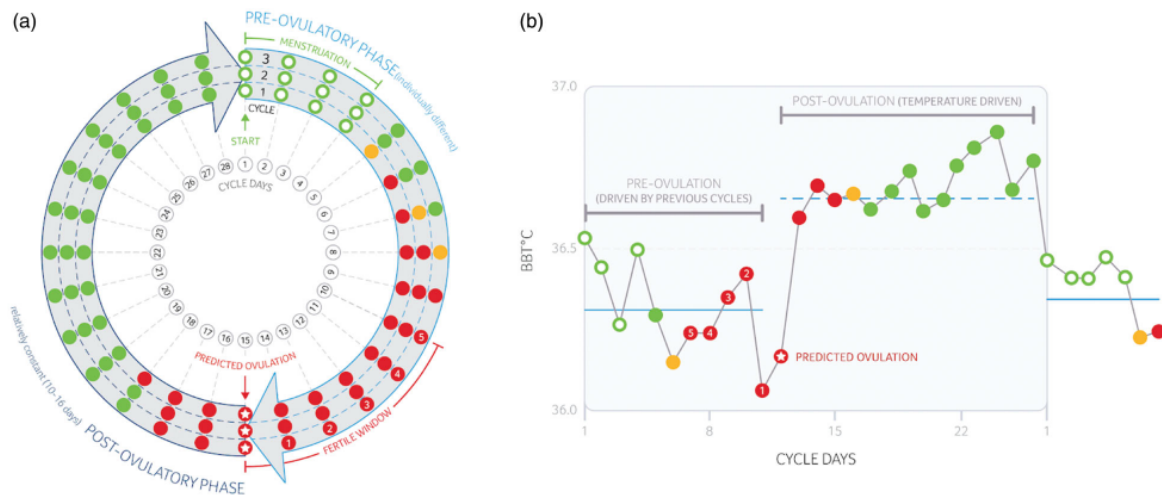
Figure 2. Modifed BBT shift detection algorithm.

Study design

Retrospective data analysis was conducted to assess how the fertility tracking device algorithm adjusted to changes in the menstrual cycle. The study protocol was reviewed and authorised by the regional ethics committee (FAU/Erlangen/13_20 Bc). Anonymized data collected from existing LadyComp, BabyComp and Pearly users between 1 January 2004 and 1 November 2014 was used for this study analysis. Women with cycles shorter than 19 days and longer than 50 days were excluded. Furthermore, datasets had to contain at least one complete cycle to be included in the analysis. Cycles in which pregnancy was assumed by a significantly high temperature (post-ovulation phase) was present for more than 25 days, were excluded from the analysis. Data sets included temperature and menstruation cycle data (e.g., menstruation date, menstrual cycle length) as well as sociodemographic variables (e.g., age, height, weight, location). The 5th and 95th percentiles were excluded from the cycle characteristics analysis.

Analysis

Available temperature and menstruation inputs were processed through Daysy firmware to receive fertility outputs (red, green, or yellow days). Fertility estimates were calculated and the key results stored in an Excel table. Raw data stored on the devices can be uploaded into an analysis program (VE Analyser) and used to generate a BBT Chart. Data from the pdf files were used for the comparison to the data in the excel sheet. Further information on the information provided in the pdf files can be found in Appendix.

We assessed the impact of missing temperature information. Fertility estimates were calculated for datasets that provided data between 0-20% of their cycle, 20-40%, 40-60%, and 80-100%. Temperature noise was also assessed in an independent model, through theoretical simulation testing. Fertility Tracker's algorithm was fed with defined standard deviation temperature values of $\sigma=0.05\,°C$, $\sigma=0.1\,°C$, $\sigma=0.2\,°C$, $\sigma=0.3\,°C$ for 3 cycles each. The data set for this analysis contained cycles that were 28-days long and included 100% measurement of BBT. Third,

fertility estimates were collected from cycles with various temperature noise and BBT measurement skipping using the real-world cohort. Cycles were divided into two groups that had measured more than 60% or more than 90%. Cycles that had measured less than 60% were removed from this analysis. We identified menstrual cycle days that were indicated as green (infertile) by the device, but were within the six day fertile window or two days after the significant temperature rise (defined as the estimated period of ovulation). Then, we compared the distribution of false positive green days from the device to previously identified daily fecundability probabilities, to better understand the gaps in measurement.

Sensitivity analysis

Lastly, an analysis was conducted with 16 unique datasets (long cycles, short cycles, etc.) with 20 users each. The randomised selection of the data samples is described in detail in the Appendix

Results

Sample characteristics

A cohort of 5328 women from Germany and Switzerland contributed 107,020 cycles. A total of 310 women were excluded from the analysis due to a lack of at least one complete cycle. On average, women contributed 21.5 complete cycles and measured their BBT 69.77% [SD 23.5] during their cycle. For this analysis, we excluded 17,040 cycles due to elevated post ovulation temperature longer than 25 days as it is an indication of pregnancy. The average number of BBT recordings per cycle was 25.26 days [SD 6.3]. Almost half of the sample, 47.2% (n = 2516) did not report their age, height and weight, therefore BMI could not be calculated. However, 52.8% (n = 2812) of the sample was within normal/healthy BMI limits, with a mean age of 30.77 [SD 5.1] and a mean BMI of 22.07 kg/mo2 [SD 2.4] (Figure 3). The mean cycle length of the remaining cycles (n = 93,569) was 29.54 [SD 3.0] days. The majority of cycles (67.8%) were 25-30 days long, with a mean cycle length of 27.5 days. We found 4.4% of cycles (n = 3589) to be
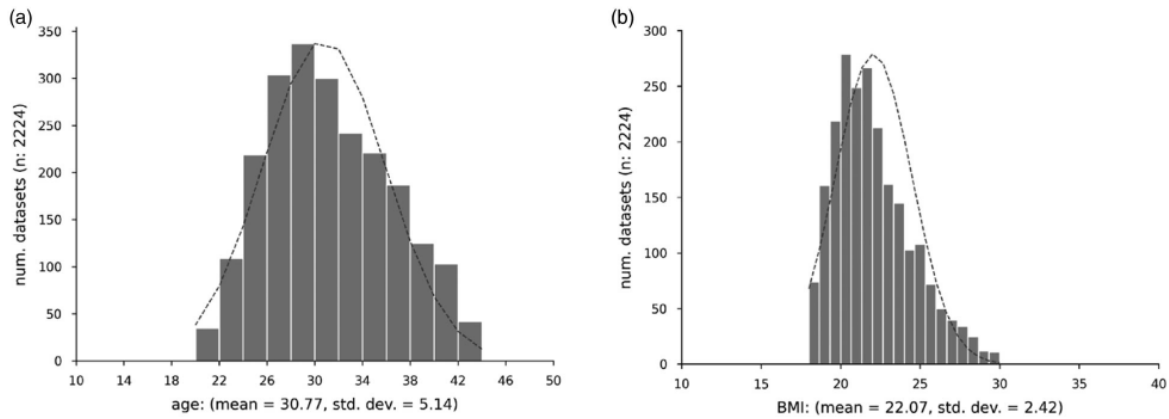
Figure 3. Mean Age and BMI of study sample.

identified as monophasic cycles (anovulatory). Of this sub-set, 49.2% of these cycles were identified as short cycles (13-20 days). The average length of the pre-ovulation phase was 16.8 [SD 2.9] days while the average post-ovulation phase length was 12.8 [SD 1.6] days long. We found that only 12.5% of cycles were 28 days long.

Fertility outputs

The majority of women were using the device 80-100% of the time during the cycle (53.1%). For this subset of women, the fertility device identified on average 41.4% [SD 6.4] possibly fertile (red) days, 42.4% [SD 8.7] infertile (green) days and 15.9% [SD 7.3] yellow days (Table 1). The number of infertile (green) days decreases proportionally to the number of measured days, whereas the number of yellow (undefined) days increases. For users measuring their BBT for 60-80% of their cycle, the device on average identifies 35.4 [SD 8.3] infertile (green), 39.9 [6.6] possibly fertile (red), and 15.9 undefined (yellow) days (Table 1).

As shown in Table 2, temperature noise has a direct influence on the output of the fertility algorithm. When sigma is very low (0.05 °C), the algorithm provides more green (fertile) days (56%) and the least yellow (undeined) (4%) days. At a very high sigma (0.30 °C), Fertility Tracker displayed relatively less green (43%) and more yellow days (17%). The percent of possible fertile (red) days is roughly unchanged in either simulation.

The real-world data set was further analysed against temperature noise and skipped BBT measurement (Table 3). The average standard deviation value of the measured temperature in total was 0.17 °C [0.05]. If users have measured more than 60% (Gr.1), a similar linear distribution of infertile (green) and undefined (yellow) days was observed as in the simulation. In contrast to the simulation, there was also a linear change in the potentially fertile (red) days for the real-world data set. The highest ratio of infertile 45.7% [SD 13.9] and fertile 40.1%[10.6] days can be found at a temperature standard deviation of 0.1 °C on average and a measuring rate of >90% (Table 3).

Sensitivity analysis

There were 300 women in this analysis who contributed a total of 9934 completed cycles (Table 4). The average age

Table 1. Distribution of fertile (green), possibly fertile (red) and undefined (yellow) days depending on the amount of measurement.

Measured (%)	Cycles	Infertile (green)%	Fertile (red)%	undefined (yellow)%
0-20	1842 (2.04%)	8.1 [4.7]	18.4 [27.1]	73.5 [17.6]
20-40	5039 (5.6%)	16.8 [5.9]	25.7 [11.2]	57.4 [15.1]
40-60	11.915 (13.2%)	25.9 [6.7]	33.6 [8.3]	40.3 [11.5]
60-80	23.379 (26.0%)	35.4 [8.3]	39.9 [6.6]	24.5 [8.9]
80-100	47.800 (53.1%)	42.4 [8.7]	41.4 [6.4]	15.9 [7.3]

Table 2. Mimic of the distribution of fertile (green), possibly fertile (red), and undefined (yellow) days depending on the standard deviation [sigma].

Cycles (n = 3) 100% measured

Sigma [C]	infertile (green)%	Fertile (red)%	Undefined (yellow)%
0.05 (low noise)	56%	40%	4%
0.10	54%	38%	7%
0.20	52%	38%	9%
0.30 (high noise)	43%	39%	17%

of this sub-sample was 32.5 [SD 8.4] years, an average of BMI 23.8 [7.3] kg/m² , and measured their BBT 87.6% of their cycle [SD 7.4] The average cycle length was 30.1 [SD 6.4] days. The mean cycle length was 30.1 [SD 6.4]. Overall, 39.6% [SD 5.8] of cycles were identified as possibly fertile (red) days, 41.8% [SD 9.4] infertile (green), and 16.5% [SD 8.1] undefined (yellow) days.

The longest pre-ovulation phase with 31.0 [3.2] days on average was found in the group with long cycles, the shortest phase with 12.3 [0.5] days on average in the group with short cycles. The post-ovulation phase is similarly stable in all groups with 12.9 [1.5] days on average. Under normal use (Gr.2; >75% measured), on average 44% of days during the menstrual cycles were identified as green days. The number of these green days is increased to a maximum of 49.4% with the 'ideal use' (Gr.1; >90% measured) of the Fertility Tracker.

For users under both normal and ideal use, 0.5% of the displayed green days were identified as 'false positive' green days in the six day fertile window or two days after the significant temperature rise. Unsurprisingly, for users with long cycles (Gr.3) (on average 44.9 days) there were 45.1% red days and 27.4% green days. Based on the earliest temperature rise observed in previous cycles, the programmed algorithm for the pre-ovulation phase assumes that all days could be potentially fertile. Compared to

Table 3. Distribution of fertile (green), possibly fertile (red) and unsafe (yellow) days depending on the standard deviation [sigma].

Sigma [C]	Gr.1 Cycles (n = 79.031) >60 %measured				Gr.2 Cycles (n = 25.085) >90 %measured			
	Cycles (n)	% infertile (green)	% fertile (red)	% uncertain	Cycles	% infertile (green)	% fertile (red)	% uncertain
0.10	27.795 (35.2%)	41.4 [13.5]	39.2 [11.3]	17.8 [19.7]	9785 (39.0%)	45.7 [13.9]	40.1 [10.6]	13.5 [7.9]
0.20	49.978 (63.2%)	38.6 [12.3]	40.3 [10.9]	19.5 [18.4]	14.868 (59.3%)	43.3 [8.3]	41.6 [5.7]	14.9 [7.2]
0.30	1250 (1.6%)	30.8 [18.3]	45.2 [14.8]	22.6 [21.7]	427 (1.7%)	33.0 [7.7]	45.6 [7.3]	20.8 [9.6]

regular cycles in Table 4, 0.3% [SD 0.1] of short and 1.5% [SD 0.2] of irregular cycles 1.5% [SD 0.6] showed significantly (****t-test, $p < 0.0001$) more false positive green days. Compared to regular cycles, significantly fewer (****t-test, $p < 0.0001$) false green days were identified among the group with a high temperature step between the pre- and post-ovulation phase. It is noticeable that half of these false positive days (51.8%) were displayed incorrectly when the fertile window was identified immediately after or during menstruation.

For the group with short cycles (Gr.4) that has (cycle length on average 23.5 days) there were 44.3% green and 39.3% red days which corresponds to the ratio of 'normal use.' This group has the most false positive green days in the analysis. However, on closer analysis shows that the majority of all false positive green days (64.3%) are related to the fifth and fourth day before the significant temperature rise respectively.

A total of 0.6% [SD 0.1] of all displayed green (infertile) days by the Fertility Algorithm were identified as false positives in the fertile window. Table 5 shows the distribution of all false green days (FGD) over the different days of the fertile window. The highest percentage of false positive green days (49.42%) was detected 5 days before the significant temperature rise (TR). On the day before the temperature rise and the expected day of ovulation the probabilities were that the lowest percentage of false green days was determined before the day of temperature rise (4.88%) and on the day of temperature rise (6.32%) itself.

Discussion

Our findings indicate that the fertility tracker was able to adapt a diverse set of conditions and cycle characteristics and provide personalised fertility statuses for users based on daily BBT readings and menstruation data. We observed a wide variation in BBT measurement with an overall average of 69.77% [SD 23.5] measured cycle days. A majority of women (53.1%) were using the device 80-100% of the time during their cycle. We identified a direct linear relationship between the number of measurements and output of the fertility tracker. As women continuously measure their BBT with the device, the less undefined (yellow) days reported to the user, thus showing the personalisation of fertility information.

For users who measured temperature 80-100% of days, the ratio of green (42.4%) to red (41.4%) days was close to balanced and closely matched to the number of fertile and infertile days displayed by other methods such as the symptothermal method [18]. This data shows that a majority of women are able to use the device correctly and measure their temperature consistently enough through the cycle over time. Less than 1% (0.6%) of green days

were identified as false positive within the six day fertile window or two days after the significant temperature rise. A closer examination of the false positive green days showed that the largest fraction (49.42%) was displayed five days before the significant temperature increase (Table 5). According to Fertility Tracker, the probability that fertile couples will successfully fertilise five days before the significant temperature rise is 6.8% [19].

Händle and Wahlström highlight that when using the output from a proposed ovulation detection algorithms (like the algorithm used in this device), users must consider not only the uncertainty in the relative time difference between the detected temperature shift and ovulation, but also the statistical uncertainty of the detection methods due to noisy measurement [20]. We found less temperature measurement noise to have a linear influence on the increase of infertile (green) days as well as undefined (yellow) days. The results also suggest that the simulation data and the real-world data are comparable. Thus, the number of green days displayed decreases with increasing standard deviation of the measured days. The opposite is true for the days displayed in yellow by the Fertility Tracker, which increase with increasing standard deviation of the measured days. For the real-world data, the number of days displayed as red increased with increasing standard deviation. This increase occurred only with a sigma of more than 0.2 °C. Interestingly, there was no difference of 60% (45.2% red) or 90% (45.6% red) measured by users. Therefore, the algorithm is fairly robust to Gaussian temperature measurement noise, while temperature noise and skipping seems to have a direct effect on fertility identification.

Performance of fertility tracker under unique conditions

Table 4 exemplifies the Fertility tracker's ability to adapt and provide fertility status information to users of different ages, BMI, cycle lengths, and user variability. The Fertility Tracker found an average of 39.6% red, 41.8% green and 16.5% yellow days in the sensitivity analysis (n = 300). The main variability in the menstrual cycle is related to changes in the follicular phase (pre-ovulation) which averages 13–16 days. The luteal phase (post-ovulation) is usually constant at 10–16 days due to the fixed lifespan of the corpus luteum [15]. The mean length of the pre-ovulation phase in this analysis was 16.8 [SD 2.9] days while the mean post-ovulation phase length was 12.8 [SD 1.6] days long. Studies have shown that the pre-ovulation phase is on average longer than generally assumed, with an average of 16.9 days for pre-ovulation and 12.4 days post ovulation. Furthermore, a strong linear correlation between menstrual cycle length and the pre-ovulation phase length with

Table 4. Numbers of Cycles, the mean of age and BMI; the percentage of infertile, possible fertile, undefined days; the percentage of false positive fertile (green) days and monophasic cycles; the mean cycle, Pre- and Post-Ovulation Phase length, the mean Pre- and Post-Ovulation temperature step in correlation to different Cycle scenarios.

Cycle Scenario	cycles	age	BMI	measured (%)	% infertile (green)	% possible fertile (red)	% undefined (yellow)	% false positive (green)	Mono (%)	cycle length (days)	Pre-O Phase (d)	Post-Ov Phase (d)	Temp Pre-Ov (C°)	Temp Post-Ov (C°)
Gr.1 ideal use (n = 20)	540	29.7 [3.1]	21.6 [2.6]	94.3 [2.5]	49.4 [4.5]	39.4 [3.1]	11.2 [3.3]	0.5[0.2*]	2.1 [0.7*]	29.6 [2.8]	16.8 [2.6]	12.8 [0.8]	36.2 [0.1]	36.5 [0.1]
Gr.2 normal use (n = 20)	589	30.1 [6.8]	21.2 [2.7]	87.0 [6.1]	44.2 [7.8]	37.9 [6.6]	15.5 [4.5]	0.5 [0.2*]	4.3 [1.9*]	29.2 [3.1]	16.1 [2.6]	13.1 [1.5]	36.2 [0.2]	36.5 [0.2]
Gr.3 long cycles (n = 20)	279	29.4 [5.2]	22.3 [3.1]	87.2 [6.4]	27.42 [9.6]	45.1 [5.5]	27.1 [10.2]	0.9 [0.3*]	10.9 [3.3*]	44.9 [4.1]	31.0 [1.9]	13.9 [1.9]	36.1 [0.1]	36.4 [0.2]
Gr.4 short cycles (n = 20)	913	36.0 [4.8]	21.6 [2.4]	87.6 [7.6]	44.3 [4.9]	39.3 [4.1]	15.7 [4.5]	1.5 [0.2*]	5.3 [1.5*]	23.5 [0.5]	12.3 [0.5]	11.2 [0.6]	36.3 [0.1]	36.5 [0.1]
Gr.5 oldest (n = 20)	841	47.5 [1.7]	23.5 [5.1]	90.5 [5.8]	44.1 [5.6]	39.6 [4.3]	16.0 [4.2]	0.6 [0.2*]	7.5 [2.4*]	28.1 [7.1]	15.2 [1.3]	12.9 [1.0]	36.2 [0.1]	36.4 [0.2]
Gr.6 youngest (n = 20)	516	19.4 [0.1]	21.1 [1.6]	87.2 [6.4]	40.1 [8.3]	41.4 [4.5]	17.3 [5.9]	1.1 [0.4*]	5.7 [1.6*]	30.1 [2.9]	18.1 [2.5]	12.9 [1.2]	36.2 [0.1]	36.4 [0.1]
Gr.7 pregnancies (n = 20)	660	28.7 [5.0]	21.8 %4.8&	78.3 [7.7]	35.8 [5.2]	31.4 [3.2]	13.2 [4.0]	0.5 [0.2*]	5.6 [1.5*]	30.1 [5.2]	17.2 [2.3]	13.0 [2.3]	36.1 [0.2]	36.4 [0.2]
Gr.8 irregular cycles (n = 20)	309	30.6 [4.3]	21.6 [2.3]	84.7 [5.8]	29.5 [10.6]	39.1 [6.7]	30.4 [12.6]	1.5 [0.6*]	10.9 [3.0*]	39.0 [5.4]	25.2 [5.0]	13.8 [2.1]	36.2 [0.1]	36.4 [0.3]
Gr.9 regular cycles (n = 20)	613	30.4 [5.1]	23.2 [4.5]	90.3 [6.0]	49.1 [3.4]	28.2 [2.5]	12.8 [2.8]	0.3 [0.1*]	2.7 [0.6*]	26.7 [1.2]	14.0 [1.3]	12.8 [1.1]	36.1 [0.1]	36.4 [0.1]
Gr.10 high temp step (n = 20)	770	33.9 [7.2]	22.4 [1.9]	90.2 [5.0]	51.5 [3.8]	38.2 [1.9]	10.3 [3.2]	0.2 [0.05*]	0.8 [/]	27.2 [1.7]	14.5 [1.3]	12.7 [1.3]	36.1 [0.2]	36.4 [0.1]
Gr.11 low temp step (n = 20)	714	36.5 [5.1]	24.2 [4.7]	89.4 [4.7]	43.8 [5.4]	39.6 [3.8]	16.6 [3.6]	0.3 [0.1*]	5.8 [1.2*]	26.6 [1.1]	13.9 [1.1]	12.7 [1.0]	36.3 [0.2]	36.5 [0.2]
Gr.12 high BMI (n = 20)	536	34.4 [6.9]	39.7 [10.6]	89.0 [6.3]	41.7 [6.7]	41.8 [4.4]	15.5 [4.8]	0.2 [0.08*]	7.6 [3.6*]	28.1 [3.0]	15.7 [2.7]	12.4 [1.36]	36.4 [0.1]	36.4 [0.1]
Gr.13 low BMI (n = 20)	767	31.3 [5.6]	17.4 [0.6]	90.3 [6.8]	43.7 [6.5]	41.0 [5.0]	14.9 [4.5]	0.5 [0.2*]	7.2 [2.6*]	29.0 [4.0]	16.2 [3.0]	12.8 [1.5]	36.3 [0.1]	36.5 [0.1]
Gr.14 long use (n = 20)	1859	35.3 [6.6]	24.3[3.8]	84.3 [11.9]	43.5 [6.7]	39.4 [4.0]	15.1 [5.6]	0.3 [0.1*]	3.1 [0.7*]	28.6 [1.9]	16.3 [2.0]	12.2 [1.0]	36.1 [0.2]	36.4 [0.1]
Gr.15 short use (n = 20)	28	29.5 [3.5]	26.7 [3.8]	84.4 [8.8]	40.1 [7.4]	42.2 [11.2]	16.0 [10.2]	0	0	27.8 [2.8]	14.4 [1.9]	13.4 [1.5]	36.2 [0.2]	36.5 [0.2]
Gr.16 All (n = 300)	9934	32.5 [8.4]	23.8 [7.3]	87.6 [7.4]	41.8 [9.4]	39.6 [5.8]	16.5 [8.1]	0.6 [0.1*]	5.3 [0.5*]	30.1 [6.4]	17.3 [5.7]	12.9 [1.5]	36.2 [0.2]	36.5 [0.2]

Table 5. Relationship between the distribution of false positive green days and the probability of pregnancy.

	−5	−4	−3	−2	−1	Rise in temperature	+1	+2
Total distribution of false green days in relation to the fertile window	49.4%	14.9%	6.9%	4.3%	4.9%	6.3%	7.5%	5.7%
Probability of pregnancy on the different days of the cycle [19]	6.8%	17.6%	23.7%	25.5%	21.2%	10.3%	0.8%	0.4%

increasing age has been demonstrated in a prior study [16]. In the present work the same trend was found. The pre-ovulation phase was significantly shortened from 17.05 [3.9] days (age 25-40) to 15.07 [4.2] days (age 40+).

Limitations

There were several limitations to this study. The retrospective nature of this study is limiting, although the data set was large and robust. Furthermore, a large proportion have not given their demographic data. Thus, for the population's characteristic, only 52.8% of the data set could be used for the evaluation. Another limitation of this study is that there is no information on whether pregnancies were desired and how long the user used the device to fulfil this desire. The determination of ovulation was done retrospectively, therefore it cannot be excluded that ovulation occurred at a later time. To minimise this limitation, the six-day fertile window in this study was extended by two days (after the significant temperature increase).

Conclusion

The data shows that women are able to use the device correctly and measure their temperature more consistently throughout the cycle, thus self-efficacy of Fertility Tracker increases over time. Consistent BBT measurement provides more data for better performance of the device and less undefined (yellow) days by Fertility Tracker. The analysis of the data has shown that the temperature shift algorithm used by Fertility Tracker is able to exclude the fertile window with very high accuracy and to detect the different phases of the menstrual cycle. Further research is needed to explore the efficacy of the device in a prospective study.

Acknowledgements

Disclosure statement

Funding

ORCID

Niels van de Roemer http://orcid.org/0000-0003-2351-5472
Liya Haile http://orcid.org/0000-0001-9170-4879

References

[1] Wilcox AJ, Weinberg CR, Baird DD. Timing of sexual intercourse in relation to ovulation. Effects on the probability of conception, survival of the pregnancy, and sex of the baby. N Engl J Med. 1995;333(23):1517–1521.

[2] Lynch CD, Jackson LW, Buck Louis GM. Estimation of the day-specific probabilities of conception: current state of the knowledge and the relevance for epidemiological research. Paediatr Perinat Epidemiol. 2006;20(Suppl 1):3–12.

[3] Pyper CM. Fertility awareness and natural family planning. Eur J Contracept Reprod Health Care. 1997;2(2):131–146.

[4] Weschler T. Taking charge of your fertility; 2006. PMID: 24988816.

[5] Daniels K, Mosher WD. Contraceptive methods women have ever used: United States, 1982–2010. Natl Health Stat Rep. 2013;62:1–15. PMID: 24988816.

[6] Manhart MD, Duane M, Lind A, et al. Fertility awareness-based methods of family planning: a review of effectiveness for avoiding pregnancy using SORT From the. Osteopath Fam Physician. 2013;5(1):2–8.

[7] Brezina PR, Haberl E, Wallach E. At home testing: optimizing management for the infertility physician. Fertil Steril. 2011; 95(6):1867–1878.

[8] Pallone SR, Bergus GR. Fertility awareness-based methods: another option for family planning. J Am Board Fam Med. 2009;22(2):147–157.

[9] Fehring RJ, Barron ML, Schneider M. Protocol for determining fertility while breastfeeding and not in cycles. Fertil Steril. 2005; 84(3):805–807.

[10] Freundl G, Frank-Herrmann P, Godehardt E, et al. Retrospective clinical trial of contraceptive effectiveness of the electronic fertility indicator Ladycomp/Babycomp. Adv Contracept Off J Soc Adv Contracept. 1998;14(2):97–108.

[11] Freundl G, Frank-Herrmann P, G C. Cycle monitors and devices in natural family planning. Endokrinol. 2010;7:90–96.

[12] Duane M, Contreras A, Jensen ET, et al. The performance of fertility awareness-based method apps marketed to avoid pregnancy. J Am Board Fam Med. 2016;29(4):508–511.

[13] Mangone ER, Lebrun V, Muessig KE. Mobile phone apps for the prevention of unintended pregnancy: a systematic review and content analysis. JMIR Mhealth Uhealth. 2016;4(1):e6.

[14] Creinin MD, Keverline S, Meyn LA. How regular is regular? An analysis of menstrual cycle regularity. Contraception. 2004; 70(4):289–292.

[15] Cole LA, Ladner DG, Byrn FW. The normal variabilities of the menstrual cycle. Fertil Steril. 2009;91(2):522–527.

[16] Bull JR, Rowland SP, Scherwitzl EB, et al. Real-world menstrual cycle characteristics of more than 600,000 menstrual cycles. NPJ Digit Med. 2019;2:83.

[17] Kleinschmidt TK, Bull JR, Lavorini V, et al. Advantages of determining the fertile window with the individualised Natural Cycles algorithm over calendar-based methods. Eur J Contracept Reprod Health Care. 2019;24(6):457–463.

[18] Lupinska-Dubicka A, Druz MJ. A comparison of popular fertility awareness methods to a DBN model of the woman's monthly cycle. Proceedings of the 6th European Workshop on Probabilistic Graphical Models, PGM; 2012. p. 219–226.

[19] Colombo B, Masarotto G. Daily fecundability: first results from a new data base. Demogr Res. 2000;3:39.

[20] Händel P, Wahlström J. Digital contraceptives based on basal body temperature measurements. Biomed. Signal Process. Control. 2019;52:141–151.

Appendix

Preselection of data set up for analysis

The following guidelines were used for selecting 16 data sets of 20 files each out of the basic 5329 Metadata files for the sensitive analysis. The columns were filtered/sorted according to the descriptions below. The files are only allowed to be in one set, and the sets were generated started with a.

a. **ideal use**
   a. 20 data sets
   b. randomly selected (Random generator)
   c. days between 365 and 1095
   d. % measured > 90%
   e. age between 18-35

b. **normal use**
   a. 20 data sets
   b. randomly selected (Random generator)
   c. days between 365 and 1095
   d. % measured > 75%

c. **long cycles**
   a. 20 data sets
   b. days between 365 and 1095
   c. % measured > 75%
   d. min. 6 completed cycles
   e. max length of the cycle 50 days on average
   f. length mean sorted descending, top values chosen

d. **short cycles**
   a. 20 data sets
   b. days between 365 and 1095
   c. % measured > 75%
   d. length mean sorted ascending, top values chosen

e. **oldest**
   a. 20 data sets
   b. days > 365
   c. % measured > 75%
   d. age sorted descending, top values chosen

f. **youngest**
   a. 20 data sets
   b. days > 365
   c. % measured > 75%
   d. age sorted ascending, top values chosen

g. **pregnancies**
   a. 20 data sets
   b. days > 365
   c. % measured > 75%
   d. pregnancy sorted descending, top values chosen

h. **irregular cycles**
   a. 20 data sets
   b. days > 365
   c. % measured > 75%
   d. cycle length std. dev. sorted descending, top values chosen

i. **regular cycles**
   a. 20 data sets
   b. days > 365
   c. % measured > 75%
   d. cycle length std. dev. sorted ascending, top values chosen

j. **high temp step**
   a. 20 data sets
   b. days > 365
   c. % measured > 80%
   d. temp. step avg. sorted descending, top values chosen

k. **low temp step**
   a. 20 data sets
   b. days > 365
   c. % measured > 80%
   d. temp step avg sorted ascending, top values chosen

l. **high BMI**
   a. 20 data sets
   b. days > 365
   c. % measured > 80%
   d. BMI sorted descending, top values chosen

m. **low BMI**

   a.   20 data sets
   b.   days > 365
   c.   % measured > 80%
   d.   BMI sorted ascending, top values chosen
n.   **long use**
   a.   20 data sets

   b.   % measured > 60%
   c.   days sorted ascending, top values chosen
o.   **short use**
   a.   20 data sets
   b.   days in use < 90
   c.   randomly selected (Random generator)