Internal Q&A document on the study „The Performance of a Fertility Tracking Device“? (2021)

As of: 12.01.2021

To be used by persons authorized to answer media inquiries regarding the 2021 study „The Performance of a Fertility Tracking Device“. This document is not suitable for distribution to third parties!

Study

What was the Background of the study “The Performance of a Fertility Tracking Device“ ?

Fertility Awareness-Based Methods (FABM) are a set of family planning methods based on a woman‘s periodic fertility. bFABMs 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). Research has found that 20% of all women in the United States have tried a FABM at some point in their lives.

Typically, use of FABM requires an understanding of one‘s own biology, discipline and the proper educational training. Furthermore, reading and interpreting data is open to human error, potentially reducing the method’s efficacy. 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. 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. 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

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.

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 delimited. 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).

What is the aim of the study "The Performance of a Fertility Tracking Device" ?

The objective of this study was to better understand and identify any errors or discrepancies in Fertility Tracker's results. Further, we aimed to analyze different physiological factors that directly could influence the outcome of the algorithm (i.e. age, BMI, cycle length, measurement skipping, high vs low average temperature, temperature steps).

The results of this study provides insights on how machine learning algorithms in fertility tracking devices respond to physiological changes For the analysis, 107,000 cycles from 5328 women were evaluated!

How was the study designed?

For the scientific study, 107,000 cycles (basal temperature, menstrual input) of 5,328 women from Germany and Switzerland were evaluated over a period of ten years

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.

What was analyzed?

1. 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%.

2. The impact of temperature noise (temperature outlier) was also assessed in an independent model, through theoretical simulation testing. Fertility Tracker's

algorithm was fed with defined standard deviation temperature values o 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.

3. 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.

What are the results of the study?

Sample Characteristics of the study

- The average age of the women was 30 years, with a body mass index (BMI) of 22.07 (normal weight is 18.5 to 25).
- For the Analysis, the fertility tracker was used for 22 complete cycles on average
- The average cycle length was 29.5 days
- The average length of the follicular phase (before ovulation) was 16.8 days, while the average length of the lutheal phase was 12.8 days. The study shows that only 12.5% of all cycles were 28 days long.

Analysis of the fertility Output (1)

- Of the 53.1% of users who used the Fertility Tracker 80-100% of their cycle, an average of 41% potentially fertile (red) days, 42% infertile (green) days were calculated. For users measuring their BBT for 60-80% of their cycle, the device on average identifies 35.4 infertile (green), 39.9 possibly fertile (red), and 15.9 undefined (yellow) days.

Analysis of temperature noise (2)

- Temperature Outliers has a direct influence on the output of the fertility algorithm. When this outliers are very low (0.05°C), the algorithm provides more green (infertile) days (56%) and the least yellow (undifined) (4%) days. If these outliers are extremely strong (0.30°C), Fertility Tracker displayed relatively less green (43%) and more yellow days (17%).

Sensitivity Analysis (3)

- Overall, 0.6% of all green (infertile) days displayed by the Fertility Tracker were calculated incorrectly, therefore in the user's fertile window. Thus, the Fertility Tracker has an accuracy in calculating infertile (green) days of 99.4%.
- Of these 0.6% incorrect days designated as infertile, more than half were five days before the expected day of ovulation and thus have only a 6.8% chance of pregnancy.

What is the conclusion of the Study?

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.