Analyzing Commercial Fertility Apps: Algorithmic Predictions and Users’ Perceptions

Mayara Costa Figueiredo¹, H. Irene Su², Yunan Chen¹
¹University of California, Irvine, ²University of California, San Diego

Motivation

More often patients are bringing apps to clinical encounters; it becomes critical to analyze the support apps offer and how patients interact with predictions. The goal of this study is to evaluate current commercially available fertility apps focusing on their algorithmic feedback and users’ experiences.

App Selection

<table>
<thead>
<tr>
<th>Apps selected for eligibility (n = 50)</th>
<th>Apps excluded (n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 out of 30 apps (60%)</td>
<td></td>
</tr>
<tr>
<td>Changes in 29 out of 30 apps (96.6%)</td>
<td></td>
</tr>
</tbody>
</table>

Algorithmic Feedback

Table 1. Final list of evaluated apps. Three apps appeared in both stores but were individually analyzed as features may vary.

Dataset: Four months of data

In April 2019, three researchers entered a single dataset simulating four months of regular fertility cycles:

- Cycle length: 28 days
- Period length: 6 days
- First period: 12/07 – 12/12
- Second period: 01/04 – 01/09
- Third period: 02/01 – 02/06
- Fourth period: 03/01 – 03/06
- Temperature: ~97.5°F until 3 days before ovulation, followed by ~98.2°F
- Ovulation: positive ovulation 3 days before predicted
- Cervical Mucus: 3 days of egg-white CM 2 days before predicted ovulation

Each app was analyzed by two researchers who entered the dataset and recorded how different types of data visibly change the predictions.

What tracked data affected predictions:

- Changes in the first day of last period affected predictions in 29 out of 30 apps (96.6%)
- Changes in previous periods affected predictions in 18 out of 30 apps (60%)
- Positive ovulation tests affected predictions in 13 out of 20 apps (65%)
- Tracked temperature affected predictions in 6 out of 25 (24%)
- Tracked cervical mucus affected predictions in 1 out of 22 apps (4.5%)

User Reviews: perceptions and opinions

Two researchers qualitatively analyzed 3,214 open-ended reviews using text open and axial coding and thematic analysis, following a grounded approach.

Users have mixed reactions towards the apps, some fully trust algorithmic predictions, but many are confused about how the apps make predictions and why the predictions are incorrect comparing to the results of OPKs, or inconsistent with other fertility apps they use.

- "Best app. If you periodically track, I guarantee you will get pregnant or avoid pregnancy"

- "After trying to conceive for some months, I started using OPKs and found that my fertile window started earlier than the app predicted. […] I have been missing my fertile window all this time"

- "It gave me false hope. I am extremely upset and down. I was using the app to track my fertility. It not only got my fertile and ovulation dates wrong; it also got my period dates wrong. I was super excited cause I missed a period, so I thought I conceived. It was incredibly hard to accept that I’m not pregnant again!"

Algorithmic Feedback

Figure 2. Prediction for fertility and ovulation window. First day of fertility window varied by five days, with 13 apps predicting the same start day. Ovulation day varied by two days, with 15 apps predicting the same day and 2 apps not providing predictions. Length of fertility window ranged from 3 to 14 days (mean±SD: 8.7±2.036).

User reviews

Figure 3. user reviews selection process

Our study identified inconsistencies in fertility predictions from the popular fertility apps we reviewed. This variance was considered important to many users who described frustration when they noticed them (even when the fertility window overlapped). Besides, other than period dates, most data tracked by users do not lead to changes in predictions, which suggests that indicators that may require daily and disciplined work are not used.

Our analysis show that the lack of clear description of what data are used in making fertility predictions can cause potential tracking burden, distrust of fertility technologies, or over-trust in predictions that may not be accurate. These issues may further affect users’ fertility experiences and their interactions with healthcare providers.

This study shows fertility technologies have (i) to be designed with more transparency regarding their algorithmic feedback and (ii) to make the uncertainty intrinsic to fertility (e.g., predictions are not 100% accurate or predictive of fertility) more visible in apps’ graphs and visualizations to help users set realistic expectations.

Conclusion

Contact: Mayara Costa Figueiredo (mcostafi@uci.edu)