

An Evaluation of 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 and App Selection

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

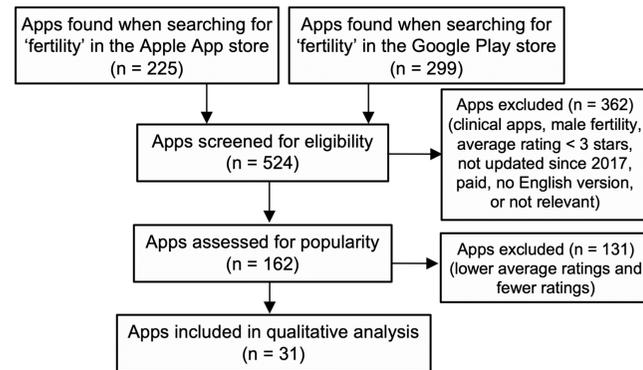


Figure 1. app selection process

| Apple App Store | Google Play Store |
|------------------------------------------|------------------------------------------|
| Flo (4.8, 365K) | Period Calendar (4.8, 4M) |
| Glow (4.7, 29K) | Flo (4.9, 795K) |
| Life (4.8, 27K) | Clue (4.8, 627K) |
| Clue (4.8, 24K) | Period Tracker (GP) (4.5, 334K) |
| Ovia (4.8, 12K) | My Calendar (4.7, 185K) |
| Cycles (4.5, 7K) | Maya (4.7, 173K) |
| Period Tracker Health Calendar (4.5, 4K) | Pepapp (4.7, 155K) |
| Kindara (4.6, 3K) | Petal (4.8, 140K) |
| Natural Cycles (4.8, 3K) | Lilly (4.5, 140K) |
| My Calendar (4.8, 2K) | WomanLog (4.5, 123K) |
| Ferdy (4.5, 2K) | Period Tracker (Amila) (4.9, 119K) |
| Dot (4.7, 1.7K) | Woman Diary (4.6, 94K) |
| Femometer (4.8, 1.6K) | My Days (4.5, 93K) |
| My Cycles (4.6, 1.4K) | Period Tracker (Leap Fitness) (4.9, 83K) |
| Premom (4.7, 1.3K) | Ladytimer (4.5, 72K) |
| Monthly Cycles (4.6, 1.2K) | |

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

User reviews

User Reviews: perceptions and opinions

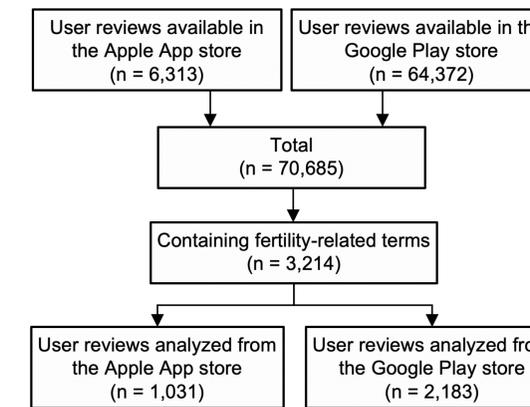


Figure 3. user reviews selection process

Two researchers qualitative analyzed all 3,214 open-ended reviews text using 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 days wrong; it also got my period days 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

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: ~97.5F until 3 days before ovulation, followed by ~98.2F
- 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 dates** 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%)

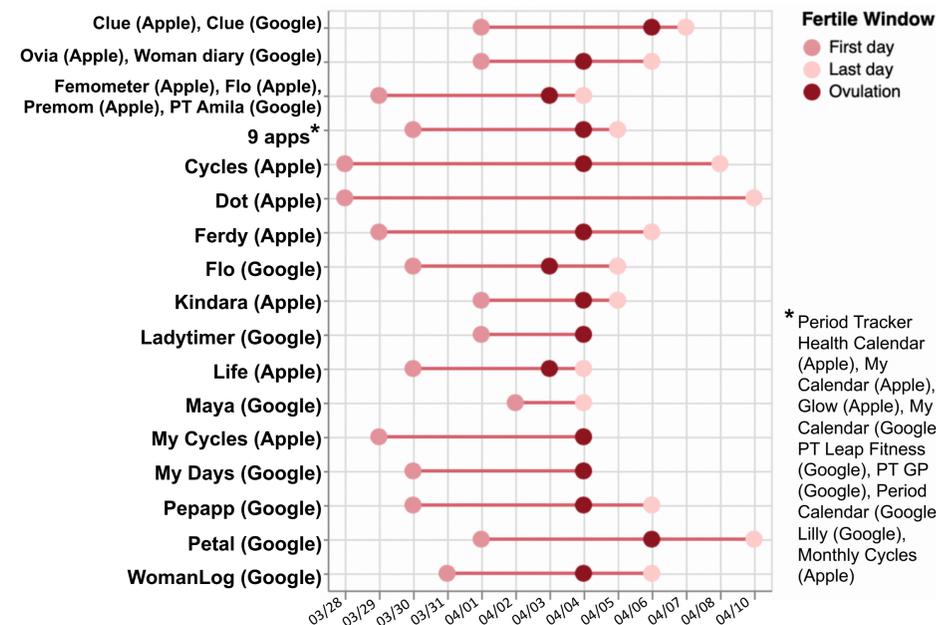


Figure 2. Predictions for ovulation and fertile window. First day of fertile window varied by five days, with 13 apps predicting the same start day. Ovulation day varied by five days, with 19 apps predicting the same day and 2 apps not providing predictions. Length of fertile window ranged from 3 to 14 days (mean=7.167, sd=2.036).

Conclusion

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

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

