"When Am I Fertile?": A Pilot Study Comparing Ovulation Prediction Accuracy of Apps and LH Kits

Molly Enenbach[1], Cassandra Haworth[1], Camille Hawkins[1] and Matthias Kochmann[1]

[1]A.T. Still University, School of Osteopathic Medicine in Arizona, Mesa, AZ, USA

ABSTRACT

Prediction of peak fertility is critical yet challenging in both planning and preventing pregnancy. Period tracking applications for the smartphone are ubiquitous, free of charge, and user-friendly with many providing ovulation estimates. The objective of this study was to analyze the period tracking applications' (apps) ability to accurately predict fertility windows and ovulation. Three medical students tracked their menstrual cycle over four months in seven commercially available menstrual period tracking applications. Six of the apps were analyzed for fertility window, ovulation prediction, and usability. Two home ovulation kits were utilized to confirm ovulation. The sensitivity to predict the fertility window ranged from 35% to 94% (p<0.05) while sensitivity to predict ovulation ranged from 0% to 31% (p>0.05). Four of the apps allowed for menstrual cycle lengths greater than 35 days and offered an adjustable algorithm. Apps had increased sensitivity due to an expanded fertility window with the increased number of predicted fertile days, but a low ovulation sensitivity and an inability to predict the day of ovulation. Additionally, apps allowed for additional personal information to be added with some apps sharing this data with a third party, raising the question of data protection for users. Solely using period tracking apps is not the gold standard for contraception or conception. The use of these apps in conjunction with luteinizing hormone home kits for detection of physiologic ovulation provides an accurate tool that allows a woman to take charge of her reproductive health.

Introduction

Predicting peak fertility, to either plan or prevent pregnancy, is complicated due to a brief fertility window surrounding ovulation.[1–3] The confusion and complication of predicting this fertility window pose respective challenges for different populations of women. For women not desiring pregnancy, the inability to accurately calculate their day of ovulation poses an increased risk for unplanned and potentially unwanted pregnancy. For women wanting to conceive, correctly timing intercourse based on ovulation increases the likelihood of success.[4] On one hand, 45% of pregnancies are unintended while on the other hand, 13.1% of those purposefully attempting to become pregnant between the ages of 15-49 have decreased fecundity.[5] Tracking ovulation remains a difficult task despite the high demand to decrease subfertility and unintended pregnancies.

Only 12.7% of couples correctly identified their fertile window in a study among 282 highly motivated patients seeking care from assisted reproductive technology clinics due to subfertility.[6,7] Additionally, variation in the menstrual cycle makes it difficult to predict ovulation.[7,8]

Medical physiology textbooks state that the duration of the adult menstrual cycle varies from 20-45 days with an average cycle lasting 28 days.[9] This is due to a highly variable follicular phase of roughly 14 to 21 days before ovulation, and a minimally variable luteal phase of 14 days after ovulation (Figure 1). However, in a 2019 study of real-world menstrual cycle characteristics (n=600,000), variation in cycle length and both phases were demonstrated: Cycle lengths ranged from 10-90 days; the mean follicular phase was 16.9 days; the mean luteal phase was found to be shorter at 12.4 days.[10]
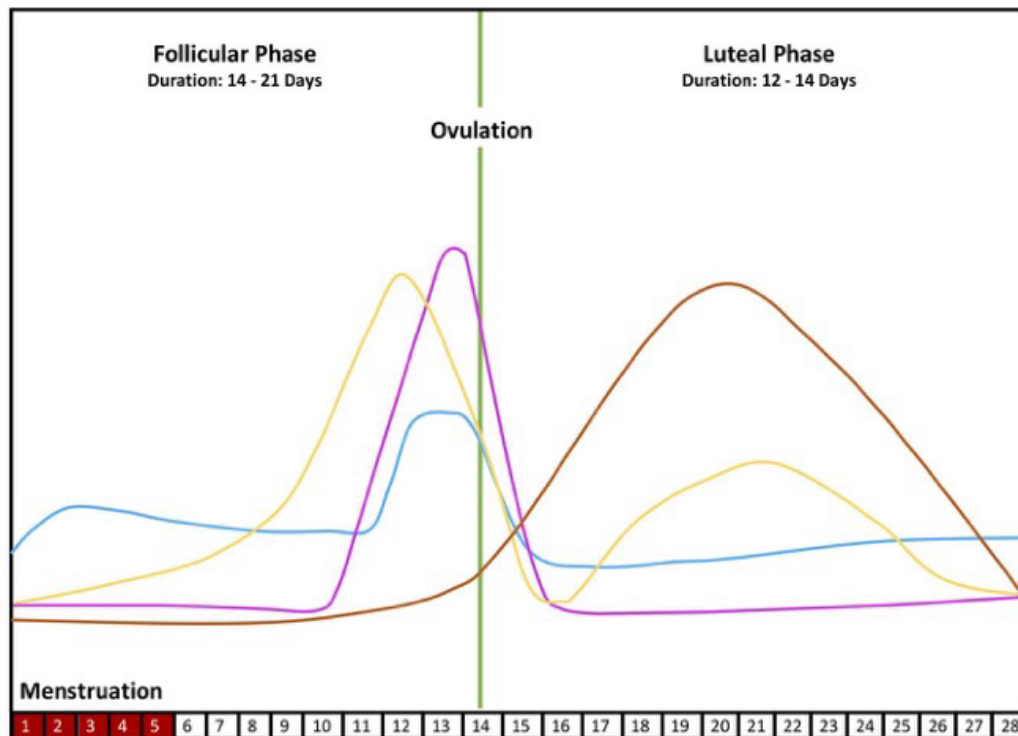
Figure 1: Physiologic Menstrual Cycle

The transition from the follicular phase to the luteal phase marks the day of ovulation and peak fertility, but the fertility window begins approximately 3–5 days before ovulation (accounting for sperm lifespan) and continues approximately 1–2 days after ovulation (accounting for oocyte lifespan).[11]

Although many methods exist to assist in predicting ovulation, as well as menstruation, no current studies have compared period and ovulation tracking apps vs home luteinizing hormone (LH) ovulation kits. Both methods are widely accessible and provide users with a noninvasive means of predicting their fertile window to be used for contraception or conception. The apps predict the fertility window, ovulation, and the start of the next period based on manually entered user data whereas the LH kit measures the surge of luteinizing hormone in the urine to predict ovulation.

Within one month, the keywords ovulation, conception, and pregnancy had over a quarter of a million internet searches in the United States demonstrating the interest for information about fertility planning and the menstrual cycle.[12] We conducted a pilot study of adult females in which we compared the accuracy of ovulation prediction between home LH ovulation kits and period tracking apps from the Apple App Store. In addition, we also examined the usability and services provided by each of the apps. In this paper, we discuss our results with the hope that our data will provide information as well as grounds for additional studies that aim to empower women by giving them better control over their reproductive health.

Materials and Methods

Seven commercially available menstrual period tracking applications were chosen from the Apple app store based on user rating (1 to 5 stars) and cost. All seven apps could be downloaded at no cost. Three apps were chosen based on a

user rating of four stars or higher. Next, two apps were selected in the three-star range. Finally, two apps were chosen with less than three stars. Attention was also paid to the number of total reviews and selections were made based on a high number of reviews (over 500,000), middle point (almost 8,000), and a small number of reviews (below 100). Additional characteristics are included in Table 1.

Two luteinizing hormone (LH) home kits were chosen from Amazon based on user reviews and cost. The first, was a well-known brand with four stars while the second kit was chosen based on a much lower cost and one-star rating (Table 1).

Table 1: Selection Criteria of Ovulation Prediction Methods.

Method	Cost	Stars	Total reviews	Category
Apps				
B	Free	4.8	524K	Hybrid
G	Free	4.6	7.8K	Calendar-based
A	Free	4.1	115	Hybrid
D	Free	3.8	240	Calendar-based
C	Free	3.2	22	Calendar-based
F	Free	2.8	10	Calendar-based
E	Free	1	2	Calendar-based
LH Kits				
Clearblue	35.98	4.1	2908	Quantitative hormone assay
Healva	$11.99	1.9	5	Quantitative hormone assay

Initially, four female graduate students from A.T. Still University (ATSU) were both participants and investigators in the study. One participant dropped out of the study during the first month. The remaining three participants tracked their menstrual period in all seven apps for three menstrual cycles. The first cycle was used to estimate the duration of each individual cycle to allow for more accurate testing with the LH kit. For a 28-day cycle, testing was recommended on day 11 or 12 with the Clearblue and Healva LH kits, respectively. Cycles shorter than 28 days would need to be tested earlier while longer cycles would be tested later (See Table 2). The LH kits were used as the gold standard for ovulation as a previous study showed a 96% accuracy at prediction with LH kits, confirmed by ultrasonography.[13]

Table 2: Cycle Chart.

Cycle Length	Clearblue (Day of Cycle to Begin Testing)	Healva (Day of Cycle to Begin Testing)
21 days	5	5
22 days	5	6
23 days	6	7
24 days	7	8
25 days	8	9
26 days	9	10
27 days	10	11
28 days	11	12
29 days	12	13
30 days	13	14
31 days	14	15
32 days	15	16
33 days	16	17
34 days	17	18
35 days	18	19
36 days	19	20

Contingency tables constructed from the data from each app were analyzed for accuracy, sensitivity, and specificity with the application of the Chi-square test to evaluate the hypothesis. Additionally, apps and LH kits were analyzed for usability and services which can be found in the results section. The study proposal was reviewed by the ATSU Institutional Review Board.

Results

Assessment results are summarized in Table 3. The two most popular apps are labeled A and B, while the comparable apps are labeled C-G. Apps varied in data gathering methods, diagnostic quality, and services provided.

Table 3: App Interface User Friendliness, Services, and Diagnostic Options.

	App A Lily	App B Flo	App C Period Tracker: Ovulation App - by Content Arcade	App D MyDaysX	App E Menstrual Period Tracker V	App F The best Period Tracker-Ovulation Intimate Fertile and Sex Tracker!	App G Period Plus
Interface User Friendliness							
Available on iPhone and Android	iPhone	Both	iPhone	Both	iPhone	iPhone	iPhone
Stars and Number of Ratings in Apple App store	4.7 Stars 115 Ratings	4.8 Stars 572 K Ratings	3.1 Stars 23 Ratings	3.8 Stars 241 Ratings	1.0 Star 2 Ratings	2.8 Stars 10 Ratings	4.6 Stars 8.3 K Ratings
Ability to Enter Own Data	Yes - Easy	Yes - easy	No - set length and cycle length	Yes	Yes - but limited timeframe	Yes - but limited timeframe	Yes
Legend/Key /Guide	Yes	Yes	No	Yes-difficult to find	Yes	Yes	Yes
Price	Free. $5.99 one-time fee for premium.	Free. $9.99 per month for premium	Free. No option for upgrade.	Free. $2.99 one-time fee for premium.	Free. No option to upgrade.	Free. No option to upgrade.	Free $1.49 per month for premium
Advertisements	None	None	Yes	Yes - timed ones	No	None	Yes
Adaptable Algorithms	Yes	Yes	No	Yes	No	No	Yes
Medical questionnaire	None	Yes	No	No	No	No	No
Symptoms Data	Yes	Yes	No	Weight/Temp	No	No	Yes-a few, then more costs extra

Shareable Data	Yes - costs	Yes - costs	No	Yes	No	No	No
>35-day period interval	Yes	Yes	No	Yes	No	No	Yes
Reminders	Yes - costs	Yes	Yes	Yes	No	No	Yes
Health Literature included	None	Yes	No	No	No	No	Yes
Passcode	Yes	Yes	Yes	Yes	No	No	Yes
Account	No	Yes- costs	Yes	Yes	No	No	Yes- costs
Services							
Contact Support	Yes	Yes	Yes	Yes	No	No	Yes
Diagnostic Options							
Period Prediction	Yes	Yes	Yes	Yes	Yes	No	Yes
Ovulation Prediction	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Notes:		Shares information with third party	No day of ovulation given		App was not analyzed for diagnostic accuracy due to inability to enter data		

Data Collection Interface

All tested apps had a calendar type interface upon opening, showing either recent or upcoming menstrual dates and predicted fertility window. Four of the seven apps (A, B, D, and G) allowed for the user to click on the day the cycle began and enter data about the period, including the quantity of blood flow, color, cramps, and other related symptoms. These four apps also included a brief summary on the main page, predicting the next period and specified which day the user was on in the current cycle. App E allowed for cycles to be entered by tapping on the day, but the dates changed as the user tapped on the day, causing the calendar to have random dates displayed. This caused the user to spend a long stretch of time entering data, often with no success. App C and F asked for period start day, length of period, and the typical length of cycle to provide ovulation prediction and then created a static calendar that did not allow for future modification.

Both ovulation kits included printed directions and required small amounts of urine. Clearblue allowed for a direct urinary stream to device method to obtain an appropriate sample. The Healva kit required the collection of urine in a provided cup and then the strip was dipped into the collected urine. The cup was inappropriately sized with user difficulty in collecting urine to obtain sufficient samples. Clearblue prevents subjective interpretation of the ovulation results due to the automated image of a smiley face (ovulating) or an empty circle (not ovulating) on the

reusable digital reader. Healva is a more traditional test with subjective interpretation of a positive and a control line, leading to possible errors in what may or may not be a positive test. The price point between the two home LH kits is appreciable. Clearblue Digital Ovulation Test provides 20 ovulation strips and a reusable digital result reader for $35.98. Healva provides 60 ovulation strips, 20 pregnancy strips, and 80 urine cups for $11.99. Neither company disclosed the threshold of LH that was used to determine ovulation.

Diagnostic Quality

Contingency tables were created to examine the relationship between the result of the gold standard (LH Kits) and the result of each fertility app. These tables were used to compute the accuracy, sensitivity, and specificity of each application for the fertility window and day of ovulation. The results are summarized in Table 4 and Table 5.

Table 4: Accuracy, Sensitivity, and Specificity of Ovulation and Fertility.

Fertile or Ovulating	Accuracy	Sensitivity	Specificity
App F	0.83	0.94	0.81
App G	0.84	0.62	0.89
App B	0.85	0.92	0.84
App C	0.79	0.48	0.86
App D	0.83	0.35	0.93
App A	0.76	0.94	0.72

Ovulating	Accuracy	Sensitivity	Specificity
App F	0.94	0.25	0.98
App G	0.92	0.06	0.97
App B	0.92	0.06	0.97
App C	0.94	0.00	1.00
App D	0.94	0.06	0.99
App A	0.95	0.31	0.98

Table 5: Chi-square Test.

Fertile or Ovulating	Value	df	Asymptotic Significance (2-sided)
Pearson chi-square	11.891[a]	5	0.036
Likelihood ratio	11.591	5	0.041
N of Valid cases	1800		

Ovulating	Value	df	Asymptotic Significance (2-sided)
Pearson chi-square	3.337[b]	5	0.648
Likelihood ratio	3.280	5	0.657
N of Valid cases	1800		

For the fertility window, the accuracy of the applications ranged from 76% to 85%. App B had the highest accuracy (85%) followed by App G (84%) and App D (83%). There was a substantial range between apps in respect to sensitivity (35% to 94%). App F and A had the highest sensitivity at 94%. App D was the least sensitive at 35%. App D had the highest specificity at 93% followed by App G and C, at 89%, and 86% respectively. However, all of these three applications had low sensitivity. A chi-square test of independence was used to analyze the hypothesis that application is independent of the accuracy of tracking results. The p-value was 0.036 which is less than 0.05, thus the null hypothesis was rejected.

Regarding ovulation, all applications had an accuracy of greater than 90% (range: 92-95%). App A had the highest accuracy, at 95%, while Apps G and B had the lowest, at 92%. None of the applications were very sensitive but App A was the most sensitive at 31%. All of the applications had a specificity greater than 96% (range: 97-100%). App C was 100% specific but had 0% sensitivity. App G and B had the lowest specificity (97%) with 6% sensitivity for both. No statistically significant (p= 0.648) association was found between the application used and the accuracy of ovulation.

Services

All of the seven apps were free to download from the Apple App Store. Three of the seven apps (C, D, and G) displayed advertisements, with two having the option to upgrade to an ad-free experience (Apps D and G). The remaining four apps delivered an ad-free experience. Four of the apps (A, B, D, and G) offered additional features for a price: account and username, symptom data, shareable data, and reminders. App C offered reminders, account/username, and passcode for free. Apps E and F did not offer any additional features.

Apps A, B, D, and G used adaptable algorithms that readjusted cycle length and fertility window prediction based on previous period lengths and newly entered data. These same four apps allowed for period lengths greater than 35 days to be entered and/or 10 days of menses. The remaining apps (C, E, and F) had a maximum menstrual cycle length of 35 days and/or 10 days of menses.

In regards to shareability and educational resources, apps B and G provided health literature to help answer questions about women's health issues, for example, breast cancer screening and endometriosis symptoms. App A and B allowed for shareable and downloadable data while the other apps with passcode or account settings offered no method to print or email data from the phone to other devices. All apps, except for App C, provided a legend or a key to determine what the colors or symbols on the calendar represented but most were difficult to find in the app (See Table 3 above).

Discussion

In regards to predicting the fertility window, App F and A had the highest sensitivity at 94%. As previously discussed, the physiologic fertility window averages around 5 days. However, App F predicted a total of nine fertile days while App A predicted 11 fertile days surrounding ovulation. It is a reasonable assumption that having an expanded fertility window increased the sensitivity results due to an increased number of predicted fertile days. With so many potentially fertile days, the apps essentially ensured that the prediction included the physiologic fertility window plus a buffer of potentially fertile days. Having an expanded fertility window may decrease compliance while using natural family planning, including abstinence.

All of the apps are poor predictors of psychological ovulation. App A had the highest sensitivity at 31% while other apps ranged from 0 up to 25%. The apps all had a specificity of greater than 96%, showing the reliability of the apps to indicate when not ovulating. App C had 100% specificity but 0% sensitivity due to the fact that the app did not indicate a day of ovulation but a window in which ovulation occurred. App A had the highest accuracy at 95% with all other apps ranging from 92-94%. Due to the fact that the apps had low sensitivity, high specificity, and high accuracy, the apps were able to determine nonfertile days and detect true ovulation but inconsistently. While there was statistical significance (p=0.036) between the application used and fertility window predictions, no statistical significance (p=0.648) was found between the application used and ovulation prediction. These results indicated that a woman should not use these apps for ovulation, but should instead use another technique such as LH kits. The best method of ovulation prediction for contraception or conception is to use the apps to calculate periods, non-fertile days, and the fertility window with the LH kits applied during the fertility window to detect physiologic ovulation. If the LH kits are used, the cycle chart listed in the ovulation kits (Table 2 above) generally provided dates early enough to catch the LH surge; however, given the variation in cycle length for some subjects, it proved difficult to remember when to start testing in order to not miss the LH surge.

Interface

Only four of the apps allowed for period lengths greater than 35 days to be entered and/or 10 days of menses even though irregular periods are common and can cause difficulty in being prepared for unexpected menstruation.
This limitation discriminates against the many women who fall outside of the "normal" cycle as cycles can vary greatly. With this variation, it is important for an app to have an adjustable algorithm to accommodate the irregularities. Understanding one's own cycle is empowering as is understanding other health issues. Only apps B and G provided health literature to help answer questions about women's health issues.

Additionally, being able to share this data with a medical provider to cultivate a partnership for improved care is paramount and an aspect that most of the apps lacked. Only app A and B allowed for shareable and downloadable data.

Privacy

Many of the apps ask for additional information and symptoms, such as cramps, mood, and dates of intercourse, but the algorithm did not use this information to recalculate ovulation or menstruation days. With no apparent application of this extra information, the user is left wondering how this data is used. According to a 2019 study published by Privacy International, some menstrual tracking apps share information with social media websites without permission from the user, including days of intercourse to improve targeted advertisements.15 Apps that ask for additional information should inform the user of the utility of the information. For example, App B requires the user to opt-out of

third-party data sharing [ex. Facebook Analytics] by adjusting in your phone's security settings, the Identifier For Advertiser (IDFA), a setting hidden and set to allow tracking as the phone's default setting.

Under guidelines created by the Apple App Store for app developers providing healthcare, these apps must be tested before becoming available for download or purchase. In those guidelines, the company agrees to comply with Apple's user interface, consumer protection, and license agreement; however, no method is utilized to monitor or regulate the quality of the services the apps provide.16 With high demand and the necessity for clear and accurate instructions and data collection, healthcare apps should be tested for the quality of their program as well as compliance with the Apple App Store's other guidelines.

Strengths and Weaknesses

This study had several limiting factors: limited resources, number of participants, and time frame. The research was funded by the investigators which allowed for a small budget; however, the resources used were meant to represent those that the average woman could afford to use. With the limited time frame of three months, distinguishing regular menstrual and ovulation cycles was difficult. One researcher dropped out during the course of the study, demonstrating the likely pattern of noncompliance if this study were to be applied to a larger group. Another limitation is the generalizability of the results as all investigators, highly educated medical students, were the study subjects. Ultrasound could have been used as a more objective measurement of ovulation than using LH kits alone. The age ranges for the researchers (24 - 33 years) represented the fraction of women at peak reproductive age during which the average American woman delivers her first child (26.5 years).14 Due to the low cost of the apps and kits used, this study is easily reproducible with a minimal budget.

The researchers, as highly educated and dedicated individuals using these apps felt the commitment to track periods, symptoms, and tests for ovulation was extensive. For the majority of app and testing kit users, compliance would potentially be low and may not provide a time-worthy investment.

Conclusion

Period tracking apps are an easy, abundant, and low-cost resource that empowers women to take charge of their reproductive health. Evaluation of a small sample of period tracking apps in the Apple App store with three users showed application data collection interfaces, services, and cost variation, with some risk of user information misused for targeted advertisements. Further research with additional subjects is needed. Since the three subjects were university educated and still experienced problems, research should account for each subject's age and education.

The results indicated that a woman should not use these apps for ovulation, but should instead use another technique such as LH kits. The best method of ovulation prediction for contraception or conception is to use the apps to calculate periods, non-fertile days, and the fertility window with the LH kits applied during the fertility window to detect physiologic ovulation. These results are based on women with predictable cycle lengths and these methods may pose to be less reliable for those without "regular" menstrual cycles.

Menstrual cycles continue to be highly variable and difficult to predict, especially in women with irregular periods. Apps that do not account for real-world menstruation and cycle length discriminate against women who fall out of the app's predetermined "normal." Because of this high variability in the menstrual cycle for a woman and amongst women, a study with more subjects, a longer time frame, and a diagnostic technique to confirm ovulation will allow for improved examination of the accuracy of the menstrual tracking apps and ovulation kits.

The study also suggests that few apps provide health literature even though the opportunity is abundant due to high daily viewability which can be capitalized to further educate the user and cultivate a culture of partnership. A partnership that should be based on trust and privacy. Some of these apps allow for further tracking of personal reproductive health, but the safety and privacy of this deeply personal information and how this is being used is unclear.

This study raises questions of whether additional guidelines should be enforced to regulate not only the privacy of information but also the quality of the information provided by the applications. Each app should also clearly state if and to whom the user's information is shared or an easy way to opt-out of sharing private information because every woman has the right to free, private, accurate tools that allow her to take charge of her reproductive health.

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. doi:10.1056/NEJM199512073332301

2. Wilcox AJ, Dunson D, Baird DD. The timing of the "fertile window" in the menstrual cycle: Day specific estimates from a prospective study. Br Med J. 2000;321(7271):1259-1262. doi:10.1136/bmj.321.7271.1259

3. Thijssen A, Meier A, Panis K, Ombelet W. "Fertility Awareness-Based Methods" and subfertility: a systematic review. Facts, views Vis ObGyn. 2014;6(3):113-123. http://www.ncbi.nlm.nih.gov/pubmed/25374654. Accessed April 24, 2020.

4. Stanford, Joseph B. MD, MSPH; White, George L. Jr PhD, MSPH; Hatasaka HM. Timing Intercourse to Achieve Pregnancy: Current Evidence. Obstet Gynecol. 2002;100(6):1333-1341. doi:10.1016/s0029-7844(02)02382-7

5. Finer LB, Zolna MR. Declines in unintended pregnancy in the United States, 2008-2011. N Engl J Med. 2016. doi:10.1056/NEJMsa1506575

6. Hampton K, Mazza D. Should spontaneous or timed intercourse guide couples trying to conceive? Hum Reprod. 2009. doi:10.1093/humrep/dep322

7. Hampton KD, Mazza D, Newton JM. Fertility-awareness knowledge, attitudes, and practices of women seeking fertility assistance. J Adv Nurs. 2013;69(5):1076-1084. doi:10.1111/j.1365-2648.2012.06095.x

8. Frank-Herrmann P, Jacobs C, Jenetzky E, et al. Natural conception rates in subfertile couples following fertility awareness training. Arch Gynecol Obstet. 2017;295(4):1015-1024. doi:10.1007/s00404-017-4294-z

9. Hall J. Female Physiology Before Pregnancy and Female Hormones- ClinicalKey. In: Guyton and Hall Textbook of Medical Physiology. 13th ed. Elsevier Inc; 2016:1037-1054. https://www.clinicalkey.com/#!/content/book/3-s2.0-B9781455770052000822?indexOverride=GLOBAL. Accessed April 15, 2020.

10. Bull JR, Rowland SP, Scherwitzl EB, Scherwitzl R, Danielsson KG, Harper J. Real-world menstrual cycle characteristics of more than 600,000 menstrual cycles. npj Digit Med. 2019;2(1):83. doi:10.1038/s41746-019-0152-7

11. Su H-W, Yi Y-C, Wei T-Y, Chang T-C, Cheng C-M. Detection of ovulation, a review of currently available methods. Bioeng Transl Med. 2017;2(3):238-246. doi:10.1002/btm2.10058

12. Ahrefs - SEO Tools & Resources To Grow Your Search Traffic. https://ahrefs.com/. Accessed February 23, 2020.

13. Guermandi E, Vegetti W, Bianchi MM, Uglietti A, Ragni G, Crosignani P. Reliability of ovulation tests in infertile women. Obstet Gynecol. 2001. doi:10.1016/S0029-7844(00)01083-8

14. Natality, 2016-2018 expanded Results Form. https://wonder.cdc.gov/controller/datarequest/D149;jsessionid=6AAFCC43E03CA2868AA38A36EC16C1A5. Accessed April 8, 2020.

15. No Body's Business But Mine: How Menstruation Apps Are Sharing Your Data | Privacy International. https://privacyinternational.org/long-read/3196/no-bodys-business-mine-how-menstruations-apps-are-sharing-your-data. Accessed April 8, 2020.

16. App Store Review Guidelines - Apple Developer. https://developer.apple.com/app-store/review/guidelines/#terms-conditions. Accessed April 15, 2020.