

Original research article

Perfect-use and typical-use Pearl Index of a contraceptive mobile app[☆]

E. Berglund Scherwitzl^a, O. Lundberg^a, H. Kopp Kallner^{b,c}, K. Gemzell Danielsson^c,
J. Trussell^{d,e}, R. Scherwitzl^{a,*}

^aNaturalCycles Nordic AB, Luntmakargatan 26, S-111 37 Stockholm, Sweden

^bDepartment of Clinical Sciences at Danderyd Hospital, Karolinska Institutet, S-182 88 Stockholm, Sweden

^cDepartment of Women's and Children's Health, Karolinska Institutet and Karolinska University Hospital, S-171 77 Stockholm, Sweden

^dOffice of Population Research, Princeton University, Princeton, NJ 08544, USA

^eThe University of Edinburgh, Edinburgh EH8 9YL, Scotland, UK

Received 21 March 2017; revised 22 August 2017; accepted 24 August 2017

Abstract

Objectives: The Natural Cycles application is a fertility awareness-based contraceptive method that uses dates of menstruation and basal body temperature to inform couples whether protected intercourse is needed to prevent pregnancies. Our purpose with this study is to investigate the contraceptive efficacy of the mobile application by evaluating the perfect- and typical-use Pearl Index.

Study design: In this prospective observational study, 22,785 users of the application logged a total of 18,548 woman-years of data into the application. We used these data to calculate typical- and perfect-use Pearl Indexes, as well as 13-cycle pregnancy rates using life-table analysis.

Results: We found a typical-use Pearl Index of 6.9 pregnancies per 100 woman-years [95% confidence interval (CI): 6.5–7.2], corrected to 6.8 (95% CI: 6.4–7.2) when truncating users after 12 months. We estimated a 13-cycle typical-use failure rate of 8.3% (95% CI: 7.8–8.9). We found that the perfect-use Pearl Index was 1.0 pregnancy per 100 woman-years (95% CI: 0.5–1.5). Finally, we estimated that the rate of pregnancies from cycles where the application erroneously flagged a fertile day as infertile was 0.5 (95% CI: 0.4–0.7) per 100 woman-years. We estimated a discontinuation rate over 12 months of 54%.

Conclusions: This study shows that the efficacy of a contraceptive mobile application is higher than usually reported for traditional fertility awareness-based methods. The application may contribute to reducing the unmet need for contraception.

Implications: The measured typical- and perfect-use efficacies of the mobile application Natural Cycles are important parameters for women considering their contraceptive options as well as for the clinicians advising them. The large available data set in this paper allows for future studies on acceptability, for example, by studying the efficacy for different cohorts and geographic regions.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Fertility awareness; Mobile application; Contraceptive efficacy; Pearl Index

1. Introduction

1.1. Fertility-awareness based methods for contraception

Fertility awareness-based (FAB) methods of contraception are methods based on identifying a woman's approximately six fertile days of each menstrual cycle and a couple adapting their sexual behavior according to this knowledge [1,2]. Various prospective studies have identified the number of fertile days in a cycle and day-by-day conception probabilities [3,4].

The input parameters for the identification and prediction of fertile days in FAB methods are a woman's cycle lengths,

[☆] Conflict of interest: E.B.S. and R.S. are the scientists behind the application Natural Cycles and the founders of the company with stock ownership. O.L. is employed by NaturalCycles Nordic AB. K.G.D. and H.K.K. serve on the medical advisory board of NC and have received honorarium for participating in advisory boards and/or as giving presentations for matters related to contraception and fertility regulation for MSD/Merck, Bayer AG, Gedeon Richter, Exeltis, Actavis, Ferring (K.G.D.), Exelgyn (K.G.D.) and Mithra (K.G.D.). J.T. declares explicitly that there are no conflicts of interest in connection with this article.

* Corresponding author. Tel.: +46 707 17 48 66.

E-mail address: raoul.scherwitzl@naturalcycles.com (R. Scherwitzl).

basal body temperature (BBT), quality of the cervical mucus, urine- or saliva-based fertility markers or combinations of these [5]. The mobile application under study in this paper uses BBT and dates of menstruation as input, with optional input of luteinizing hormone (LH) test results.

FAB methods provide an option for women who do not want to use hormonal contraceptives for medical or personal reasons. Several FAB methods provide comparable contraceptive effectiveness to other modern contraceptive methods [6], though generally require higher dedication from the users [6,7].

The perfect-use failure rates of FAB methods that have been computed correctly vary from 0.4 per 100 woman-years to 4.8% annual pregnancy probability [8–11]. Published typical-use failure rates for FAB methods range from 1.6% per year up to above 20% [10,12].

Several fertility monitoring devices based on FAB methods are available on the market, but their contraceptive efficacies are not known with a few notable exceptions [13–15].

Berglund Scherwitzl et al. [16] estimated the typical-use efficacy of the Natural Cycles mobile application to be 7.0 pregnancies per 100 woman-years and the conception probability over 13 cycles to be 7.5%. These calculations were based on 2053 woman-years of data collected from 4054 women. This was the first study to calculate the contraceptive efficacy of a FAB method supported by a mobile application, but did not fully adhere to the guidelines set out by Trussell and Kost [17]. The method failure rate, that is, a conservative estimate of the rate of pregnancies in cycles where the algorithm erroneously mislabeled a fertile day as infertile, was computed similar to studies of other fertility monitoring devices [18]. However, the study lacked a correctly computed perfect-use failure rate, included retrospective elements and did not have a 1-year follow-up. In this prospective study, we employ a larger data set and longer follow-up times to calculate the perfect-use efficacy of the application, as well as to re-investigate the typical-use efficacy and the method failure rate.

2. Materials and methods

2.1. The mobile application

Users of the Natural Cycles mobile application log menstrual cycle dates and BBT data into a device such as a smartphone, tablet or computer. The underlying technology is a statistical algorithm that uses a color code to signal to the user whether a given day is likely to be in the fertile window [16,19]. The predictions improve over time as the algorithm incorporates more user data into its calculations.

The algorithm can identify the ovulation day with high precision and concomitantly has very low probability of falsely attributing a fertile day as infertile [19].

In addition to menstruation and BBT data, users are encouraged to take a urine LH test on days close to predicted ovulation, as well as to take a pregnancy test if the data indicate a possible pregnancy. The application detects pregnancy by looking for a combination of (i) delayed menses and (ii) a

longer period of elevated temperature levels consistent with the implantation of a fertilized egg in the uterus [20]. Users may also log sexual activity. On a given day, they can log protected, unprotected or no intercourse. This information was used as input for calculation of the perfect-use failure rate. Furthermore, users were asked in-app “What form of contraception do you use on red days?”. Answering this question was voluntary and complemented the intercourse information logged by the user.

2.2. Study design

This prospective observational analysis included all women who registered as paying users of Natural Cycles between August 1 2014 and August 1 2016 with the intent of preventing pregnancies. We included no data from menstrual cycles starting after January 31, 2017 in the study. We followed users until they had reported a positive pregnancy test result, answered a follow-up e-mail or reported menstruation in the application, up until a final cutoff at March 9, 2017.

Recruitment used end-consumer marketing techniques. All users included in the study agreed at registration to share data anonymously for research and were free to withdraw their consent at any time through their profile settings in the application. Each participant had to log at least 20 days of data in total (such a daily log can contain any combination of menstruation, BBT, LH test, pregnancy test result, sexual activity and personal notes) to be included in the data analysis. We imposed this lower limit mainly to exclude not only women who registered with no real intention of using the application but also users who were already pregnant when registering.

We detected pregnancies primarily through the users logging a positive pregnancy test in the application. We attempted to follow up with women who discontinued the method through in-app messages and via e-mail. We then classified all users lost to follow-up in one of three categories, based on the BBT at discontinuation, and at what time in the menstrual cycle they discontinued:

1. We considered users *likely* to be pregnant if they discontinued at a late day of their luteal phase and/or reported high BBT when discontinuing after ovulation. We treated these women as pregnant for all calculations unless they stated explicitly in follow-up that they were not.
2. We considered users *very unlikely* to be pregnant if they discontinued at a point in their cycle when it was very improbable that they were pregnant (e.g., between menstruation and ovulation). We do not find it accurate to denote a possible later pregnancy of such a user as due to contraceptive failure. We treated these women as not pregnant for all calculations unless they stated explicitly in follow-up that they were.
3. We considered users *possibly* pregnant if we could not easily place them in either of the above categories (e.g., a user who quit early in the luteal phase with BBT

incompatible with their follicular phase). We treated such users as having unknown pregnancy status. We treated these users as not pregnant in the main efficacy analysis. However, we conducted a sensitivity analysis in which all possibly pregnant women were treated as pregnant, as a worst-case scenario.

We censored the exposure time of pregnant women at the time of the reported pregnancy test.

We removed women from the study who stated in follow-up that their pregnancy was planned despite using the application in the prevention mode. However, we additionally performed a sensitivity analysis by including these women and counting their planned pregnancies as contraceptive failures.

The study protocol was reviewed and approved by the regional ethics committee (EPN, Stockholm, diary number 2017/563-31).

2.3. Data analysis

We estimated efficacies using Kaplan–Meier life-table analysis to find 13-cycle typical-use pregnancy rates [21], as well as by calculating the Pearl Index for typical and perfect usage [22]. We calculated confidence intervals (CIs) for life-tables as suggested by Kalbfleisch and Prentice [23]. Calculating a perfect-use efficacy of a FAB method requires knowledge of when the women had intercourse, in order to distinguish between perfect and imperfect usage.

To achieve a definition of perfect use that provides enough statistical precision while making minimal assumptions, we used an approach similar to Frank-Herrmann et al. [10]. We treated certain modes of sexual behavior as constituting perfect use, and then related all pregnancies occurring during perfect use to all perfect-use cycles. A perfect-use cycle in this study is a cycle in which the user was sexually active but had no unprotected intercourse on days indicated as fertile. The calculation of perfect-use efficacy thus requires detailed intercourse information for such days. As intercourse logging was voluntary, the information needed was available for only a subset of cycles. In estimating the perfect-use efficacy we counted cycles in which a user logged either *protected intercourse* or *no intercourse* on any of the days indicated as fertile, while not logging *unprotected intercourse* on these fertile days.

Besides calculating a Pearl Index, we followed the method of Trussell and Grummer-Strawn [11] in calculating the 13-cycle probability of contraceptive failure as $1-(1-p)^{13}$, where p is the per-cycle probability of failure for a user, calculated using all perfect-use cycles and pregnancies. For the calculation of p , each woman was allowed to contribute a maximum of 13 cycles.

Since a woman may register at any time during her cycle, we calculated her exposure from registration to censoring. We added up these individual contributions to calculate the typical-use Pearl Index using a woman-year consisting of 365 days. We additionally calculated a 1-year Pearl Index

for which each woman was allowed to contribute a maximum exposure of 365 days.

Finally, we calculated how often a pregnancy occurred in a cycle where the application failed by erroneously indicating a fertile day as infertile. This estimate is what was labeled as the perfect-use failure rate in the previous study and is here instead labeled as the method failure rate [15]. We calculate this as a Pearl Index after first removing pregnancies where the user logged unprotected intercourse on a correctly indicated fertile day closer to ovulation.

3. Results

The study comprised 22,785 women fulfilling the inclusion criteria, who on average contributed 9.8 months of data, yielding a total of 18,548 women-years of exposure. Four thousand one hundred eighty-two additional women had registered without logging the 20 data points required for inclusion—778 of these registered without logging any data at all. Table 1 lists the exposure times of the women studied. Six thousand nine hundred forty-four (30%) of them contributed more than a year of data.

The average age of the users was 29.2 years (1 standard deviation=5.0 years). The women registered as paying users of the application in 37 different countries, with a large majority (79%) from Sweden. We registered a total of 1273 pregnancies. These numbers are all reported after we first removed the exposure and pregnancies of 25 women who in our follow-up reported that their pregnancy was planned. Out of the 1273 pregnancies, 62 were registered in the first cycle of usage. Fig. 1 shows a flowchart describing how these pregnancies were determined. Table 2 shows the total number of cycles, woman-years and pregnancies during perfect use and typical use. In 32% of all menstrual cycles, the users registered some type of intercourse information (protected, unprotected or no intercourse), and 9.6% of the cycles were considered perfect-use cycles. In 603 (47%) of all cycles with pregnancies, no intercourse was reported.

Using the data in Table 2, we calculated a perfect-use Pearl Index of 1.0 (95% CI: 0.5–1.5) pregnancies per 100 woman-years and a 13-cycle pregnancy probability of 1.0% (95% CI: 0.5–1.5). Fig. 2 shows the Kaplan–Meier

Table 1
Included women and exposure times

	Number of women
Registered	26,967
Included in the study	22,785
Contributed >3 months	19,534
Contributed >6 months	15,224
Contributed >12 months	6944
Contributed >18 months	2684

Women included in the study logged at least 20 data points. The times “contributed” are between the woman’s start date and the censoring due to being lost to follow-up, pregnant or reaching the end of the study.

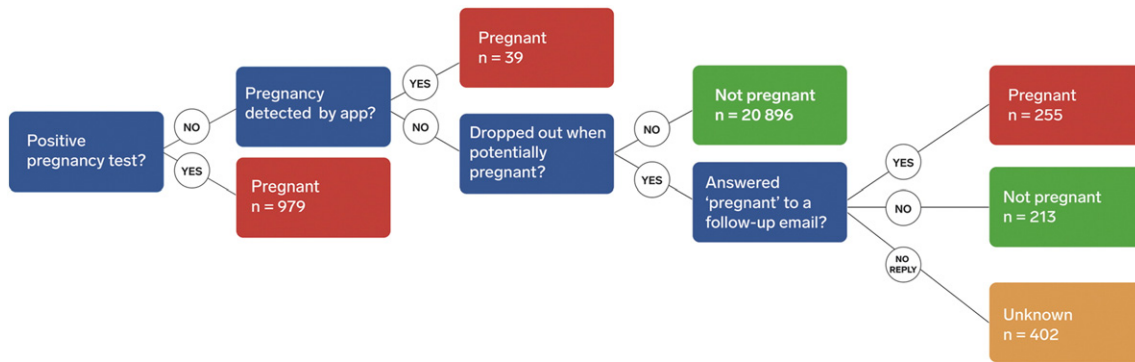


Fig. 1. Flowchart describing how pregnancies were detected in the study. Each box showing an “end-state” also contains the number *n* of women who are found in this state. Red and green boxes contain the women who are considered pregnant and not pregnant in all calculations of the study. The yellow “unknown” box is considered not pregnant for the main calculations but pregnant for the worst-case scenario. There were 25 additional pregnancies detected in any of the three channels illustrated in the figure. These pregnancies, as well as the full exposure of the pregnant women, were removed from the study due to the users’ follow-up statement that the pregnancy was planned.

cumulated probability of nonpregnancy as a function of the number of cycles for typical use. The 13-cycle typical-use failure rate was 8.3% (95% CI: 7.8–8.9).

The typical-use Pearl Index was 6.9 pregnancies per 100 woman-years (95% CI: 6.5–7.2). Truncating each user at a maximum of 365 days yielded a Pearl Index of 6.8 (95% CI: 6.4–7.2). The difference between the 1-year Pearl Index and the life-table figure stems from the fact that many women, especially those who previously used hormonal contraceptives, have longer-than-average early cycles.

At the end of follow-up, we considered 402 women possibly pregnant but with pregnancy status unknown. In the worst-case scenario where all these women were pregnant, the typical-use Pearl Index was 9.0 if all exposure times were taken into account and 9.3 if users were censored after contributing 13 cycles.

The method failure rate was 0.5 pregnancies (95% CI: 0.4–0.7) per 100 woman-years.

The discontinuation rate over 12 months was 54%. The discontinuation rate per month was constant over the year, with the exception of cycle 0 in which very few of the selected women discontinued (due to the 20 data points requirement).

A sensitivity analysis including the 25 women removed due to having reported the pregnancy as planned showed that the typical-use Pearl Index increased from 6.9 to 7.0, the life-table pregnancy probability increased from 8.3 to 8.5

and the method failure rate increased from 0.5 to 0.6. The perfect-use failure rate decreased very slightly, from 1.0253 to 1.02232, well within the rounding accuracy.

4. Discussion

We found that the Pearl Index including all cycles is 6.9 pregnancies per 100 woman-years and that the 1-year Pearl Index is 6.8 per 100 woman-years, while the pregnancy rate at typical use is 8.3% per year for the first 13 cycles. All these results are consistent with the 2016 study [16], but the precision is higher in this study. While Pearl Indexes during typical use could have been expected to be significantly lower in this study than in the previous one due to the longer exposure time, it is possible that the larger user base with a smaller fraction of more dedicated users (e.g., “early-adopters”) cancels this effect.

Of the 1273 pregnant women, 259 actively changed the mode of operation to “plan a pregnancy” within the application on the same date as they reported the pregnancy. While this change is in part due to the inherent software design, it is likely that many of these women planned a pregnancy without changing their stated intent prior to their pregnancy. In order to be as conservative as possible, we considered all these pregnancies as failures during typical use. As a result, the typical-use Pearl Index is almost certainly biased upward. In the future, women who suddenly start logging a lot of unprotected intercourse around ovulation will be asked to either switch to the “plan a pregnancy” mode or actively state that they still do not plan a pregnancy.

The perfect-use Pearl Index was 1.0 pregnancy per 100 woman-years, and the 13-cycle pregnancy probability was 1.0%. These figures make it possible to compare the efficacy of Natural Cycles to other means of contraception. The definition of perfect use of the application in this study matches what the users of the application are instructed to do,

Table 2

Number of menstrual cycles and pregnancies during perfect use and typical use and the number of method failures, and the Pearl Index calculated from these numbers

	Pearl Index	Pregnancies	Woman-years	Cycles
Perfect use	1.0	17	1661 ^a	21,597
Typical use	6.9	1273	18,548 ^b	224,563
Method failure	0.5	102	18,548 ^b	224,563

^a Calculated using the number of registered cycles.

^b Calculated using the number of days since starting to use the application.

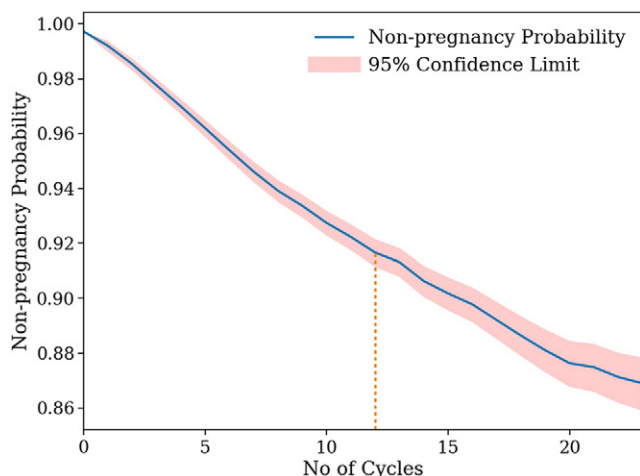


Fig. 2. The life-table nonpregnancy probability as a function of the number of cycles. The 95% CI calculated as suggested by Kalbfleisch and Prentice [23] is indicated (pink band). The dotted orange line indicates the 13th cycle.

that is, to abstain or practice protected intercourse on red days.

Other studies have calculated comparable perfect-use efficacies of FAB methods when complemented by other methods on fertile days. Frank-Herrmann et al. [10] calculated the efficacy of a symptothermal method complemented by condom to be 0.6, and Arévalo et al. [9] calculated the efficacy of the TwoDay method to be 6.3 when complemented by condom or withdrawal.

As we lacked sufficient knowledge of our users' behavior over time, we could not calculate life-table results for perfect use. The definition of perfect use we use in this study is likely to yield an estimate that is biased upward. As we only included cycles where we are confident users did not have unprotected intercourse on a fertile day, we disregard cycles where users in fact abstained but did not log this information. We have no other way in practice to estimate efficacy during perfect use.

The method failure rate remains consistent with the 2016 study [16]. This quantity, which is based on an exact measurement of how often the method fails to protect the user, cannot easily be compared to other contraceptives. We nevertheless believe that it is of interest, as it makes it possible to compare the performance of the Natural Cycles algorithm to other algorithm-based FAB methods.

The 54% discontinuation rate is consistent with what has been reported elsewhere for FAB methods [24]. Future studies will scrutinize this number in greater detail, as well as investigate which women are most likely to discontinue.

This study is one of the largest prospective studies ever performed on a FAB method, leading to very high precision in the estimates. The real-life nature, free from any possible bias from contact with clinics and health care providers, allows us to investigate the typical use of the application in a more direct sense than many other studies of contraceptive effectiveness. On the other hand, the real-life nature means

that it is less clear who is to be considered pregnant, and any such estimate must rely to some extent on credible assumptions about the pregnancy status. In this study, we have complemented the estimate of efficacy during typical use with an estimate based on the worst-case assumption that everyone who was lost to follow-up while possibly pregnant was in fact pregnant. These two estimates indicate the possible span of typical-use failure rates.

The low thresholds for entering and exiting the study contribute to a relatively high dropout rate. It is possible that because of the similarity between using FAB methods for prevention and for planning a pregnancy, several of the pregnant users were actually planning their pregnancies but never changed to the pregnancy-planning mode in the application.

The lack of information on intercourse from most users is a challenge when calculating perfect-use efficacy of the application. Future investigators of similar data need to be aware of this issue.

Funding

The study was funded by NaturalCycles Nordic AB. Partial support was provided from an infrastructure grant for population research from the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health P2C HD047879 (J.T.).

References

- [1] Jennings V, Burke AE. Fertility awareness-based methods. In: Hatcher RA, Trussell J, Nelson AL, Cates W, Kowal D, & Policar MS, editors. *Contraceptive technology*. Atlanta, GA: Ardent Media Inc.; 2011. p. 417–34.
- [2] 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. *Med* 1995;333: 1517–21.
- [3] Dunson D, Baird DD, Wilcox AJ, Weinberg CR. Day-specific probabilities of clinical pregnancy based on two studies with imperfect measures of ovulation. *Hum Reprod* 1999;14:1835–9.
- [4] Wilcox AJ, Dunson D, Baird DD. The timing of the “fertile window” in the menstrual cycle: day specific estimates from a prospective study. *BMJ* 2000;321:1259–62.
- [5] Freundl G, Sivin I, Batár I. State-of-the-art of non-hormonal methods of contraception: IV. Natural family planning. *Contracept Reprod* 2010;15:113–23.
- [6] Hatcher RA, Trussell J, Nelson AL, Cates W, Kowal D, Policar MS. *Contraceptive technology*. 20th ed. Atlanta, GA: Ardent Media, Inc.; 2011.
- [7] Trussell J. Contraceptive failure in the United States. *Contraception* 2011;83:397–404.
- [8] Arévalo M, Jennings V, Sinai I. Efficacy of a new method of family planning: the standard days method. *Contraception* 2001;65:333–8.
- [9] Arévalo M, Jennings V, Nikula M, Sinai I. Efficacy of the new TwoDay method of family planning. *Hum Reprod* 2004;82:885–92.
- [10] Frank-Herrmann P, Heil J, Gnoth C, et al. The effectiveness of a fertility awareness based method to avoid pregnancy in relation to a

- couple's sexual behaviour during the fertile time: a prospective longitudinal study. *Hum Reprod* 2007;22:1310–9.
- [11] Trussell J, Grummer-Strawn L. Contraceptive failure of the ovulation method of periodic abstinence. *Fam Plan Perspect* 1990;22:65–75.
- [12] Kost K, Singh S, Vaughan B, Trussell J, Bankole A. Estimates of contraceptive failure from the 2002 National Survey of Family Growth. *Contraception* 2008;77:10–21.
- [13] Freundl G, Godehardt E, Kern PA, Frank-Herrmann P, Koubenec HJ, Gnath C. Estimated maximum failure rates of cycle monitors using daily conception probabilities in the menstrual cycle. *Hum Reprod* 2003;18:2628–33.
- [14] Freundl G, Frank-Herrmann P, Godehardt E, Klemm R, Bachhofer M. Retrospective clinical trial of contraceptive effectiveness of the electronic fertility indicator Ladycomp/Babycomp. *Adv Contracept* 1998;14:97–108.
- [15] Drouin J, Guilberg EE, Désaulniers G. Contraceptive application of the Bioself fertility indicator. *Contraception* 1994;50:229–38.
- [16] Berglund Scherwitzl E, Gemzell Danielsson K, Sellberg JA, Scherwitzl R. Fertility awareness-based mobile application for contraception. *Contracept Reprod* 2016;21:234–41.
- [17] Trussell J, Kost K. Contraceptive failure in the United States: a critical review of the literature. *Stud Fam Plan* 1987;18:237–83.
- [18] Bonnar J, Flynn A, Freundl G, Kirkman R, Royston P, Snowden R. Personal hormone monitoring for contraception. *Fam Plann* 1999;24:128–34.
- [19] Berglund Scherwitzl E, Lindén Hirschberg A, Scherwitzl R. Identification and prediction of the fertile window using Natural-Cycles. *Contracept Reprod* 2015;20:403–8.
- [20] Baird DD, Weinberg CR, McConaughy DR, Wilcox AJ. Rescue of the corpus luteum in human pregnancy. *Biol Reprod* 2003;68:448–56.
- [21] Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;53:457–81.
- [22] Pearl R. Factors in human fertility and their statistical evaluation. *Lancet* 1933;222:607–11.
- [23] Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. New York: Wiley; 1980.
- [24] Vaughan B, Trussell J, Kost K, Singh S, Jones R. Discontinuation and resumption of contraceptive use: results from the 2002 National Survey of Family Growth. *Contraception* 2008;78:271–83.