

Original Paper

# Relationship Between the Menstrual Cycle and Timing of Ovulation Revealed by New Protocols: Analysis of Data from a Self-Tracking Health App

Satoshi Sohda<sup>1</sup>, MD, PhD; Kenta Suzuki<sup>2,3</sup>, PhD; Ichiro Igari<sup>3</sup>, PhD

<sup>1</sup>Department of Obstetrics and Gynaecology, Graduate School of Medical Science, University of Tsukuba, Tsukuba, Japan

<sup>2</sup>Biodiversity Conservation Planning Section, Center for Environmental Biology and Ecosystem Studies, National Institute for Environmental Studies, Tsukuba, Japan

<sup>3</sup>Advanced Technology Section, R&D Division, MTI Ltd, Shinjyuku, Japan

**Corresponding Author:**

Kenta Suzuki, PhD

Biodiversity Conservation Planning Section

Center for Environmental Biology and Ecosystem Studies

National Institute for Environmental Studies

EEF, 3rd Floor

16-2 Onogawa

Tsukuba, 305-8506

Japan

Phone: 81 029 850 2747

Fax: 81 029 850 2747

Email: [kenta11514201@gmail.com](mailto:kenta11514201@gmail.com)

## Abstract

**Background:** There are many mobile phone apps aimed at helping women map their ovulation and menstrual cycles and facilitating successful conception (or avoiding pregnancy). These apps usually ask users to input various biological features and have accumulated the menstrual cycle data of a vast number of women.

**Objective:** The purpose of our study was to clarify how the data obtained from a self-tracking health app for female mobile phone users can be used to improve the accuracy of prediction of the date of next ovulation.

**Methods:** Using the data of 7043 women who had reliable menstrual and ovulation records out of 8,000,000 users of a mobile phone app of a health care service, we analyzed the relationship between the menstrual cycle length, follicular phase length, and luteal phase length. Then we fitted a linear function to the relationship between the length of the menstrual cycle and timing of ovulation and compared it with the existing calendar-based methods.

**Results:** The correlation between the length of the menstrual cycle and the length of the follicular phase was stronger than the correlation between the length of the menstrual cycle and the length of the luteal phase, and there was a positive correlation between the lengths of past and future menstrual cycles. A strong positive correlation was also found between the mean length of past cycles and the length of the follicular phase. The correlation between the mean cycle length and the luteal phase length was also statistically significant. In most of the subjects, our method (ie, the calendar-based method based on the optimized function) outperformed the Ogino method of predicting the next ovulation date. Our method also outperformed the ovulation date prediction method that assumes the middle day of a mean menstrual cycle as the date of the next ovulation.

**Conclusions:** The large number of subjects allowed us to capture the relationships between the lengths of the menstrual cycle, follicular phase, and luteal phase in more detail than previous studies. We then demonstrated how the present calendar methods could be improved by the better grouping of women. This study suggested that even without integrating various biological metrics, the dataset collected by a self-tracking app can be used to develop formulas that predict the ovulation day when the data are aggregated. Because the method that we developed requires data only on the first day of menstruation, it would be the best option for couples during the early stages of their attempt to have a baby or for those who want to avoid the cost associated with other methods. Moreover, the result will be the baseline for more advanced methods that integrate other biological metrics.

(*J Med Internet Res* 2017;19(11):e391) doi: [10.2196/jmir.7468](https://doi.org/10.2196/jmir.7468)

**KEYWORDS**

self-tracking; person generated health data; calendar calculation; fertility; menstrual cycle

## Introduction

Awareness of one's own fertility is considered important in helping women become pregnant in a shorter period of time [1-3]. Fertility depends on the menstrual (ovarian) cycle, and in each cycle there is a "fertile window" during which women can conceive [1,4,5]. Studies on standard menstrual cycles suggest that the fertile window starts 5 days prior to ovulation and ends on the day of ovulation [4]. This is essentially consistent with the results of statistical studies [1,5] that estimated the day-specific pregnancy rates; the pregnancy rate starts to increase 8 days prior to ovulation, peaks after 6 days (2 days prior to ovulation), and reaches almost 0 at 2-3 days after ovulation. Hence, in order to be aware of the fertility window, it is important for a woman to be able to predict the next ovulation date in the course of her menstrual cycles.

The "calendar method" of predicting the next ovulation date was developed based on the recognition of cycles in the menstrual period and fertility, in which women record their menstrual cycles for family planning [6]. It may have been one of the most widely recorded personal health information details before the spread of mobile phones and computers. On the other hand, the recent popularity of self-tracking tools realized by ubiquitous and wearable technologies has led people to gather various kinds of self-information ranging from financial behaviors to physical activities [7,8]. These technologies are used to "help people collect personally relevant information for the purpose of self-reflection and gaining self-knowledge" and are referred to as personal informatics systems [9]. Currently, the classical calendar method of predicting the next ovulation date is integrated into personal informatics systems. There are many mobile phone apps aimed at helping women map their ovulation and menstrual cycles and facilitating successful conception (or avoiding pregnancy) [10,11]. Apps available for these purposes include Ovia Fertility (Ovuline), Glow, OvuView, Ovulation Calendar, Fertility Calendar, My Days, Period Diary, Period Tracker, Maybe Baby, and Fertility Friend [12]. As Lupton [12] mentioned, while these apps are grounded on traditional gynecological knowledge, the advantage that some of them claim is a data analytic approach that can provide greater accuracy than more traditional forms of self-tracking. These apps usually ask users to input various biological features (eg, ovulation, sexual intercourse, basal body temperature, state of cervical mucus, body weight, and the timing of menstrual bleeding).

However, it is not known how to process these features numerically to improve the ovulation prediction error. A mixture of knowledge on biological mechanisms and a statistical approach using the newly enabled biological metrics is promising [13,14], although it is still an open problem. Instead of pursuing the usage of various biological metrics, we believe that more attention should be paid to other aspects of this dataset, that is, its massiveness. It is a remarkable achievement that mobile phone apps have been able to accumulate menstrual cycle data of a vast number of women. Currently, records of

more than 10,000 individuals can be a target of statistical analysis. In this study, we start from calendar-based methods that require only the recording of menstruation to predict the ovulation date [15-17]. For many couples, the calendar-based methods are the simplest options of determining the timing of the menstrual cycle [2]. Even within the simple prediction framework, a large amount of data potentially allows us to figure out individual differences better than in traditional understanding [18,19].

This study aims to clarify the above points using data obtained from a commercial women's health care service provided as a mobile phone app in Japan. We extracted approximately 0.1% of all users of the app, resulting in 7043 subjects after data screening. This paper reports on the progressive health data ecosystem in which commercial health care mobile apps generate massive amounts of data. The results of the data analysis give feedback to the app and can be used to improve public health as well.

## Methods

### Data

We screened 150,000 users who wanted to conceive out of a total of 8,000,000 users of a mobile phone app from a commercial women's health care service, Luna Luna. We used the data of 7043 women who had each recorded at least one menstrual cycle with ovulation date, suggesting that about 5% of women utilize ovulation tests to support their conception. The ovulation date had been determined by one of the methods described in the next paragraph. The total number of cycles was 135,666, and there were 12,731 cycles with an ovulation date. Any cycles in the record that were less than 20 days or greater than 45 days were removed to rule out unnatural cycle length that is due to erroneous or defective input. About 57% (7285) of the cycles had more than 8 records of past menstrual cycles after the screening, which allowed us to analyze the relationship between past and future menstrual cycles. The age distribution of the 7043 women ranged from 20-45 years with a mean of 32.94 years (95% CI 32.04-33.85), which is slightly higher than the mean maternal age of Japanese women at the time of first birth (which was 30.1 years in 2010).

Each user's personal records consisted of the dates when they recognized menstrual bleeding (onset of menstruation) and the dates when they detected ovulation. In the log file, an identifier is attached to each ovulation record to distinguish its basis (clinical diagnosis/ovulation test kit/other reliable method). In our analysis, only clinical diagnosis based (31%) and ovulation test kit based (54%) ovulation records were used. Luna Luna does not ask women to record which clinical diagnostic test they used to determine the ovulation date. However, it is noted that the ovulation day in Japan is commonly determined by ultrasound scanning and occasionally with testing of blood luteinizing hormone or estrogen level.

The Luna Luna data are the property of MTI Ltd (Shinjyuku, Tokyo, Japan). The authors (KS and II) are employees of MTI Ltd and are permitted to access the Luna Luna data server. The information security committee of MTI Ltd concluded that this study does not require approval by an ethics committee because the data are anonymized appropriately; the data server used in the study is a backup of the original data server, on which anonymous IDs are placed on personally identifiable information. Hence, it was impossible for the authors to access personally identifiable information, which was controlled separately. Consent for data use and information acquisition was obtained from Luna Luna users, as stipulated in the terms of use.

## Luna Luna

Luna Luna is a total health care service for female mobile phone users in Japan. Luna Luna offers its users predictions of menstrual cycles, fertility, ovulation, and related health care information, based on user-inputted personal records that are sent and stored in its data server. The data are securely stored separately from personally identifiable information. Luna Luna has been provided as a commercial service for more than 10 years since the year 2000. Luna Luna has 7 million subscribers as of 2016 and occupies a leading position among mobile health care services for female users in the Japanese market.

## Menstrual Cycles, Timing of Ovulation, Follicular Phases, and Luteal Phases

We express the records of the first day of menstruation of woman  $i$  as,

$$M_i = (m_{i1}, m_{i2}, \dots, m_{iT}),$$

where  $m_{i1}$  is the first day of the most recent menstruation of woman  $i$  ( $i=1, 2, \dots, N$ ),  $m_{i2}$  is the first day of her second most recent menstruation, and so on. Then, we defined  $C_i$  as a series of menstrual cycle lengths of woman  $i$  by,

$$C_i = (c_{i1}, \dots, c_{iT-1}) = (D(m_{i1}, m_{i2}), \dots, D(m_{iT-1}, m_{iT})),$$

where  $c_{ij}$  denotes the  $j$ th most recent menstrual cycle of woman  $i$  and  $D(m_{ij}, m_{ij+1})$  denotes a function that gives the number of elapsed days between  $m_{ij+1}$  and  $m_{ij}$ . For simplicity of notation, we define the mean length of the menstrual cycles of woman  $i$  over the  $j$ th to  $j'$ th cycles as  $c_i^*(j, j') = \sum_{t=j, \dots, j'} c_{it} / (j' - j + 1)$ .

$F_i$  is the series of follicular phase lengths of woman  $i$ , where each follicular phase length,  $f_{ij}$ , is defined as follows: assuming we have records of ovulation  $o_{ij}$  ( $j=1, 2, \dots$ ) between  $m_{ij}$  and  $m_{ij+1}$ ,

$$F_i = (\dots, f_{ij}, \dots) = (\dots, D(o_{ij}, m_{ij+1}), \dots).$$

Thus, the timing of ovulation is  $f_{ij}$  days after the day of the previous menstruation  $m_{ij+1}$ . Similarly, the series of luteal phase lengths of woman  $i$ ,  $L_i$ , where each luteal phase length is  $l_{ij}$ , is defined as,

$$L_i = (\dots, l_{ij}, \dots) = (\dots, D(m_{ij} - 1, o_{ij} + 1), \dots).$$

Here,  $o_{ij} + 1$  is the day after ovulation and  $m_{ij} - 1$  is the day before the next menstruation. The timing of ovulation is  $c_{ij} - l_{ij} - 1$  days

after the first day of the previous menstruation  $m_{ij+1}$ . It should be mentioned that the length of records varied among the women. We used  $T_i$  to indicate the length of records of woman  $i$ .

We investigated the relationships between the length of menstrual cycles and the length of the follicular phases or that of the luteal phases. We also analyzed the relationships between the mean length of past menstrual cycles and the length of the follicular phases or that of the luteal phases because prediction of ovulation date requires an unknown length of the next menstrual cycle.

## Calendar Calculations

We evaluated the relevance of three calendar-based methods using our data. The first was the Ogino method [15], which assumes a fixed length of the luteal phase of 14 days and predicts the ovulation date as  $[c^*] - 15$  days after the onset of the previous menstruation for a particular woman when the mean length of her menstrual cycles is  $c^*$  (here  $[c^*]$  represents  $c^*$  rounded down to the nearest integer). This implies that the length of the follicular phase in the next cycle is  $[c^*] - 15$ . The Ogino method is the most widespread calendar-based method relied on by Japanese women. The second method is the method proposed by Lamprecht and Grummer-Strawn [16], which assumes that the length of the next follicular phase is  $[c^*/2]$ , and thus this method predicts a woman's next ovulation date as  $[c^*/2]$  days after the onset of the previous menstruation. Here, we call this method as the half cycle length (HCL) method. Because of its simplicity, we chose the HCL method over other calendar-based methods that reflect individual differences in the length of the luteal phase in a menstrual cycle. The third method is the method that was developed in this study and predicts a woman's next ovulation date as  $f_i(c^*)$  days after the onset of the previous menstruation. As explained in the next section,  $f_i$  is a linear function that is optimized by using the relationship between the follicular phase length and  $c^*$ . Hence, we call the third method as the Optimized method.

To analyze the relationship between the timing of ovulation and the mean length of past menstrual cycles, we evaluated the three prediction models for follicular phase length, that is,  $\mu_{\text{Ogino}}$ ,  $\mu_{\text{HCL}}$ , and  $\mu_{\text{OPT}}$ , which predict the length of the next follicular phase as  $\mu_{\text{Ogino}}(C_i, j, k) = [c_{ij}^*(k)] - 15$ ,  $\mu_{\text{HCL}}(C_i, j, k) = [c_{ij}^*(k)/2]$ , and  $\mu_{\text{OPT}}(C_i, j, k) = [f_{ij}(c_{ij}^*(k))]$ , respectively. Here,  $c_{ij}^*(k)$  is defined as  $c_{ij}^*(k) = c_i^*(j+1, J+k)$ . It should be noted that if  $k=L$ ,  $c_{ij}^*(k)$  is identical to the nearest cycle length,  $c_{ij+1}$ .

To predict the timing of ovulation, we used the results of least square fitting between mean cycle length and follicular phase length. The prediction performance of the timing of ovulation using the obtained model,  $\mu_{\text{OPT}}$ , and the prediction performances using the Ogino and HCL methods,  $\mu_{\text{Ogino}}$  and  $\mu_{\text{HCL}}$ , respectively, were compared with different allowable prediction error levels,  $|\mu - f_{ij}| \leq 0, 1, 2$ .

## Linear Models

In our analysis, we used a linear model to describe the relationship between an explanatory variable  $x$  and a response

variable  $y$ . For data that consist of multiple data points from each individual, linear models are generally categorized into two types: fixed effect models and random effect models [20]. A fixed effect model is formalized as follows:

$$y_{ij} = \alpha_1 \delta_{i1} + \dots + \alpha_n \delta_{in} + x_{ij} \beta + \mu_{ij}$$

$$\delta_{ij} = 1 \text{ if } i=j \text{ else } 0,$$

where  $\delta_{ij}$  s are dummy variables. The least square estimate of parameters including dummy variables is obtained as,

$$\beta_{lsdv} = \sum_{i=1, \dots, N} \sum_{t=1, \dots, T} (x_{ij} - x_i^*) (x_{ij} - x_i^*)' / \sum_{i=1, \dots, N} \sum_{t=1, \dots, T} (x_{ij} - x_i^*) (x_{ij} - x_i^*)'$$

$$\alpha_i = y_i^* - x_i^* \beta_{lsdv}$$

where  $x_i^* = 1/T_i \sum_{t=1, \dots, T} x_{it}$  and  $y_i^* = 1/T_i \sum_{t=1, \dots, T} y_{it}$ . Hence, in this model, different women have different  $\alpha_i$  for the relationship between  $x$  and  $y$ , and there is a correlation between  $\alpha$  and  $x$ . On the other hand, in random effect models, the random component of  $\alpha_i$  s is included in the random variable  $\mu_{ij}$  as follows:

$$y_{ij} = \alpha + x_{ij} \beta + \mu_{ij} = x_{ij}' \beta' + \mu_{ij}$$

Here, we assumed that  $x_{ij} = (1, x_{ij})$  and  $\beta' = (\alpha, \beta)$ . A pooled ordinary least square estimate,  $\beta_p' = (\alpha_p, \beta_p)$ , is obtained as,

$$\beta_p' = \sum_{i=1, \dots, N} \sum_{t=1, \dots, T} x_{it}' y_{it} / \sum_{i=1, \dots, N} \sum_{t=1, \dots, T} x_{it}' x_{it}$$

However, the result underestimates the covariance structure in  $\mu_{ij}$  s, which is described as a matrix,

$$\Omega = \{\omega_{ij}\}_{i, j=1, \dots, N}$$

$$\omega_{ij} = \xi_{\alpha}^2 + \xi^2 \text{ if } i=j \text{ else } \xi_{\alpha}^2$$

Using the result of pooled ordinary least square, the generalized least square estimate of parameters that include the effect of covariance,  $\beta_{gls}' = (\alpha_{gls}, \beta_{gls})$ , is obtained as,

$$\beta_{gls}' = \sum_{i=1, \dots, N} \sum_{t=1, \dots, T} \sum_{s=1, \dots, T} \omega_{ts}^{(-1)} x_{it}' y_{it} / \sum_{i=1, \dots, N} \sum_{t=1, \dots, T} \sum_{s=1, \dots, T} \omega_{ts}^{(-1)} x_{it}' x_{is}$$

where  $\omega_{ts}^{(-1)}$  is an element of  $\Omega^{(-1)}$ , which is the inverse matrix of  $\Omega$ . To obtain  $\xi_{\alpha}$  and  $\xi$ , we first calculate,

$$\xi_{\mu} = 1/(NT-p) \sum_{i=1, \dots, N} \sum_{t=1, \dots, T} \mu_{it}^2$$

and

$$\xi_{\alpha} = 1/\{NT(T-1)/2-p\} \sum_{i=1, \dots, N} \sum_{t=1, \dots, T-1} \sum_{s=t+1, \dots, T} \mu_{it} \mu_{is}$$

then,

$$\xi = \xi_{\mu} - \xi_{\alpha}$$

In this paper, we used the Hausman test to determine which of the models better explains the data. In the Hausman test, the percentile value of  $H = (\beta_{gls} - \beta_{lsdv})^2$  in a  $X^2$  distribution with one degree of freedom is calculated, and a random effect model is rejected if the  $P$  value is greater than .05.

## Results

In our dataset, the mean (95% confidence interval) of the menstrual cycle length, the follicular phase length, and the luteal phase length of the 7043 women over all cycles was 29.76 (24-38), 14.84 (10-23), and 13.91 (10-19) days, respectively. Both the length of the follicular phases and of the luteal phases had a positive correlation with the length of the menstrual cycles (Table 1). The Pearson correlation coefficient between the length of the menstrual cycles and the length of the follicular phases or of the luteal phases was .75 ( $P < .001$ ) and .37 ( $P < .001$ ), respectively. Hence, both the follicular phase length and luteal phase length had significant positive correlations with the menstrual cycle length. For analysis of the relationship between the length of the menstrual cycles and the length of the follicular phases or of the luteal phases, we applied the random effect model because the  $P$  value of the Hausman test was .001. The coefficient ( $\alpha, \beta$ ) of the generalized least square estimate was (.501, -.088) for the follicular phase length and (.466, .088) for the luteal phase length.

We then investigated the relationship between the mean length of past cycles and the cycle length, follicular phase length, and luteal phase length of the next menstrual cycle. Table 2 shows the number of cycles having enough records to calculate  $c_{ij}^*(k)$  for each number of  $k$ . For example, there were 11,640 cycles with at least one previous cycle, and there were 7285 cycles having records of 8 past cycles. Table 3 shows the Pearson correlation coefficients between  $c_{ij}^*(k)$  and the next menstrual cycle length, follicular phase length, and luteal phase length. Both the next cycle length and the follicular phase length had strong correlations with the mean cycle length. Only a weak correlation was found between the mean cycle length and the luteal phase length, although it was statistically significant. We applied the random effect model because the  $P$  value of the Hausman test was  $< .050$  for all cases (Table 4). The coefficient ( $\alpha, \beta$ ) of the generalized least square estimate was similar to that calculated for the actual cycle length. In summary, the menstrual cycle length had positive correlations with both the follicular phase length and luteal phase length, although the correlation was less strong with luteal phase length.

The prediction performance of the timing of ovulation using the obtained model,  $\mu_{OPT}$ , was compared with that of  $\mu_{Ogino}$  and  $\mu_{HCL}$ .  $\mu_{OPT}$  outperformed  $\mu_{Ogino}$  when a woman's mean cycle length was shorter than 27 days or longer than 31 days (Figure 1).  $\mu_{OPT}$  outperformed  $\mu_{HCL}$  when a woman's mean cycle length was less than 28 days. As for the mean accuracy over different mean cycle lengths (Figure 2),  $\mu_{OPT}$  outperformed  $\mu_{Ogino}$  in all cases. The prediction performances of  $\mu_{OPT}$  and  $\mu_{HCL}$  were similar when only small numbers of cycles were available to calculate the mean cycle length, whereas  $\mu_{OPT}$  showed an advantage with increasing values of  $k$ .

**Table 1.** Relationship between cycle length, follicular phase length, and luteal phase length.

Cycle length, days	Cycles, n	Mean follicular phase length, days	95% CI	Mean luteal phase length, days	95% CI
23	120	10.5	(7-15)	11.5	(7-15)
24	324	11.1	(7-15)	11.9	(8-16)
25	657	11.5	(8-16)	12.5	(8-16)
26	1065	12.1	(9-16)	12.9	(9-16)
27	1407	12.7	(10-16)	13.3	(10-16)
28	1637	13.4	(10-17)	13.6	(10-17)
29	1516	14.1	(11-17)	13.9	(11-17)
30	1392	15	(11-19)	14	(10-18)
31	1144	15.7	(12-20)	14.3	(10-18)
32	875	16.4	(12-20)	14.6	(11-19)
33	707	17.4	(13-21)	14.6	(11-19)
34	557	18.1	(12-23)	14.9	(10-21)
35	395	18.8	(11-26)	15.2	(8-23)
36	300	19.6	(12-25)	15.4	(10-23)
37	204	20.1	(12-25)	15.9	(11-24)
38	162	21.5	(14-28)	15.5	(9-23)
39	145	21.6	(12-29)	16.4	(9-26)
40	124	22.3	(12-29)	16.7	(10-27)

**Table 2.** Number of cycles having records of k past cycles.

Mean cycle length, days	k=1	2	3	4	5	6	7	8
24	176	168	117	121	90	91	67	68
25	438	496	384	385	325	313	282	272
26	878	918	805	798	694	700	625	608
27	1253	1291	1194	1222	1129	1083	1003	911
28	1569	1529	1507	1440	1382	1293	1196	1154
29	1620	1541	1472	1361	1306	1270	1178	1078
30	1396	1328	1292	1238	1181	1074	998	934
31	1101	1024	1036	979	964	895	860	785
32	862	856	867	728	715	647	624	560
33	721	637	564	569	520	463	429	378
34	532	448	461	371	362	301	276	225
35	377	353	314	253	226	188	195	174
36	290	214	194	179	141	110	88	76
37	219	160	125	90	74	65	47	42
Total	11,640	11093	10421	9804	9154	8529	7898	7285

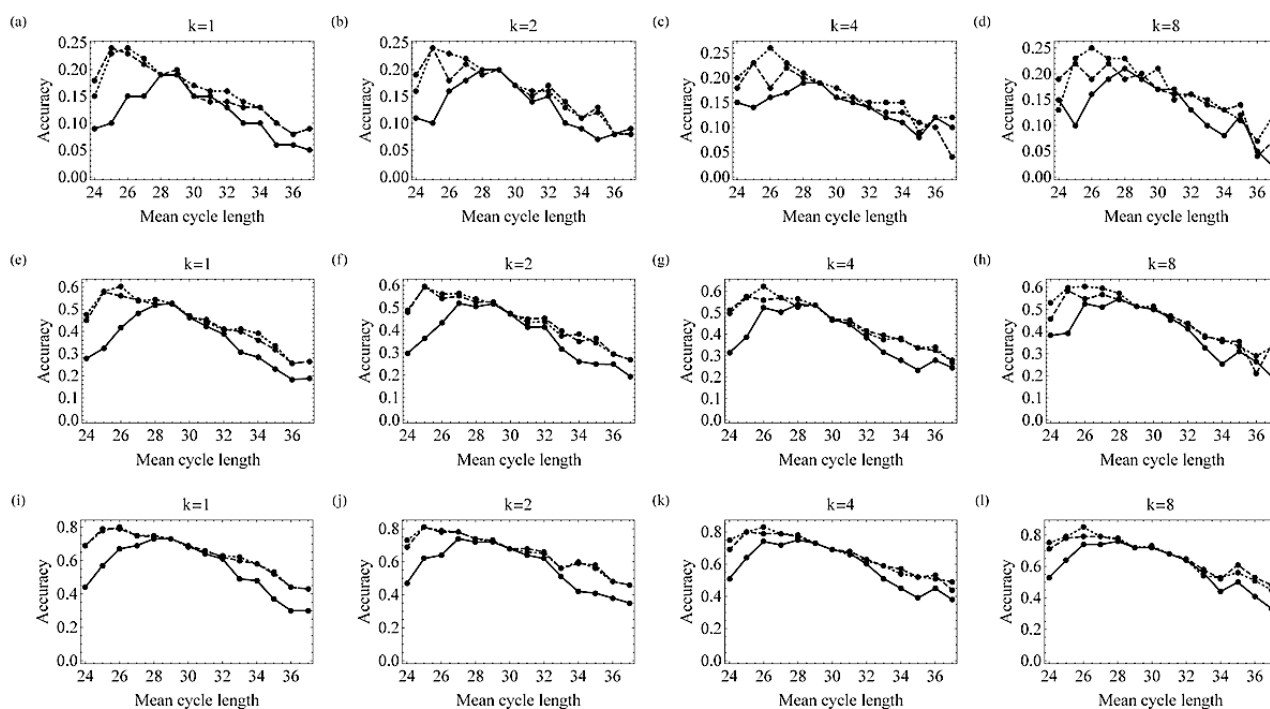


**Table 3.** Pearson correlation coefficients between mean cycle length and next cycle length, follicular phase length, or luteal phase length ( $P < .001$  for all cases).

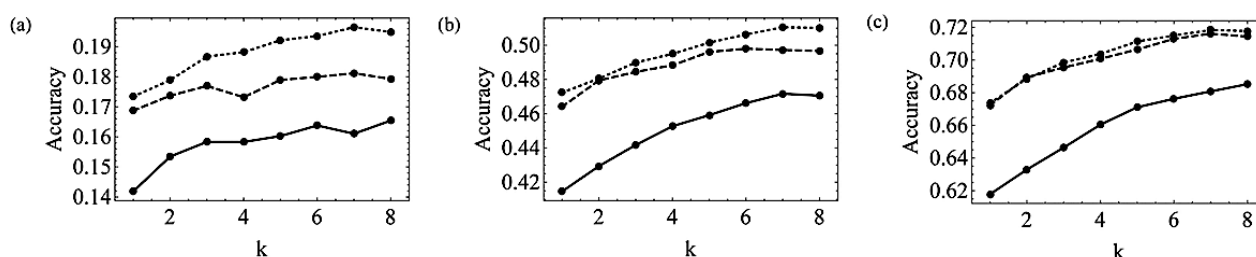
	k=1	2	3	4	5	6	7	8
Next cycle length, days	.543	.574	.584	.589	.589	.593	.592	.592
Follicular phase length, days	.506	.526	.537	.547	.547	.546	.548	.547
Luteal phase length, days	.109	.129	.138	.135	.137	.144	.140	.140

**Table 4.** Results of least square fitting and the Hausman test.

	k=1	2	3	4	5	6	7	8
<b>Follicular phase length</b>								
A	.528	.528	.527	.526	.526	.525	.525	.523
B	.039	.017	.012	.003	-.002	-.010	-.011	-.011
P value	.001	.001	.001	.001	.001	.001	.001	.001
<b>Luteal phase length</b>								
A	.465	.467	.470	.471	.471	.472	.472	.473
B	.039	.017	.012	.003	-.002	-.010	-.011	-.011
P value	.001	.001	.001	.001	.001	.001	.001	.001

**Figure 1.** Accuracy of the Ogino, HCL, and the Optimized methods in predicting the day of the next ovulation date with allowable prediction error=0 (a-d), 1 (e-h), and 2 (i-j) for different mean cycle lengths (solid line, dashed line, and dotted line indicate the Ogino, HCL, and the Optimized methods, respectively).

**Figure 2.** Mean accuracy over different mean cycle lengths with allowable prediction error=0 (a), 1 (b), and 2 (c) (solid line, dashed line, and dotted line indicate the Ogino, HCL, and the Optimized methods, respectively).



## Discussion

### Principal Findings

In this paper, using the data obtained from the Luna Luna service, we evaluated how menstrual cycle length is related to luteal phase length and follicular phase length. As suggested in a previous study [21], the correlation between length of menstrual cycles and length of follicular phases was stronger than the correlation between length of menstrual cycles and length of luteal phases (Table 1). Furthermore, there was a positive correlation between the lengths of past and future menstrual cycles (Tables 3 and 4); this suggests that there is regularity in menstrual cycles as claimed by Creinin et al [22]. A strong positive correlation was also found between the mean length of past cycles and the mean length of follicular phases. Thus, the follicular phase length can be mostly estimated from each woman's menstrual cycle length. However, the correlation between mean cycle length and luteal phase length was also statistically significant. Our results showed that the random effect model was relevant for modeling the relationship between menstrual cycle length and follicular phase length as well as the relationship between menstrual cycle length and luteal phase length. Furthermore, the random effect model was also relevant for modeling the relationships between mean length of past menstrual cycles and mean length of follicular phases or luteal phases. These results suggest that these relationships are shared among all women without being influenced by personal factors. The existing calendar-based methods (Ogino and HCL methods) did not explain our data well. Against the assumption of the Ogino method, the weak positive correlation between mean cycle length and luteal phase length (Tables 3 and 4) suggests that the luteal phase does not have a constant length among women with mean menstrual cycles of different lengths. This is why the Ogino method was less accurate when a woman's mean cycle length was shorter than 27 days or longer than 32 days (Figure 1). The HCL method was less accurate when a woman's mean cycle length was less than 28 days (Figure 1). By taking the average over different mean cycle lengths, the accuracy of the Ogino method was worse than that of the other methods as well (Figure 2). The accuracy of the HCL method was close to that of the Optimized method when there were only a small number of cycles available to calculate the mean cycle length (Figure 2). The Optimized method showed better performance with increasing numbers of available cycles. The Optimized method outperformed the HCL method in explaining the relationship between the mean cycle length and the follicular phase length when a large number of past cycles were available

to calculate the mean cycle length. These results showed that the performance of the Optimized method was equivalent to or better than that of the HCL method. Hence, we recommend using the Optimized method to predict the timing of ovulation from the mean length of menstrual cycles if these data are available.

The novelty of our findings is essentially accounted for by the large number of participants. The Ogino method was developed based on the anatomical observation of ovarian follicles of 81 women with cycle lengths of between 23 and 45 days [15]. They reported that there was a variable luteal phase length (12-16 days), while the number of subjects studied was not large enough to identify the more detailed relationship between the lengths of the menstrual cycle, follicular phase, and luteal phase. Three studies with more data reported the mean (95% CI) of the follicular phase length and the luteal phase length as 16.5 (9-23) and 12.4 (8-17) [21], 15.5 (9-22) and 12.6 (9-17) [23], and 15.0 (2-21) and 13.5 (7-20) days [24], respectively. These studies reported only the mean value over all women, except for Fehring et al [21] who reported positive correlations between menstrual cycle length and follicular phase length or luteal phase length. However, their analysis was based on only 1060 cycles in 141 women. Lamprecht and Grummer-Strawn [16] suggested that follicular phase length was better estimated by  $c^*/2$  based on an analysis of 7514 cycles in 1062 women. In contrast, our study analyzed more than 65,000 cycles, which is 61 and 8.5 times greater than what was used in the studies of Fehring et al [21] and Lamprecht and Grummer-Strawn [16], respectively. This study allowed us to capture the relationships between length of menstrual cycle, follicular phase, and luteal phase in more detail.

Recently, mobile health information technology—known as “digital health,” “eHealth,” or “mHealth”—has been used in medicine and public health in various ways [25-28]. With the recent spread of smartphones and other mobile devices (eg, 77% of US adults [29], 62.6% of Japanese [30], and 62% of Chinese [31] own smartphones), digital health technologies have begun to be adapted for a wide variety of purposes. The mHealth information technologies are one of the origins of the advanced movement called “Quantified Self”, which stresses the role of patients or consumers in medicine and public health [18,19,32]. Quantified-self refers to an individual who is engaged in the self-tracking of any kind of biological, physical, behavioral, or environmental information [19]. These movements are now spreading among people who were not familiar with such technologies before [7]. For example, 60% of US adults are currently tracking their weight, diet, or exercise routine, and 33% are monitoring other factors such as blood sugar, blood

pressure, headaches, or sleep patterns [33]. There are 165,000 mobile phone health apps available on the market [34]. The self-tracked health data are regarded as the key to realizing personalized medicine and health maintenance [19].

On the other hand, there are several concerns about these technologies. First, Lupton [35] argued that they might allow our personal health data be used as “both an object of surveillance and persuasion” [35]. The author in particular claimed that the sharing of health data on social networking services (SNSs) would result in “public surveillance” where everyone would be a subject of surveillance by others. However, it is less beneficial for women to share their menstrual cycle records on SNSs because menstrual cycles strongly depend on biological and physiological factors. Furthermore, issues about menstrual cycles are “sensitive” personal information that would not likely be a topic of conversation among friends. Hence, it is unlikely that women would start sharing information on their menstrual cycles on SNSs. On the other hand, health informatics systems can integrate these data both systematically and anonymously and provide feedback knowledge at a scale that is not achieved by any person-to-person communications. These systems allow women to maximize the benefit of sharing data on their menstrual cycles (or other sensitive health information) without publicizing the data themselves. Sharon [36] argued that self-tracking health technologies reduce phenomena to numbers and “that this simultaneously displaces other, non-quantifiable yet highly insightful means of knowing and expression”. For example, a small percentage of women are aware that they have somatic symptoms around the time of ovulation, for example, ovulation pain. This self-awareness would help women recognize the timing of ovulation. However, not all women have or are aware of these symptoms. Hence, there is a benefit of summarizing regularity behind menstrual cycles in a simple rule and sharing it as social knowledge [17]. We claim that mHealth technology puts this knowledge sharing onto a new stage because it can relax the restriction that the rules must be simple enough to be handled by anyone. Of course, there is loss of information in the rule extraction process. The providers of health informatics systems should take this point seriously and should aim at designing their systems [7-9] so that users can maintain an appropriate distance between their body and its data representation [35,37]. For this purpose, the interaction between a system and its users should be studied in terms of user behavior to find a better way of presenting predictive performance, and these insights should be incorporated into the design of the system [37-39]. Most people are still sensitive about unintended use of data by for-profit companies. In a recent survey on attitudes toward personal health care data [40], more than 80% of the respondents

answered that they were willing to share health and medical information with a personal physician or health care provider (88%) and with nonprofit research organizations (84%). Only 24% answered that they were willing to share their data with for-profit companies. Nevertheless, it is worth noting that 88% answered that their motivation for sharing health information was to make new health discoveries. With increasing numbers of people contributing their health data to analyzable datasets, health information data will not only empower companies and consumers but also benefit health and social systems [18].

## Limitations

Our analysis lacked complete profile data for all subjects and the dataset had inevitable selection bias. Moreover, menstrual bleeding started in the middle of the night in some cases—the date taken as the menstrual start day depended on the user’s subjective choice. However, we believe that having a large-scale dataset available to carry out investigations on women’s health overcomes such limitations.

## Conclusions

Our study demonstrated how the present calendar methods of predicting the ovulation date were improved by the better grouping of women, which can be supported statistically only with massive numbers of subjects. Because calendar methods require only the timing of menstruation, one of the calendar methods would be the best option for couples during the early stages of their attempt to have a baby or for those who want to avoid the cost associated with other methods. Moreover, the results of the calendar method will be the baseline for more advanced methods that integrate other biological metrics. The mobile phone based health care services are very efficient in obtaining large datasets because they offer easy ways for users to input and manage their personal data. This recently enabled data collection framework is complementary to existing well-controlled experimental methods and will contribute to the testing of medical hypotheses that previously could not be studied due to insufficient numbers of subjects. Users benefit from newly developed medical knowledge by using mobile phone based services without the need to learn intricate calculations. For medical personnel and researchers, the records accumulated by these commercial services can be a useful source of data for analysis after appropriate anonymity processing. Thus, in the mobile phone based services that aim at facilitating conception (or contraception), medical specialists and users form a knowledge-improving cycle that can provide quick feedback to the users from the emerging analysis results. Such systems, including other mobile phone health care services, are strongly expected to contribute to comprehensive health care for people of all ages.

## Acknowledgments

We would like to thank Cornelis B Lambalk for valuable comments and suggestions. MTI Ltd provided support in the form of salaries for authors KS and II but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.



## Authors' Contributions

The study idea was conceived by all authors. The study protocol and analysis were developed by KS and II. The first draft was written by KS and II. SS participated in the revision of the paper and critical discussions. All authors approved the final draft.

## Conflicts of Interest

KS and II are employees of MTI Ltd. In addition, KS and II have a patent (Japan Patent Office, No. P5998307) pending.

## References

- Dunson DB, Colombo B, Baird DD. Changes with age in the level and duration of fertility in the menstrual cycle. *Hum Reprod*. May 2002;17(5):1399-1403. [Medline: [11980771](#)]
- Stanford JB, White GL, Hatasaka H. Timing intercourse to achieve pregnancy: current evidence. *Obstet Gynecol*. Dec 2002;100(6):1333-1341. [Medline: [12468181](#)]
- 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*. Dec 07, 1995;333(23):1517-1521. [doi: [10.1056/NEJM199512073332301](#)] [Medline: [7477165](#)]
- Mihm M, Gangooly S, Muttukrishna S. The normal menstrual cycle in women. *Anim Reprod Sci*. Apr 2011;124(3-4):229-236. [doi: [10.1016/j.anireprosci.2010.08.030](#)] [Medline: [20869180](#)]
- Colombo B, Masarotto G. Daily Fecundability. *DemRes*. Sep 06, 2000;3:5. [FREE Full text] [doi: [10.4054/DemRes.2000.3.5](#)]
- Spieler JM, Collins WP. Potential fertility - defining the window of opportunity. *J Intern Med Res*. 2001;29(1\_suppl):3A-13A. [FREE Full text] [doi: [10.1177/14732300010290S102](#)]
- Rapp A, Cena F. Personal informatics for everyday life: How users without prior self-tracking experience engage with personal data. *International Journal of Human-Computer Studies*. Oct 2016;94:1-17. [FREE Full text] [doi: [10.1016/j.ijhcs.2016.05.006](#)]
- Epstein DA, Ping A, Fogarty J, Munson SA. A lived informatics model of personal informatics. 2015. Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2015:731-742; Osaka, Japan.
- Li I, Dey A, Forlizzi J. A stage-based model of personal informatics systems. 2010. Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2010:557-566; Atlanta, GA. URL: <https://pdfs.semanticscholar.org/f90d/ffcab9850a03a4733abb31c123f244ff5f6c.pdf>
- Simmons RG, Shattuck DC, Jennings VH. Assessing the Efficacy of an App-Based Method of Family Planning: The Dot Study Protocol. *JMIR Res Protoc*. Jan 18, 2017;6(1):e5. [doi: [10.2196/resprot.6886](#)]
- Mangone ER, Lebrun V, Muessig KE. Mobile Phone Apps for the Prevention of Unintended Pregnancy: A Systematic Review and Content Analysis. *JMIR Mhealth Uhealth*. Jan 19, 2016;4(1):e6. [FREE Full text] [doi: [10.2196/mhealth.4846](#)] [Medline: [26787311](#)]
- Lupton D. Quantified sex: a critical analysis of sexual and reproductive self-tracking using apps. *Cult Health Sex*. 2015;17(4):440-453. [doi: [10.1080/13691058.2014.920528](#)] [Medline: [24917459](#)]
- Pallone SR, Bergus GR. Fertility awareness-based methods: another option for family planning. *J Am Board Fam Med*. 2009;22(2):147-157. [FREE Full text] [doi: [10.3122/jabfm.2009.02.080038](#)] [Medline: [19264938](#)]
- Fehring RJ. New low- and high-tech calendar methods of family planning. *J Midwifery Womens Health*. 2005;50(1):31-38. [doi: [10.1016/j.jmwh.2004.07.001](#)] [Medline: [15637512](#)]
- Ogino K. Über den konzeptionstermin des weibes und seine anwendung in der praxis. *Zentralbl Gynakol*. 1932;56:721-732.
- Lamprecht VM, Grummer-Strawn L. Development of new formulas to identify the fertile time of the menstrual cycle. *Contraception*. Dec 1996;54(6):339-343. [Medline: [8968662](#)]
- Arévalo M, Jennings V, Sinai I. Efficacy of a new method of family planning: the Standard Days Method. *Contraception*. May 2002;65(5):333-338. [Medline: [12057784](#)]
- Swan M. Health 2050: The realization of personalized medicine through crowdsourcing, the quantified self, and the participatory biocitizen. *Journal of Personalized Medicine*. 2012;2:93-118. [doi: [10.3390/jpm2030093](#)]
- Swan M. The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data*. 2013;1:85-99. [doi: [10.1089/big.2012.0002](#)] [Medline: [27442063](#)]
- Schmidheiny K, Basel U. Panel data: fixed and random effects. In: *Short Guides to Microeconometrics*. Basel, Switzerland. Unversitat Basel; 2011:2-7.
- Fehring R, Schneider M, Raviele K. Variability in the phases of the menstrual cycle. *J Obstet Gynecol Neonatal Nurs*. 2006;35(3):376-384. [doi: [10.1111/j.1552-6909.2006.00051.x](#)] [Medline: [16700687](#)]
- Creinin MD, Keverline S, Meyn LA. How regular is regular? An analysis of menstrual cycle regularity. *Contraception*. Oct 2004;70(4):289-292. [doi: [10.1016/j.contraception.2004.04.012](#)] [Medline: [15451332](#)]
- France JT, Graham FM, Gosling L, Hair P, Knox BS. Characteristics of natural conceptual cycles occurring in a prospective study of sex preselection: fertility awareness symptoms, hormone levels, sperm survival, and pregnancy outcome. *International Journal of Fertility*. 1991;37:244-255.

24. World Health Organization. A prospective multicentre study of the ovulation method of natural family planning. IV. The outcome of pregnancy. World Health Organization. Fertil Steril. Apr 1984;41(4):593-598. [Medline: [6705935](#)]
25. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res. May 21, 2013;15(5):e95. [FREE Full text] [doi: [10.2196/jmir.2430](#)] [Medline: [23697600](#)]
26. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med. 2013;10(1):e1001362. [FREE Full text] [doi: [10.1371/journal.pmed.1001362](#)] [Medline: [23349621](#)]
27. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform. Feb 2012;45(1):184-198. [FREE Full text] [doi: [10.1016/j.jbi.2011.08.017](#)] [Medline: [21925288](#)]
28. Boulos MN, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online. Apr 05, 2011;10:24. [FREE Full text] [doi: [10.1186/1475-925X-10-24](#)] [Medline: [21466669](#)]
29. Pew Research Center. Device ownership over time. URL: <http://www.pewinternet.org/fact-sheet/mobile/> [accessed 2017-02-03] [WebCite Cache ID 6nzTpWmFv]
30. Economic Research Office, General Policy Division, Information and Communications Policy Bureau. White Paper. Information and Communications in Japan. Tokyo. Ministry of Internal Affairs and Communications, Japan; 2015.
31. Hsu J, Liu D, Yu Y, Zhao HT, Chen ZR, Li J, et al. The Top Chinese Mobile Health Apps: A Systematic Investigation. J Med Internet Res. Aug 29, 2016;18(8):e222. [FREE Full text] [doi: [10.2196/jmir.5955](#)] [Medline: [27573724](#)]
32. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int J Environ Res Public Health. Dec 2009;6(2):492-525. [FREE Full text] [doi: [10.3390/ijerph6020492](#)] [Medline: [19440396](#)]
33. Fox S, Duggan M. Mobile health 2012. Washington, DC. Pew Internet & American Life Project; 2012.
34. IMS Institute for Healthcare Informatics. URL: <https://www.imshealth.com/files/web/TMSH%20Institute/Reports/Patient%20Adoption%20of%20mHealth/IIHI-Patient-Adoption-mhealth-Exhibits-Full.pdf> [accessed 2017-02-03] [WebCite Cache ID 6nzUmYHJS]
35. Lupton D. M-health and health promotion: The digital cyborg and surveillance society. Soc Theory Health. Jun 27, 2012;10(3):229-244. [FREE Full text] [doi: [10.1057/sth.2012.6](#)]
36. Sharon T. Self-Tracking for Health and the Quantified Self: Re-Articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare. Philos Technol. Apr 18, 2016;30(1):93-121. [doi: [10.1007/s13347-016-0215-5](#)]
37. Schüll ND. Data for life: Wearable technology and the design of self-care. BioSocieties. Oct 13, 2016;11(3):317-333. [doi: [10.1057/biosoc.2015.47](#)]
38. Bentley F, Tollmar K, Stephenson P, Levy L, Jones B, Robertson S, et al. Health Mashups. ACM Trans Comput-Hum Interact. Nov 01, 2013;20(5):1-27. [doi: [10.1145/2503823](#)]
39. Rapp A, Tirassa M. Know Thyself: A Theory of the Self for Personal Informatics. Human-Computer Interaction. Jan 27, 2017;32(5-6):335-380. [doi: [10.1080/07370024.2017.1285704](#)]
40. Pickard K, Swan M. Big desire to share big health data: A shift in consumer attitudes toward personal health information. 2014. Presented at: AAAI Spring Symposium; 2014:51-58; Palo Alto, CA. URL: <https://www.aaai.org/ocs/index.php/SSS/SSS14/paper/viewFile/7765/7783>

## Abbreviations

**HCL:** half cycle length

**OPT:** optimal

**SNS:** social networking service

*Edited by G Eysenbach; submitted 07.Feb.2017; peer-reviewed by K Blake, A Rapp; comments to author 20.Apr.2017; revised version received 11.Aug.2017; accepted 08.Sep.2017; published 27.Nov.2017*

*Please cite as:*

Sohda S, Suzuki K, Igari I

Relationship Between the Menstrual Cycle and Timing of Ovulation Revealed by New Protocols: Analysis of Data from a Self-Tracking Health App

J Med Internet Res 2017;19(11):e391

URL: <http://www.jmir.org/2017/11/e391/>

doi: [10.2196/jmir.7468](#)

PMID: [29180346](#)

©Satoshi Sohda, Kenta Suzuki, Ichiro Igari. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 27.Nov.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.