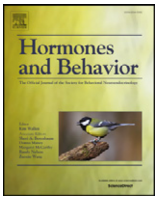
Hormones and Behavior 81 (2016) 74–83

Contents lists available at ScienceDirect

Hormones and Behavior

journal homepage: www.elsevier.com/locate/yhbeh



Standardized protocols for characterizing women's fertility: A data-driven approach

Khandis R. Blake [a,*], Barnaby J.W. Dixson [b], Siobhan M. O'Dean [a], Thomas F. Denson [a]

[a] *School of Psychology, Mathews Building, The University of New South Wales, UNSW, Sydney, NSW 2052, Australia*
[b] *School of Psychology, The University of Queensland, Brisbane St Lucia, QLD 4072, Australia*

ARTICLE INFO

Article history:
Received 29 September 2015
Revised 9 March 2016
Accepted 18 March 2016
Available online 10 April 2016

Keywords:
Ovulation
Luteinizing hormone
Menstrual cycle
Ovulatory shift
Forward-counting method
Backward-counting method
Fertility
Fertile window

ABSTRACT

Experts are divided on whether women's cognition and behavior differs between fertile and non-fertile phases of the menstrual cycle. One of the biggest criticisms of this literature concerns the use of indirect, imprecise, and flexible methodologies between studies to characterize women's fertility. To resolve this problem, we provide a data-driven method of best practices for characterizing women's fertile phase. We compared the accuracy of self-reported methods and counting procedures (i.e., the forward- and backward-counting methods) in estimating ovulation using data from 140 women whose fertility was verified with luteinizing hormone tests. Results revealed that no counting method was associated with ovulation with >30% accuracy. A minimum of 39.5% of the days in the six-day fertile window predicted by the counting methods were non-fertile, and correlations between counting method conception probabilities and actual conception probability were weak to moderate, $rs = 0.11$–0.30. Poor results persisted when using a lenient window for predicting ovulation, across alternative estimators of the onset of the next cycle, and when removing outliers to increase the homogeneity of the sample. By contrast, combining counting methods with a relatively inexpensive test of luteinizing hormone predicted fertility with accuracy >95%, but only when specific guidelines were followed. To this end, herein we provide a cost-effective, pragmatic, and standardized protocol that will allow researchers to test whether fertility effects exist or not.

Introduction

More than a century has passed since Walter Heape (1900) observed that in some mammals, females were only sexually receptive to males during a brief period prior to menstruation. He named this period *estrus*. In rats, cats, cattle, and sheep, female sexual activity is indeed rigidly controlled by ovarian hormones, with ovulation accompanied by striking behavioral changes (Dixson, 2012; Feder, 1981). This pattern, however, is not universal in mammals (Dixson, 2012; Wallen, 2001). For instance, in some anthropoid primates copulations continue after ovariectomy, suggesting a decoupling of hormonal control and sexual behavior (Dixson, 2012; Wallen, 2001).

Women are sexually receptive across the ovulatory cycle; they engage in sexual activity in the fertile (periovulatory) and non-fertile (luteal and menses) phases of the menstrual cycle. This fact has led to considerable debate surrounding a putative loss of estrus in modern humans (Havlíček et al., 2015). Scholars have suggested that women "lost" estrus, either before or after the evolutionary divergence between humans and nonhuman primates (Dixson, 2009; Symons, 1979). Because biological concepts have historically been used to oppress women (e.g., Chrisler and Caplan, 2002), women's apparent loss of estrus has been interpreted by some as a needed departure from biological determinism and a demonstration of women's capacity to regulate their own sexual behavior (Wu, 2015).

An alternative view that has gained considerable attention suggests that fertility-induced variations in women's sexual behavior and mate preferences are evolved adaptations (Thornhill and Gangestad, 2008). For example, there is evidence that sexual attractiveness judgments, women's sexual proceptivity, and self-reported sexual receptivity are all heightened during high-fertility periods of the menstrual cycle (Gangestad and Haselton, 2015). Researchers in these fields suggest that such ovulatory shifts are adaptive as they may increase the likelihood of women passing on certain genetic benefits to their offspring, ultimately increasing their own reproductive success (Thornhill and Gangestad, 2008). These findings also suggest that modern females do experience estrus, albeit in a more subtle form than many other mammalian species (Thornhill and Gangestad, 2008).

Whether women possess estrus (or whether it is concealed, see Puts et al., 2013) has powerful implications for understanding the evolution of women's mating strategies. As a result, much research has been

---

* Corresponding author at: School of Psychology, University of New South Wales, UNSW, Sydney, NSW 2052, Australia.
E-mail address: k.blake@psy.unsw.edu.au (K.R. Blake).

http://dx.doi.org/10.1016/j.yhbeh.2016.03.004
0018-506X/

devoted to testing hypotheses that women's sexual behavior and mate preferences shift when they are fertile. Two independently conducted meta-analyses of this literature have cast doubt on the robustness of the research findings in this field. One meta-analysis found no overall support for the hypothesis that women's mate preferences shift across the ovulatory cycle (Wood et al., 2014). The second meta-analysis reported significant and small-to-moderate ovulatory shift effects in preferences for short-term romantic partners (Gildersleeve et al., 2014a). Numerous commentaries followed these meta-analyses, some of which defended that periovulatory shifts were, at least in some cases, adaptations (Jones, 2014; Gangestad, 2015). Others suggested that positive findings were artifacts of poorly defined and variably applied methodology (Harris et al., 2014; Wood, 2015; Wood and Carden, 2014).

The conflicting findings generated by these meta-analyses highlighted a recurring concern in the ovulatory cycle shift literature: the lack of an evidence-based, a priori methodology for characterizing women's fertile phase (Harris et al., 2013; Wood, 2015). Scientific methods vary and change over time, and it is frequently noted that the methods used to quantify women's behavior across the menstrual cycle vary between studies (Harris et al., 2013, 2014; Gildersleeve et al., 2013, 2014a, 2014b). Because of their cost effectiveness and ease of use, most researchers have relied on self-report methods and counting procedures to determine when participants are fertile and non-fertile. However, due to memory biases, the reliability of such measures remains in question (Jukic et al., 2008; Small et al., 2007; Wood et al., 2014). The absence of best practice standardized protocols for determining women's fertility is a major impediment to understanding whether the menstrual cycle influences psychological outcomes.

One recommended strategy to curtail criticisms of the ovulatory shift literature is to provide more detailed and validated methodology that can be readily used by other researchers seeking to test effects of fertility on behavior (Harris et al., 2014). Such an endeavor would allow researchers to apply preregistration criteria and analyze results using validated and replicable methodology that explicitly test a priori hypotheses (Wagenmakers et al., 2012). Recently, two studies employed statistical simulations and large published datasets on women's menstrual cycles to test the efficacy of counting methods on characterizing women's fertility (Gonzales and Ferrer, 2015; Gangestad et al., 2015). While these two studies differ in their approaches, methodology, and to some extent their goals, both independently concluded that weaknesses in design (between-subject), procedure (counting methods), and statistical power (small sample sizes) have together contributed to mixed findings in the literature. Both Gonzales and Ferrer (2015) and Gangestad et al. (2015) provided power analyses and designated the required sample size for researchers to detect large effects when using indirect counting methods to estimate fertility. Both research teams also recommended that studies seeking to compare changes at high and low fertility should ideally use within-subjects designs and measure fertility directly (e.g., through hormonal or ultrasound verification; Gonzales and Ferrer, 2015; Gangestad et al., 2015).

In the current study, we build upon the recent methodological advancements by Gangestad et al. (2015) and Gonzales and Ferrer (2015) by providing a concise set of data-driven methodological considerations for academic researchers who wish to test hypotheses that necessitate comparisons between fertile and non-fertile phases of the menstrual cycle. The current study provides multiple evidence-based comparisons of the accuracy of counting methods used to predict women's fertility with a well-validated, relatively inexpensive hormonal measure of ovulation (luteinizing hormone; LH). We further provide recommendations for: a) increasing sample homogeneity to improve ovulation prediction accuracy; b) the cycle length estimator that best prospectively estimates the next cycle onset; c) the number of days prior to their predicted ovulation when women should start LH testing; and d) the duration of LH testing. The ultimate contribution of this research is to provide pragmatic guidelines for future studies testing effects of fertility on women's behavior that are ideal for researchers within university settings who may have limited time and funding.

Method

Ethics statement

This study was carried out in accordance with the Declaration of Helsinki and was part of a larger two-session experiment examining fertility effects on women's behavior. Written informed consent was obtained from all participants and the UNSW Ethics Committee approved the protocol. Participants were informed that they were contributing to a study investigating reproductive biology and self-presentation.

Exclusion criteria

Participants were recruited in two waves (outlined below). Inclusion criteria for both waves excluded use of hormonal birth control (current or within the past two months; e.g., birth control pills, Norplant, vaginal ring, birth control patch, Depo-Provera, Mirena IUD); highly irregular menstrual cycles; pregnancy/breastfeeding (current or recent); immune, cardiovascular, metabolic, or kidney disorders; anabolic steroid use; cancer/tumors, or recreational drug use within the past 30 days.[1] Inclusion criteria in Wave 2 additionally excluded medically diagnosed fertility problems, polycystic ovarian syndrome, and endometriosis.

Participants

One hundred and seventy-eight women ($M_{\text{age}} = 21.56$, $SD = 4.42$) participated in the study. Participants in Wave 1 ($n = 91$) were recruited from August 2014 to May 2015 and participated as part of their course requirements or in exchange for AUD$45. Wave 2 participants ($n = 87$) were recruited from August 2015 to December 2015 and participated for course credit or in exchange for AUD$70. Data were excluded from 12 participants who did not follow or received incorrect LH testing instructions. Four participants were excluded as they reported no LH surge but skipped two or more consecutive days of LH testing, and a further four participants were excluded due to lying about test results. All women in Wave 1 reported cycles lasting 23–35 days; 18 participants from Wave 2 were withdrawn before attempting any LH testing for prior cycles lasting <23 or >36 days, leaving a final sample of 140 participants ($M_{\text{age}} = 21.76$, $SD = 4.60$).

Sample characteristics

Participants ranged from 18 to 38 years old; the majority (81.0%) was aged 18–24 years. Participants were primarily Asian (65%) and Caucasian (25%); the remainder was Middle Eastern (1.4%), African American (0.7%), or other ethnicities (7.9%). The majority of women (92.1%) reported being at or below the mid-point on a 7-point sexuality scale (1 = attracted only to men, 7 = attracted only to women), with 62.9% reporting attraction to men only. Fifty-five percent of participants were single, 35.7% were in committed romantic relationships, and the remainder did not report their relationship status (or reported it as 'other').

---

[1] The prescreening items were a requirement of the larger experiment which sampled steroid hormones; the following items are not necessarily required for urinary LH test prescreening: immune, cardiovascular, metabolic, or kidney disorders; anabolic steroid use; cancer/tumors, or recreational drug use within the past 30 days.

Materials and procedure

Menstrual cycle self-report

After pre-screening, participants completed a questionnaire on their average menstrual cycle length (which we refer to as *reported length*), their confidence in that length, the date of their last menses onset, and the date of the predicted onset of their next menses. Women indicated their reported length by text response or on a 13-point scale (‘23 days or under’, then in one-day intervals to ‘35 days or over’[2]). Participants reported their confidence in their typical menstrual cycle their length (‘1’ = not at all confident to ‘9’ = very confident).

Ovulation prediction procedure

The majority of behavioral research across the menstrual cycle used indirect counting methods to estimate ovulation (Gildersleeve et al., 2014a; Gildersleeve et al., 2014b; Harris et al., 2013; Wood et al., 2014). These methods include the forward-counting and backward-counting methods and are usually used in concert with recollected cycle length and menses onset (alternative methods record actual cycle length and onset within experimental conditions, though in practice this method is infrequent).

The *forward-counting* method predicts ovulation to occur 14–15 days after the menstrual cycle onset (counting the cycle onset day as Day 1). This method relies on hormonal and ultrasound data which indicate that the mean follicular phase length for women aged 19–42 years is 14.6 days (ovulation occurs at the end of the follicular phase; Ecochard and Gougeon, 2000). However, the follicular phase remains more variable than the luteal phase (Baird et al., 1995; Waller et al., 1998). The *backward-counting* method bypasses this variability and estimates ovulation by subtracting 14 days from the next predicted menses onset (Dixon, 1980). The backward-counting method may therefore be more reliable than the forward-counting method when using self-reported data (Gildersleeve et al., 2013).

We estimated the day of ovulation using the backward-counting method (Dixon, 1980). Due to participants' natural variation in cycle phase upon entering the study and the design of the larger experiment, 14 women in Wave 1 began LH testing during their immediate menstrual cycle (C1), 52 women tested during their next menstrual cycle (C2), 12 women tested in their third menstrual cycle (C3), and two women tested in their fourth menstrual cycle (C4).[3] Wave 2 participants tested in C2 ($n = 30$), C3 ($n = 29$), or C4 ($n = 1$), assignment was random for all except eight participants who requested to test during their subsequent cycle. For women testing in C1, we used their reported length to estimate the onset of their next cycle, predicting ovulation to occur 14 days prior to this date. We followed the same process for women testing in their second, third, and fourth cycle, but substituted their reported length with their previous cycle length or with a truncated average of their C1, C2, and C3 lengths. Given that women vary in their accuracy in recollecting the lengths of their menstrual cycles (Jukic et al., 2008; Small et al., 2007), we believed that cycle lengths measured within the experiment were more likely to be accurate than those recollected from cycles prior to participating in the study. All women testing in C2–C4 had their cycle onset and offset recorded in the experiment, thus bypassing recall inaccuracies.

Ovulation validation procedure

The forward- and backward-counting methods are potentially problematic as participants may inaccurately recall the details of their menstrual cycles (Jukic et al., 2008; Small et al., 2007). Similarly, although women are only fertile around ovulation, there is considerable variability between- and within-women regarding the onset of the period of fertility (Fehring et al., 2006; Wilcox et al., 1995). Some researchers have attempted to circumvent these issues by testing ovulatory shift effects against multiple available methods (Dixson and Brooks, 2013; Dixson et al., 2013; Harris, 2011, 2013; Zietsch et al., 2015). However, co-variation in lifestyle and medical factors may lead to incorrect self-report responses (Jukic et al., 2008), rendering any conclusions drawn from estimated fertility generated from self-reports problematic.

To validate women's fertility and bypass issues with indirect ovulation prediction methods, we used commercially available urinary LH tests (Blue Cross Bio-Medical Co. LTD, CE/FDA Registered). LH tests are relatively inexpensive (approximately USD$0.50 each) and considered the most validated method for estimating ovulation (Guida et al., 1999); following an LH surge, ovulation is expected to occur approximately 24–48 h later (Collins et al., 1991; Direito et al., 2013; Guermandi et al., 2001; Pearlstone and Surrey, 1994; Testart et al., 1981; Testart and Frydman, 1982). All tests preceded their expiry date and detected LH surges with >99% specificity at 25 ml U/ml sensitivity. Participants met with a female member of the research team and were trained in LH test usage. The experimenter provided participants with a detailed instruction handout with diagrams, urine sample cups, and LH test strips. Participants were instructed to test between 10:00 h to 20:00 h daily and report the result immediately to the research team via SMS or email. If no result was reported, the experimenter followed up with the participant either that day or the next day via SMS or email. We also sent emails or SMS messages prior to the first testing day to remind women to begin testing.

Wave 1 participants were instructed to begin LH testing at least three days prior[4] to their predicted day of ovulation and continue testing daily until a positive result was reported, or for a minimum of seven days (i.e., at least three days after the predicted date of ovulation). Wave 2 participants tested five days prior to their predicted day of ovulation and continued testing for an 11-day period. In Wave 2, participants additionally watched an LH testing instruction video designed by the authors, received daily SMS messages to remind them to test, and were required to submit daily photographs of test results including the time the test was taken. Photograph results were assessed by two coders until consensus was reached. After reporting a positive result, all participants were instructed to report the onset of their next menstrual cycle, allowing us to estimate their actual backward-counted day of ovulation.

Results

Reported menstrual cycle results

Across waves, the self-reported length of participants' cycles ranged from 20 to 35 days ($M = 29.02$, $SD = 2.87$). Confidence in their self-

---

[2] Of the 80 participants who reported their cycle length using the 13-point scale, six reported cycles lasting 23 days or under or 35 days or over. To determine whether we should record their reported length as 23 days, 35 days, or as a different figure, we asked these participants whether their cycles lasted for generally 23 or 35 days, or were usually much shorter or much longer. All participants reported their cycles lasted for specifically 23 or 35 days. The correlation between their reported length (recorded as ‘23’ or ‘35’) and their C1 length was strong, $r(4) = 0.94$, $p = 0.006$, and four of these six participants reported positive LH test results.

[3] Five women reported negative results in their first round of testing (one woman testing in C1, four women testing in C2) and re-tested in their next cycle. Of these, two women reported negative results again and concluded their participation, the remaining three women reported a positive result. In all cases we retained the data from the second round of testing only.

[4] We recommended that 15 women begin testing earlier than three days prior to their predicted ovulation. These recommendations were an ad hoc adjustment to the study protocol because we saw a large discrepancy between a woman's reported cycle length and her C1 length, or when a woman reported that her cycle length confidence was low (≤6). Four participants reported a negative result but missed their first test day (they were instructed to begin testing three days' before their predicted ovulation, but began testing two days' prior to their predicted ovulation).

reported length ranged from ‘2’ to ‘9’, with 79.3% reporting their length confidence at ‘7’ or above. Correlations between self-reported length, C1–C4 lengths, and truncated averages of C1–C4 lengths are an indication of the accuracy of self-reported versus actual cycle lengths. We found significant correlations between self-reported length and C1 length, ($r(136) = 0.46$, $p < 0.001$), C2 length ($r(112) = 0.50$, $p < 0.001$), and C3 length ($r(39) = 0.59$, $p < 0.001$). We also found a very large correlation ($r(39) = 0.68$, $p < 0.001$) between self-reported length and the averaged C1, C2, C3 length. These data indicated that cycle length recall was most accurate when cycle length was averaged across time. The majority of participants (62.5%) had an intra-cycle variability of three days or less, 25.0% had an intra-cycle variability of 4–9 days, and 5.0% had an intra-cycle variability that exceeded 10 days (this was despite all participants reporting regular menstrual cycles).

LH test results

One hundred and three participants (73.6%) reported an LH surge during their testing phase. The average number of days testing prior to reporting an LH surge was 5.56 days ($SD = 2.48$; range = 1–11 days). Participants who did not report a surge tested for an average of 9.08 days ($SD = 1.52$; range = 6–11 days). Of the 103 participants who reported an LH surge, eight did not report their next cycle onset (7.8%), leaving a sample of 95 women for investigating the accuracy of the backward-counting method. Positive test days ranged from 4 to 33 days prior to the next actual cycle onset ($M = 14.26$, $SD = 3.81$). According to the forward-counting method, positive test days ranged from Day 9 to Day 27 ($M = 16.80$, $SD = 3.16$).

Ovulation prediction accuracy: a comparison of counting methods versus LH tests

Indirect counting methods have yet to be validated against objectively determined hormonal measures in non-simulated data. To examine the validity of the counting methods for predicting ovulation, we quantified how accurately the forward- and backward-counting methods predicted the positive LH test results in the current sample. To do this, we analyzed differences between the actual positive surge day and the predicted ovulation day according to the forward- and backward-counting methods. We first conducted a literature review of 26 studies on the timing of ovulation relative to the LH surge (see Appendix A). This review concluded that most studies report ovulation occurs approximately 24–48 h after an LH surge (e.g., Collins et al., 1991; Direito et al., 2013; Guermandi et al., 2001; Pearlstone and Surrey, 1994; Testart et al., 1981; Testart and Frydman, 1982). Hence, if the predicted ovulation day was one or two days after the actual positive surge day we recorded this as an accurate ovulation prediction. If the predicted ovulation day was on or prior to the LH surge day, or more than two days after the LH surge day, we recorded this as an inaccurate prediction. These analyses only included ovulatory cycles ($n = 103$).

Forward-counting method accuracy

To examine the accuracy of the forward-counting method, we determined the percentage of ovulatory cycles with positive LH surges on forward-counted Days 12–13 inclusive. Eleven percent (10.7%) of participants reported a positive LH surge on these days and an additional 11.7% reported a positive surge on forward-counted Day 14. Given that the forward-counting method predicts ovulation will occur somewhere between Days 14–15, we also examined the percentage of ovulatory cycles with positive LH surges on Days 13–14. The percentage of positive surges occurring in this window was 16.5%, and an additional 14.6% of ovulatory cycles had positive surges on Day 15.

Backward-counting method accuracy

To predict ovulation prospectively using the backward-counting method, the next menses onset must be predicted. Thus, the method used to estimate cycle length and predict the onset of the next menses are key elements affecting the accuracy of the backward-counting method. To test the validity of the backward-counting method and determine the most accurate predictor of the next menses, we created four cycle length estimators. These cycle length estimators allowed us to compare whether cycle lengths measured within the study, recalled cycle lengths, or a combination of the two most accurately predicted the length of the LH testing cycle (and thus the next menses onset).

The first estimator consisted of participants' self-reported length from the menstrual cycle questionnaire, which is the most commonly used estimator of the next menstrual cycle onset (Wood et al., 2014). The next estimator was the length of the most recent menstrual cycle (the cycle immediately preceding the testing cycle), hereafter termed ‘length of prior cycle’. This value was derived by recording cycle onset then measuring the cycle offset within in the experiment. The next estimator was an average of the prior cycle and reported cycle length estimators. The final estimator was an average of all available cycle data, including the reported length. A summary of these estimators is provided in Table 1.

An analysis of the correlations between the actual testing cycle length and the cycle length estimators provided an indication of the accuracy of the estimators in predicting the next menses onset (Table 1). Reported length showed a moderate-to-large correlation with the testing cycle length, $r(112) = 0.44$, $p < 0.001$. All other correlations were large, $rs(100) \geq 0.53$, $ps < 0.001$.

Next, to approximate how accurately the cycle length estimators predicted the likely day of ovulation, we added the estimated lengths to the most recent cycle onset then subtracted the actual LH surge cycle day (forward-counted). This provided four alternative backward-counted LH surge days, one for each length estimator. We then determined the percentage of ovulatory cycles with positive LH surges 15–16 days prior to the next predicted cycle onset according to these estimators (these participants were most likely to have ovulated on backward-counted Day 14). We further examined the percentage of ovulatory cycles reporting positive LH surges on backward-counted Day 14, thus accounting for the small percentage of women who ovulate on the same day as their LH surge. As shown in Table 2, the maximum percentage of ovulatory cycles with a positive LH surge 15–16 days

Table 1
Alternate cycle length estimators for next cycle onset prediction (including non-ovulatory cycles).

Parameter	Notes	Correlation with actual testing cycle length
Reported length	Reported length from the pre-screening survey ($n = 140$)	0.44***
Length of prior cycle	Length of the last cycle prior to LH testing (excludes reported cycle length, only where a positive surge was reported in C2, C3, or C4; $n = 126$)	0.53***
Prior cycle + reported average	Length of the last cycle prior to LH testing plus the reported cycle length, averaged (where a positive surge was reported in C2, C3, or C4; $n = 126$)	0.57***
Average of all measured and reported lengths	Length of all cycle data (only for participants who reported both measured and reported cycles prior to testing—where a positive surge was reported in C2, C3, or C4; $n = 126$)	0.54***

*** $p < 0.001$.

Table 2
Percentage of positive LH surge results occurring in specific windows according to four alternative cycle length estimators (ovulatory cycles only).

Cycle length estimator	Percentage of positive LH surge results		
	15 or 16 days prior to the next cycle onset	14 days prior to the next cycle onset	Within 0–2 days prior to the actual backward-counted LH surge day
Reported length ($n = 103$)	28.2%	11.7%	23.2%
Length of prior cycle ($n = 91$)	15.4%	15.4%	21.7%
Prior cycle + reported average ($n = 91$)	19.8%	16.5%	26.5%
Average of all measured and reported lengths ($n = 91$)	20.9%	14.3%	26.5%

Table 3
Percentage of luteinizing hormone-verified fertile days in the fertile windows outlined by the counting methods (ovulatory cycles only).

Forward-counting method			Backward-counting methods								
				Reported length		Prior length		Reported and prior averaged		All months and reported averaged	
	NF days	F days		NF days	F days	NF days	F days	NF days	F days	NF days	F days
Day 9	90	13	Day 19	60	43	61	30	63	28	60	31
Day 10	78	25	Day 18	50	53	47	44	49	42	48	43
Day 11	64	39	Day 17	41	62	35	56	39	52	36	55
Day 12	56	47	Day 16	32	71	27	64	25	66	26	65
Day 13	44	59	Day 15	30	73	28	63	23	68	24	67
Day 14	35	68	Day 14	31	72	29	62	27	64	29	62
Days predicted fertile		618			618		546		546		546
Days actually fertile		251			374		319		320		323
Percentage		40.6%			60.5%		58.4%		58.6%		59.2%

Note. NF = non-fertile. F = fertile.

prior to their next cycle onset was 28.2%. The maximum percentage of ovulatory cycles reporting a positive surge on backward-counted Day 14 was 16.5%.

As a secondary test of the accuracy of the backward-counting method, we investigated how closely the backward-counted LH surge day of each cycle length estimator reflected the actual backward-counted LH surge day. We recoded the estimated backward-counted surge days according to whether they occurred one or two days prior to the actual LH surge day. Note that this method did not account for the day of the cycle on which the surge occurred; only whether the cycle length estimator accurately predicted the LH surge to occur in a window when the participant was likely to be ovulating. As shown in Table 2, no cycle length estimator was accurately associated with ovulation above 26.5%.

Six-day fertile window accuracy

As a secondary indication of indirect counting method accuracy, we examined the accuracy of the six-day fertile window procedure (Dunson et al., 1999). The fertile window is operationalized as the six days during which conception is most likely: The day of ovulation, and the five days prior (Dunson et al., 1999; Wilcox, Dunson, & Baird, 2000). Calculating the six-day window without hormonal tests still requires that ovulation be predicted by either forward- or backward-counting, but provides an estimation of fertility that is often more relevant for research purposes than the day of ovulation itself.

To estimate the effectiveness of the six-day fertile window for capturing fertility, we first determined the predicted six-day fertile windows according to the forward-counting and four backward-counting methods. We counted the LH surge day as the fifth day of the fertile window, coding the day after the LH surge and the four days prior to the LH surge as actual fertile days. We determined the percentage of fertile days in the predicted fertile windows by dividing the number of actual fertile days by the total number of days in the predicted fertile windows. We excluded the 37 women who did not report a positive LH surge during their testing phase as we could not rule out that they may have ovulated on a day outside their testing period, potentially rendering a day in the predicted six-day window fertile. The following data is thus for ovulatory cycles only.

As shown in Table 3, the maximum percentage of days in the predicted fertile windows that were actually fertile was 60.5% (backward-counting from reported length); the other methods yielded percentages ranging from 40.6–59.2%. Said in another way, if counting methods and the six-day window procedure are used to estimate fertility, a minimum of approximately 40% of the days in the six-day fertile window predicted by the counting methods are likely to be non-fertile (assuming all cycles are ovulatory). We then calculated the percentage of fertile days that occurred between actual backward-counted Days 14–19, using the actual cycle length and the confirmed (not predicted) date of the next cycle onset. Of the 570 testing days occurring between actual backward-counted Days 14–19, 349 were fertile (61.2%). Thus, if backward-counted methods and validation of the next cycle onset are used to estimate the fertile window, an almost identical percentage of days estimated to be fertile will be non-fertile (almost 40%, again assuming all cycles were ovulatory).

Counting method conception probability estimates

As a final indicator of counting method accuracy, we compared the actual and predicted conception probability according to each of the counting methods. We conducted a literature review which located a total of eight studies outlining the probability of conception resulting from sexual intercourse on days relative to ovulation (Barrett and Marshall, 1969; Bremme, 1991; Colombo and Masarotto, 2000; Schwartz et al., 1980; Schwartz et al., 1979; Vollman, 1977; Wilcox et al., 1995; Wilcox et al., 1998). These studies provided conception probability estimates for a maximum of 14 days, ranging from 8 days before to 5 days after ovulation. After removing duplicate datasets (Schwartz et al., 1980 and Wilcox et al., 1995), we averaged the six remaining estimates,[5] then assigned predicted conception probabilities to each day in the 14-day window using forward-counted or backward-counted Day 14 as the predicted day of ovulation (a table of

---

[5] The correlation between conception probabilities averaged from all eight studies versus those from the six unique datasets we averaged for our analyses was almost perfect, $r(12) = 0.997$, $p < 0.001$.

Table 4
Associations between predicted and actual conception probabilities of alternate cycle length estimators for ovulatory cycles.

Cycle length estimate	Conception probability estimates for the predicted day of ovulation			Conception probability estimates for the six-day fertile window			Correlation between actual and predicted conception probability
	M (*SD*)	Total *N*	T^a	*M* (*SD*)	Total *N*	T^b	*R*
Forward-counting method	0.106 (0.079)	103	-6.48^{***}	0.061 (0.055)	103	-14.75^{***}	0.11^{***}
Backward-counting methods							
Reported length	0.123 (0.073)	103	-4.63^{***}	0.096 (0.053)	103	-8.40^{***}	0.30^{***}
Length of prior cycle	0.126 (0.074)	91	-3.95^{***}	0.095 (0.048)	91	-8.98^{***}	0.23^{***}
Prior cycle + reported average	0.130 (0.072)	91	-3.57^{***}	0.095 (0.050)	91	-8.65^{***}	0.23^{***}
Average of all measured and reported lengths	0.127 (0.072)	91	-3.85^{***}	0.095 (0.050)	91	-8.65^{***}	0.24^{***}

[a] Test value = 0.157; the average actual conception probability on the day of ovulation (see Appendix A for values averaged).
[b] Test value = 0.140; the average conception probability of the six high-fertility cycle days (see Appendix A for values averaged).
*** $p < 0.001$.

conception probability estimates is in Appendix A).[6] We then assigned actual conception probabilities to these same days, with the LH surge day referencing the day prior to ovulation. Data from the 37 women who did not report an LH surge during their testing phase were excluded.

Table 4 shows the conception probability yielded by each counting method for the predicted day of ovulation and the predicted six-day fertile window. The average probability of conception from sexual intercourse on the day of ovulation was 0.157; one-sample *t*-tests indicated that all counting method conception probabilities were significantly lower than this critical value ($ps < 0.001$). The averaged conception probability from sexual intercourse during the six-day fertile window was 0.140; all counting methods yielded average conception probabilities that were significantly lower than this value ($ps \leq 0.001$). We then determined correlations between the actual and predicted conception probabilities across the 14-day window. We used the method outlined by Bland and Altman (1995), which yields correlation estimates that account for subject-level variation and repeated measures data. As shown in Table 4, correlations between the predicted and actual conception probabilities for the backward- and forward-counting methods were significant but relatively small ($rs = 0.11$–0.30, $ps < 0.001$).

Improving counting method accuracy by removing potential outliers

To potentially improve the accuracy of the counting methods, we examined outliers on age,[7] reported and prior cycle length, and confidence in that length. We first examined group differences between participants whose LH surges were and were not accurately predicted by each of the cycle length estimators. We ran four one-way ANOVAs (cycle length estimator accuracy: accurate [yes], non-accurate [no]) on each variable. The means and standard deviations of the accurate and non-accurate groups depending on the estimator are in Appendix A. Participants whose fertility was predicted accurately by the backward-counting method using all available cycle data were slightly older than those whose fertility was inaccurately predicted, $F(1, 81) = 7.96$, $p = 0.006$, $\eta_p^2 = 0.09$. Similarly, participants whose fertility was predicted accurately by that same method and by the backward-counting method using the prior + reported cycle average reported slightly shorter cycles, $F(1, 81) = 4.86$, $p = 0.030$, $\eta_p^2 = 0.06$, and $F(1, 81) = 5.89$, $p = 0.017$, $\eta_p^2 = 0.07$, respectively. In all cases, means between groups were within ±1 *SD* so no outliers could be removed. All other differences were not significant (all $ps \geq 0.100$, all $\eta_p^2 s < 0.04$).

We then investigated group differences on these same variables, but between participants who reported a positive LH surge versus those who did not, ignoring the cycle length estimators. Two one-way ANOVAs (LH surge detected, no LH surge) indicated that group differences on age and length confidence were non-significant ($ps \geq 0.228$, $\eta_p^2 s < 0.01$). There were significant differences between groups on both reported length, $F(1, 138) = 4.05$, $p = 0.046$, $\eta_p^2 = 0.03$, and prior length, $F(1, 124) = 3.99$, $p = 0.048$, $\eta_p^2 = 0.03$. In both cases, women who reported an LH surge had slightly longer cycles (descriptive data and analyses are in Appendix A). Because group differences were statistically significant, we next investigated whether imposing restrictions on reported length or prior length increased the ovulation prediction accuracy of the counting methods.

Improving counting method accuracy by imposing restrictions on cycle lengths

In an attempt to improve the accuracy of estimating ovulation via the counting methods, we restricted our sample to participants reporting specific ranges in cycle lengths. This allowed us to increase the homogeneity of the sample by removing cycle length outliers. In the first restriction, we limited the sample to women whose self-reported cycles lasted 25 to 33 days; in the second restriction, we limited the sample to women whose prior cycle measured 25 to 33 days (bypassing any potential errors in cycle length recall). A 25–33 day range fell within ±2 *SDs* of the mean reported length in the current sample and approximated the cycle length range in a normal population (Chiazze, 1968; Creinin et al., 2004; Fehring et al., 2006; Waller et al., 1998). We then re-ran the above analyses to determine how these length restrictions affected the accuracy of the counting methods in estimating actual fertility as verified by LH surges (again for ovulatory cycles only).

For the first restriction, the highest percentage of LH surges occurring one to two days prior to the predicted ovulation date was 31.1% (Table 5). The highest percentage of fertile days within the predicted fertile windows was 61.4%, and correlations between actual and predicted conception probabilities for the counting methods ranged from $r = 0.11$–0.35. Of the 510 testing days from ovulatory cycles occurring between actual backward-counted Days 14–19, 308 were fertile (60.4%). For the second restriction, the highest percentage of LH surges accurately detected within the two-day window was 25% (Table 5). The highest percentage of fertile days within the predicted fertile windows was 60.7%, and conception probability correlations ranged from $r = 0.12$–0.26. Of the 420 testing days from ovulatory cycles occurring

---

[6] Conception probability estimates from these six studies included both indirect (i.e., basal body temperature) and direct (hormonal) measures of ovulation. To ensure our averaged estimates were not skewed by the inclusion of conception probabilities from indirect ovulation measures, we analyzed correlations between: (1) an average of all six studies unweighted by sample size (sample size was not available for Vollman, 1977); (2) data weighted by sample size and excluding Vollman (1977); and (3) estimates yielded from direct hormone measures only (Wilcox et al., 1998). Averaged conception probabilities from all six studies correlated with weighted averaged estimates and hormonally verified estimates at $rs(12) \geq 0.90$, $ps < 0.001$. We thus retained and report unweighted conception probability estimates from all six studies so as to include all available unique data.

[7] Three women did not report their age. All were freshman completing their first degree; we estimate they were in their early 20s but counted their non-responses as missing data.

Table 5
Ovulation prediction accuracy and conception probability of alternate cycle length estimators, by cycle length restriction (ovulatory cycles only).

Cycle length estimate	Restriction 1[a]				Restriction 2[b]			
	Prediction accuracy[c]	*N*	Correlation with actual conception probability		Prediction accuracy[c]	*N*	Correlation with actual conception probability	
			R	*N*			*R*	*N*
Forward-counting method	11.1%	90	0.11^{***}	1260	10.5%	76	0.12^{***}	1064
Backward-counting methods								
Reported length	31.1%	90	0.35^{***}	1021	25.0%	76	0.26^{***}	842
Length of prior cycle	12.7%	79	0.24^{***}	875	14.5%	76	0.26^{***}	862
Prior cycle + reported average	20.3%	79	0.28^{***}	893	19.7%	76	0.24^{***}	857
Average of all measured and reported lengths	21.5%	79	0.30^{***}	893	21.1%	76	0.26^{***}	857

[a] Reported cycle length 25–33 days.
[b] Prior cycle length 25–33 days.
[c] Percentage of LH surges 1–2 days prior to ovulation as estimated by each counting method.
*** $p < 0.001$.

between actual backward-counted Days 14–19, 264 were fertile (62.9%). Appendix A contains details of these results.

Determining best practice for LH testing protocols

Despite attempts to remove outliers, we were unable to estimate ovulation or the six-day fertile window with a high degree of accuracy using the forward-counting method or four alternative backward-counting methods. Thus, we attempted to define LH testing parameters that maximized the likelihood of detecting surges in LH. The parameters were: the counting method that should be used to estimate ovulation, the number of days prior to their predicted ovulation when participants should start LH testing, and the duration of testing. Our aim was to capture ≥95% of positive LH test results with the least number of testing days and most lenient and inclusive parameters.

To investigate the most accurate method for predicting ovulation, we included the forward-counting method and three backward-counting cycle length estimators: reported length, length of prior cycle, and prior cycle + reported average. We did not include the average of all measured lengths estimator as we reasoned that recommending researchers collect at least two months of measured cycle data from participants before those participants started LH testing was impractical. The frequency distributions of LH surge days according to each method are in Fig. 1. To investigate when participants should begin testing and for how many days, we varied the number of testing days evenly on forward- or backward-counted Day 14 for seven, nine, or 11 testing days. We used these testing day ranges because they maximized the likelihood of detecting LH surges using the least number of testing days, thus minimizing researcher investment.

A frequency analysis on the percentage of LH surges captured by testing across these ranges is shown in Table 6. This analysis showed that when using LH tests and 7- to 11-day testing ranges to estimate ovulation, the percentage of accurately captured LH surges improved greatly. Two 11-day testing procedures captured >95% of the LH surges. We then restricted the sample to participants whose cycles lasted 25 to 33 days (either their self-reported length, or prior cycle length) to see if these restrictions reduced the number of testing days required.

Restricting the sample to women who self-reported their cycles lasted 25 to 33 days reduced the sample size to 90 participants reporting a positive result. Eleven of these women reported a positive result in their first cycle, excluding them from the backward-counting methods that included prior cycle length. Restricting the sample by these parameters resulted in three testing procedures capturing >95% of the positive test results, but all required 11 days of testing. Restricting the sample to

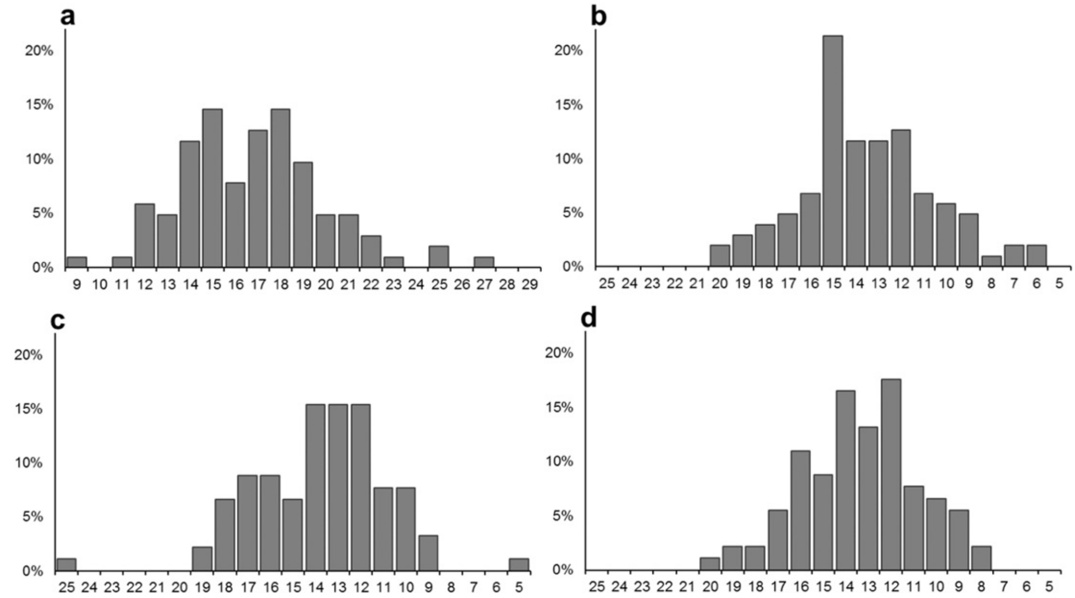


Fig. 1. Frequency distributions of detected LH surge days according to cycle method (for ovulatory cycles only). *Note.* Panel a: forward-counting method ($n = 103$). Panel b: backward-counting from reported length ($n = 103$). Panel c: backward-counting from length of prior cycle ($n = 91$). Panel d: backward-counting from prior cycle + reported average ($n = 91$).

Table 6

Percentage of positive LH surges in testing windows defined by each method, by cycle length restriction (ovulatory cycles only).

Method	No Restriction		Restriction 1[a]		Restriction 2[b]		Restriction 3[c]		Restriction 4[d]	
	% of LH surges	*N*	% of LH surges	*N*	% of LH surges	*N*	% of LH surges	*N*	% of LH surges	*N*
Forward-counting										
Days 11 to 17	58.3%	103	61.1%	90	61.8%	76	57.7%	97	57.8%	83
Days 10 to 18	72.8%	103	74.4%	90	78.9%	76	71.1%	97	74.7%	83
Days 9 to 19	83.5%	103	86.7%	90	89.5%	76	82.5%	97	86.7%	83
Backward-counting, from reported length										
11 to 17 days prior to next onset	75.7%	103	80.0%	90	75.0%	76	78.4%	97	74.7%	83
10 to 18 days prior to next onset	85.4%	103	90.0%	90	84.2%	76	88.7%	97	84.3%	83
9 to 19 days prior to next onset	93.2%	103	95.6%	90	93.4%	76	94.8%	97	92.8%	83
Backward-counting, from length of prior cycle										
11 to 17 days prior to next onset	78.0%	91	75.9%	79	82.9%	76	76.5%	85	80.7%	83
10 to 18 days prior to next onset	92.3%	91	92.4%	79	94.7%	76	91.8%	85	95.2%	83
9 to 19 days prior to next onset	97.8%	91	97.5%	79	100%	76	97.6%	85	100%	83
Backward-counting, from prior cycle + reported average										
11 to 17 days prior to next onset	80.2%	91	83.5%	79	82.9%	76	82.4%	85	83.1%	83
10 to 18 days prior to next onset	89.0%	91	89.9%	79	90.8%	76	90.6%	85	90.4%	83
9 to 19 days prior to next onset	96.7%	91	96.2%	79	100%	76	96.5%	85	98.8%	83

[a] Reported cycle length 25–33 days.

[b] Prior cycle length 25–33 days.

[c] Reported cycle length 25–35 days.

[d] Prior cycle length 25–35 days.

women whose prior cycle measured 25 to 33 days reduced the sample size to 76 participants reporting a positive result, all of whom tested in C2, C3, or C4. Two testing procedures captured >95% of the positive test results, both of which required eleven days of LH testing. One 9-day testing procedures reached 94.7% accuracy (the 9-day window using backward-counting from prior length), details are in Table 6.

Because we were able to capture 95–100% of the positive test results with multiple methods, we investigated whether we could relax the cycle length restriction, thereby increasing the sample size. We aimed to still capture >95% of the LH surges but reduce the number of testing days required. We thus retained participants whose reported and prior cycle lengths measured between 25 and 35 days. Women with 25–35 day cycles are less likely to report anovulation (Harlow et al., 2013), and this range mirrors restrictions used in past literature (Cantú et al., 2014; Fales et al., 2014; Haselton et al., 2007; Larson et al., 2012; Miller and Maner, 2011). As shown in Table 6, five methods captured >95% of the LH surges. One testing procedure required nine testing days, but only when the restriction was based on prior and not reported length. If researchers restrict the sample to women whose prior cycles lasted between 25 and 35 days, then use the prior cycle length to predict the next menses onset, instructing participants to test 10 to 18 days prior to this date will capture 95.2% of positive LH surge results. Similar accuracy percentages (98.8–100%) are also obtainable in two other testing ranges (9 to 19 days prior to the next cycle onset as predicted by the length of the prior cycle; and 9 to 19 days prior to the next cycle onset as predicted by the average of the prior and reported cycle lengths). These latter options, however, require two additional days of LH testing.

Discussion

Experts are divided on whether women's cognition and behavior differs between fertile and non-fertile phases of the menstrual cycle (Gangestad and Haselton, 2015; Gildersleeve et al., 2013; Gildersleeve et al., 2014a, 2014b; Harris, 2013; Harris et al., 2013, 2014; Wood, 2015; Wood and Carden, 2014; Wood et al., 2014). Critics of this literature highlight the indirect, imprecise, and flexible methodologies employed to determine women's fertility between studies (Harris et al., 2013; Wood, 2015; Wood et al., 2014). To resolve this problem, two recent studies have modeled the efficacy of indirect counting methods and between-subjects designs commonly used in past research, concluding that within-subjects designs with hormonally validated measures of fertility are the most accurate for testing effects at high and low fertility (Gonzales and Ferrer, 2015; Gangestad et al., 2015). In the current study, we contribute to the advancement of improved methodology by providing a data-driven and practical method of best practices that could and should be adopted by researchers. The guidelines are ideal for researchers within university settings who may have limited time and funding.

Our analysis of the accuracy of the popular backward- and forward-counting methods for predicting the day of ovulation revealed that neither method was associated with surges in LH (which precede ovulation by approximately 24–48 h) with >30% accuracy. This poor level of accuracy was retained when using a lenient window for the day of ovulation, across alternative estimators of the onset of the next cycle, and when removing outliers to increase the homogeneity of the sample. At best, only 60.5% of days in the six-day fertile windows predicted by the counting methods were in fact fertile, and correlations between actual and predicted conception probability from the counting methods were weak to moderate. We could not obtain acceptable accuracy rates by relying on self-reported menstrual cycle lengths. We further note that these poor results were despite excluding all non-ovulatory cycle data ($n = 37$; 26.4% of the sample). Should these data have been included in our calculations, the accuracy of the counting methods would have been substantially less reliable.

These findings suggest that previous research predicting fertility via counting methods and self-report are unlikely to have accurately captured fertility status. We strongly recommend that all future research testing effects of high and low fertility on women's behavior should therefore include direct measures of ovulation, such as LH tests, estradiol assays, and progesterone assays. Methods sampling ovarian hormones regularly throughout the cycle (e.g., Roney and Simmons, 2013) can also approximate fertility status and are fruitful avenues for future research. A proposed alternative is to retain the use of counting methods but recruit very large sample sizes to achieve sufficient statistical power ($ns > 500$; Gangestad et al., 2015). However, given that our data demonstrate that counting methods yield poor accuracy rates across multiple indicators, we recommend direct measures of ovulation be utilized instead.

We provide pragmatic and parsimonious sampling guidelines that will support researchers to accurately capture a high percentage of women's positive LH surges. Specifically, we recommend: 1) restricting the sample to participants whose prior cycles lasted between 25 and 35 days, and 2) instructing participants to complete daily LH tests 10 to 18 days prior to their next predicted menses (using the prior cycle length to predict this onset date). Doing so will maximize the likelihood of detecting surges in LH using the least number of LH testing days, increasing confidence that participants are fertile and not fertile when they are supposed to be. An alternative is to measure LH on every day of the cycle, thereby theoretically increasing the likelihood of capturing the LH surge to 100%, but such practices are rarely practical for research purposes. Researchers may also opt to use the 11-day testing ranges outlined in current study with or without cycle length restrictions, predicting the onset of the next cycle using the prior cycle length or the prior cycle + reported length average. However, we recommend imposing some restriction on cycle length as long and short cycles may indicate anovulation (Harlow et al., 2013). A more detailed summary of our recommendations are available in Appendix A.

Our findings have notable implications for researchers testing fertility effects on women's cognition, affect, and behavior. Because distributions in LH surge timing vary considerably, prior research predicting ovulation using self-report and counting methods is highly likely to have misclassified some women as fertile when they were not fertile. Such misclassifications reduce the likelihood of detecting a true effect of fertility on dependent variables. Consistent with recent findings in this area (Gangestad et al., 2015; Gonzales and Ferrer, 2015), our findings indicate that if counting methods are used to estimate fertility, very large sample sizes, highly reliable outcome variables, and more lenient windows of fertility would be required reach adequate statistical power and detect a true effect, should one exist. Similarly, assuming that there is no large repository of file drawer papers, the current research suggests that effect sizes in studies using counting and self-report methods are probably underestimates of the true effect size.

We believe that future work using our protocols will allow the scientific community to ascertain with greater accuracy whether or not fertility influences cognition, emotion, and behavior. To this end, we provide all the protocols, training for the LH kits, all data used in our analyses and some additional recommendations for conducting research requiring comparisons between women at high and low fertility phases of the menstrual cycles in the electronic supplementary material on the Open Science Framework (osf.io/5nc8k). Our guidelines provide practical, cost-effective protocols for researchers seeking to undertake direct replications and future research on the relationship between women's fertility and important psychological outcomes.

Acknowledgements

The research reported here and preparation of this manuscript were supported by a *Future Fellowship* (FT140100291) from the Australian Research Council to the last author. Barnaby Dixson was supported by a University of Queensland Postdoctoral Fellowship. Khandis Blake was supported by an Australian Postgraduate Award.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.yhbeh.2016.03.004.

References

Baird, D.D., McConnaughey, D.R., Weinberg, C.R., Musey, P.I., Collins, D.C., Kesner, J.S., ... Wilcox, A.J., 1995. Application of a method for estimating day of ovulation using urinary estrogen and progesterone metabolites. Epidemiology 6 (5), 547–550.

Barrett, J.C., Marshall, J., 1969. The risk of conception on different days of the menstrual cycle. Popul. Stud. 23 (3), 455–461.

Bland, J.M., Altman, D.G., 1995. Calculating correlation coefficients with repeated observations: part 1—correlation within subjects. Br. Med. J. 310 (18), 446.

Bremme, J., 1991. Sexualverhalten und Konzeptionswahrscheinlïchkeit (Auswertung einer prospektiven Studie zur Natürlichen Familienplanung) (Thesis) Medizinische Fakultät der Heinrich, Heine Universität, Düsseldorf, Germany.

Cantú, S.M., Simpson, J.A., Griskevicius, V., Weisberg, Y.J., Durante, K.M., Beal, D.J., 2014. Fertile and selectively flirty: women's behavior toward men changes across the ovulatory cycle. Psychol. Sci. 25 (2), 431–438.

Chiazze, L., 1968. The length and variability of the human menstrual cycle. JAMA 203 (6), 377–380.

Chrisler, J., Caplan, P., 2002. The strange case of Dr. Jekyll and Ms. Hyde: how PMS become a cultural phenomenon and a psychiatric disorder. Annu. Rev. Sex Res. 13 (1), 274–306.

Collins, W., Jurkovic, D., Bourne, T., Kurjak, A., Campbell, S., 1991. Ovarian morphology, endocrine function and intra-follicular blood flow during the peri-ovulatory period. Hum. Reprod. 6 (3), 319–324.

Colombo, B., Masarotto, G., 2000. Daily fecundability. Demogr. Res. 3 (5), 1–39.

Creinin, M.D., Keverline, S., Meyn, L.A., 2004. How regular is regular? An analysis of menstrual cycle regularity. Contraception 70 (4), 289–292.

Direito, A., Bailly, S., Mariani, A., Ecochard, R., 2013. Relationships between the luteinizing hormone surge and other characteristics of the menstrual cycle in normally ovulating women. Fertil. Steril. 99 (1), 279–285.

Dixon, G.W., 1980. Ethinyl estradiol and conjugated estrogens as postcoital contraceptives. JAMA 244 (12), 1336–1339.

Dixson, A.F., 2009. Sexual Selection and the Origins of Human Mating Systems. Oxford University Press, Oxford, UK.

Dixson, A.F., 2012. Primate Sexuality. Oxford University Press, Oxford, UK.

Dixson, B.J., Brooks, R.C., 2013. The role of facial hair in women's perceptions of men's attractiveness, health, masculinity and parenting abilities. Evol. Hum. Behav. 34, 236–241.

Dixson, B.J., Tam, J., Awasthy, M., 2013. Do women's preferences for men's facial hair change with reproductive status? Behav. Ecol. 24 (3), 708–716.

Dunson, D., Baird, D., Wilcox, A., Weinberg, C., 1999. Day-specific probabilities of clinical pregnancy based on two studies with imperfect measures of ovulation. Hum. Reprod. 14 (7), 1835–1839.

Ecochard, R., Gougeon, A., 2000. Side of ovulation and cycle characteristics in normally fertile women. Hum. Reprod. 15 (4), 752–755.

Fales, M., Gildersleeve, K., Haselton, M., 2014. Exposure to perceived male rivals raises men's testosterone on fertile relative to nonfertile days of their partner's ovulatory cycle. Horm. Behav. 65 (5), 454–460.

Feder, H.H., 1981. Estrous cyclicity in mammals. In: Adler, N.T. (Ed.), Neuroendocrinology of Reproduction. Plenum Press, New York, pp. 279–348.

Fehring, R.J., Schneider, M., Raviele, K., 2006. Variability in the phases of the menstrual cycle. J. Obstet. Gynecol. Neonatal. Nurs. 35 (3), 376–384.

Gangestad, S., 2015. Comment: Wood et al.'s (2014) speculations of inappropriate research practices in ovulatory cycle studies. Emot. Rev.

Gangestad, S., Haselton, M., 2015. Human estrus: implications for relationship science. Curr. Opin. Psychol. 1, 45–51.

Gangestad, S.W., Haselton, M.G., Welling, L.L., Gildersleeve, K., Pillsworth, E.G., Burriss, R.P., ... Puts, D.A., 2015. How valid are assessments of conception probability in ovulatory cycle research?: evaluations, recommendations, and theoretical implications. Evol. Hum. Behav.

Gildersleeve, K., DeBruine, L., Haselton, M., Frederick, D., Penton-Voak, I., Jones, B., Perrett, D., 2013. Shifts in women's mate preference across the ovulatory cycle: a critique of Harris (2011) and Harris (2012). Sex Roles 69 (9–10), 516–524.

Gildersleeve, K., Haselton, M., Fales, M., 2014a. Do women's mate preferences change across the ovulatory cycle? A meta-analytic review. Psychol. Bull. 140 (5), 1205–1259.

Gildersleeve, K., Haselton, M., Fales, M., 2014b. Meta-analyses and p-curves support robust cycle shifts in women's mate preferences: reply to Wood and Carden (2014) and Harris, Pashler, and Mickes (2014). Psychol. Bull. 140 (5), 1272–1280.

Gonzales, J.E., Ferrer, E., 2015. Efficacy of methods for ovulation estimation and their effect on the statistical detection of ovulation-linked behavioral fluctuations. Behav. Res. Methods.

Guermandi, E., Vegetti, W., Bianchi, M.M., Uglietti, A., Ragni, G., Crosignani, P., 2001. Reliability of ovulation tests in infertile women. Obstet. Gynecol. 97 (1), 92–96.

Guida, M., Tommaselli, G., Palomba, S., Pellicano, M., Moccia, G., Di Carlo, C., Nappi, C., 1999. Efficacy of methods for determining ovulation in a natural family planning program. Fertil. Steril. 72 (5), 900–904.

Harlow, S.D., Windham, G.C., Paramsothy, P., 2013. Menstruation and menstrual disorders. In: Goldman, M.B., Troisi, R., Rexrode, K.M. (Eds.), Women and Health, second ed. Academic Press, Oxford, UK, pp. 163–177.

Harris, C.R., 2011. Menstrual cycle and facial preferences reconsidered. Sex Roles 64 (9–10), 669–681.

Harris, C., 2013. Shifts in masculinity preferences across the menstrual cycle: Still not there. Sex Roles 69 (9–10), 507–515.

Harris, C., Chabot, A., Mickes, L., 2013. Shifts in methodology and theory in menstrual cycle research on attraction. Sex Roles 69 (9–10), 525–535.

Harris, C., Pashler, H., Mickes, L., 2014. Elastic analysis procedures: an incurable (but preventable) problem in the fertility effect literature. Comment on Gildersleeve, Haselton, and Fales (2014). Psychol. Bull. 140 (5), 1260–1264.

Haselton, M.G., Mortezaie, M., Pillsworth, E.G., Bleske-Rechek, A., Frederick, D.A., 2007. Ovulatory shifts in human female ornamentation: near ovulation, women dress to impress. Horm. Behav. 51 (1), 40–45.

Havlíček, J., Cobey, K.D., Barrett, L., Klapilová, K., Roberts, S.C., 2015. The spandrels of Santa Barbara? A new perspective on the peri-ovulation paradigm. Behav. Ecol.

Heape, W., 1900. The sexual season of mammals and the relation of the “pro-oestrum” to menstruation. Q. J. Microsc. Sci. 44, 1–70.

Jones, B.C., 2014. Comment: alternatives to Wood et al.'s conclusions. Emot. Rev. 6 (3), 254–256.

Jukic, A.M.Z., Weinberg, C.R., Wilcox, A.J., McConnaughey, D.R., Hornsby, P., Baird, D.D., 2008. Accuracy of reporting of menstrual cycle length. Am. J. Epidemiol. 167 (1), 25–33.

Larson, C.M., Pillsworth, E.G., Haselton, M.G., 2012. Ovulatory shifts in women's attractions to primary partners and other men: further evidence of the importance of primary partner sexual attractiveness. PLoS One 7 (9), 1–10.

Miller, S.L., Maner, J.K., 2011. Ovulation as a male mating prime: subtle signs of women's fertility influence men's mating cognition and behavior. J. Pers. Soc. Psychol. 100 (2), 295–308.

Pearlstone, A.C., Surrey, E.S., 1994. The temporal relation between the urine LH surge and sonographic evidence of ovulation: determinants and clinical significance. Obstet. Gynecol. 83 (2), 184–188.

Puts, D.A., Bailey, D.H., Cárdenas, R.A., Burriss, R.P., Welling, L.L.M., Wheatley, J.R., Dawood, K., 2013. Women's attractiveness changes with estradiol and progesterone across the ovulatory cycle. Horm. Behav. 63 (1), 13–19.

Roney, J.R., Simmons, Z.L., 2013. Hormonal predictors of sexual motivation in natural menstrual cycles. Horm. Behav. 63 (4), 636–645.

Schwartz, D., Mayaux, M.J., Martin-Boyce, A., Czyglik, F., David, G., 1979. Donor insemination: conception rate according to cycle day in a series of 821 cycles with a single insemination. Fertil. Steril. 31 (2), 226–229.

Schwartz, D., Macdonald, P.D., Heuchel, V., 1980. Fecundability, coital frequency and the viability of Ova. Popul. Stud. 34 (2), 397–400.

Small, C., Manatunga, A., Marcus, M., 2007. Validity of self-reported menstrual cycle length. Ann. Epidemiol. 17 (3), 163–170.

Symons, D., 1979. The Evolution of Human Sexuality. Oxford University Press, New York.

Testart, J., Frydman, R., 1982. Minimum time lapse between luteinizing hormone surge or human chorionic gonadotropin administration and follicular rupture. Fertil. Steril. 37 (1), 50–53.

Testart, J., Frydman, R., Feinstein, M.C., Thebault, A., Roger, M., Scholler, R., 1981. Interpretation of plasma luteinizing hormone assay for the collection of mature oocytes from women: definition of a luteinizing hormone surge-initiating rise. Fertil. Steril. 36 (1), 50–54.

Thornhill, R., Gangestad, S., 2008. The Evolutionary Biology of Human Female Sexuality. Oxford University Press, Oxford, UK.

Vollman, R.F., 1977. Assessment of the fertile and sterile phases of the menstrual cycle. Int. Rev. Nat. Fam. Plan. 1 (1), 40–47.

Wagenmakers, E., Wetzels, R., Borsboom, D., van der Maas, H., Kievit, R., 2012. An agenda for purely confirmatory research. Perspect. Psychol. Sci. 7 (6), 632–638.

Wallen, K., 2001. Sex and context: hormones and primate sexual motivation. Horm. Behav. 40 (2), 339–357.

Waller, K., Swan, S., Windham, G., Fenster, L., Elkin, E., Lasley, B., 1998. Use of urine biomarkers to evaluate menstrual function in healthy premenopausal women. Am. J. Epidemiol. 174 (11), 1071–1080.

Wilcox, A.J., Weinberg, C., Baird, D.D., 1995. 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. 333 (23), 1517–1521.

Wilcox, A.J., Weinberg, C.R., Baird, D.D., 1998. Post-ovulatory ageing of the human oocyte and embryo failure. Hum. Reprod. 13 (2), 394–397.

Wilcox, A.J., Dunson, D.B., Baird, D.D., 2000. The timing of the “fertile window” in the menstrual cycle: day specific estimates from a prospective study. Br. Med. J. 321 (7271), 1259–1262.

Wood, W., 2015. Reply to Gangestad's (2015) comment on Wood, Kressel, Joshi, and Louie (2014). Emot. Rev.

Wood, W., Carden, L., 2014. Elusiveness of menstrual cycle effects on mate preferences: comment on Gildersleeve, Haselton, and Fales (2014). Psychol. Bull. 140 (5), 1265–1271.

Wood, W., Kressel, L., Joshi, P., Louie, B., 2014. Meta-analysis of menstrual cycle effects on women's mate preferences. Emot. Rev. 6 (3), 229–249.

Wu, S., 2015. Despite Accepted Lore, Women's Mate Preferences Don't Shift According to Fertility, Menstrual Cycle, New Study Claims. (Retrieved from http://medicalxpress.com/news/2014-05-lore-women-dont-shift-fertility.html).

Zietsch, B.P., Lee, A.J., Sherlock, J.M., Jern, P., 2015. Variation in women's facial masculinity preference is better explained by genetic differences than by previously identified context-dependent effects. Psychol. Sci.