Investigating the Ovulatory Cycle: An Overview of Research and Methods

Lisa L. M. Welling and Robert P. Burriss
The Oxford Handbook of Evolutionary Psychology and Behavioral Endocrinology
Edited by Lisa L. M. Welling and Todd K. Shackelford

Abstract and Keywords

An abundant amount of research into the human ovulatory cycle and related adaptive shifts in preferences and behaviors has been published in recent decades. Evidence suggests that fertility in women is accompanied by increased preferences for male traits that putatively signal underlying genetic quality, adaptive increases in sexual motivation, and changes in attractiveness. Yet, these supposed adaptive shifts remain controversial and disputed, while methods across studies have been inconsistent. In this chapter, we review the research on phenotypic variation across the human menstrual cycle, focusing on the related areas of preferences for male traits, sexual behavior and motivation, and cues to ovulation. Next, we consider the various methods currently used by researchers to ascertain conception risk and review recently published recommendations intended to guide future research and facilitate comparison across studies.

Keywords: menstrual cycle, fertility, conception risk, mate preferences, mating behavior, attractiveness

Since the 1990s, the study of phenotypic variation across the human ovulatory cycle has expanded greatly. Researchers have investigated cyclical effects on women’s sexual desires, preferences, mate choice, intrasexual competitiveness, physical appearance, and partner mate retention behavior, among other constructs (reviewed by Gangestad & Thornhill, 2008; Gangestad et al., 2016). However, as recently noted by Gangestad et al. (2016), the methods used by researchers to assess women’s fertility are inconsistent, which hinders comparison across studies and led critics to question the validity and robustness of findings. In this chapter, we review the evidence for phenotypic variation across the human ovulatory cycle in three interconnected areas: preferences for male traits, sexual behavior and motivation, and cues to ovulation. We then consider the
various methods currently used by researchers to ascertain fertility status and review recently published recommendations.

**Preferences for Male Traits**

The ovulatory shift hypothesis predicts systematic variation in androphilic women’s behavior, specifically behaviors related to mating and mate preferences, over the course of the ovulatory cycle (e.g., Gangestad & Thornhill, 1998; Grammer, 1993; Thornhill & Gangestad, 1999). The likelihood of a woman conceiving after a single act of intercourse (termed her conception risk) varies over her ovulatory cycle, reaching a peak in the days preceding ovulation. As a woman’s conception risk peaks, her preferences for male traits that indicate good genes should likewise increase (Gangestad & Thornhill, 1998). Preferences are not expected to remain constant because men who possess good genes do not necessarily offer other potential benefits, such as resources and investment (Perrett et al., 1998). In fact, men who putatively possess these good genes may be less attractive as long-term mates because they invest fewer resources in their partners (e.g., Penton-Voak & Perrett, 2001; Perrett et al., 1998). Preference shifts are especially pronounced when women judge men for short-term, sexual relationships, as opposed to long-term, committed relationships (e.g., Gangestad, Simpson, Cousins, Garver-Apgar, & Christensen, 2004; Little & Jones, 2012; Penton-Voak et al., 1999; Puts, 2005), which is consistent with the hypothesis that these patterns reflect evolved strategies for obtaining good genes for offspring (Gangestad & Thornhill, 1998). Alternative functions, such as securing physical protection (Thornhill & Gangestad, 2008) or increasing the probability of conception by copulating with a fertile mate (Puts, 2006), are also consistent with these findings.

Many studies have demonstrated fertile phase increases in women’s preferences for men’s traits, including facial masculinity (V. S. Johnston, Hagel, Franklin, Fink, & Grammer, 2001; Jones, Little, et al., 2005; Little & Jones, 2012; Little, Jones, & DeBruine, 2008; Penton-Voak & Perrett, 2000; Penton-Voak et al., 1999; Roney, Simmons, & Gray, 2011; Welling et al., 2007; but see Jones et al., 2018; Marcinkowska et al., 2016; Marcinkowska, Galbarczyk, & Jasienska, 2018), facial symmetry (Little & Jones, 2012; Little, Jones, Burt, & Perrett, 2007; but see Cárdenas & Harris, 2007; Koehler, Rhodes, Simmons, & Zebrowitz, 2006; Marcinkowska, Galbarczyk, & Jasienska, 2018), the faces of bodily symmetrical men (Thornhill & Gangestad, 2003), darker facial skin color (Frost, 1994), masculine body shape (Little, Jones, & Burriss, 2007; but see Peters, Simmons, & Rhodes, 2009; Jünger, Kordsmeyer, Gerlach, & Penke, 2018; Marcinkowska, Galbarczyk, & Jasienska, 2018), voices with masculine characteristics (lower fundamental frequency and formant dispersion; Feinberg et al., 2006; Puts, 2005; but see Jünger, Motta-Mena, et al., 2018), greater height (Pawłowski & Jasienska, 2005), and the odor of men who are more dominant, symmetrical, and heterozygous at major histocompatibility complex (MHC) loci (Gangestad & Thornhill, 1998; Havliček, Roberts, & Flegr, 2005; Rikowski & Grammer, 1999; Thornhill & Gangestad, 1999; Thornhill et al., 2003). Although the ovulatory shift hypothesis predicts that women should be choosier at peak conception
risk, some studies have documented that at periovulation women’s ratings of male attractiveness generally increase, including ratings of the face (Danel & Pawlowski, 2006), body (Jünger, Kordsmeyer, et al., 2018), and voice (Jünger, Motta-Mena, et al., 2018). Furthermore, women pay more attention to attractive men (Anderson et al., 2010), and their accuracy at classifying faces as male is greatest at peak fertility (Macrae, Alnwick, Milne, & Schloerscheidt, 2002), particularly when those faces are masculine (L. Johnston, Miles, & Macrae, 2008). That said, the sexual relevance of the target face is important, with lesbian women categorizing female faces more accurately at peak fertility (Brinsmead-Stockham, Johnston, Miles, & Macrae, 2008).

Nevertheless, the validity of cyclic shifts in women’s preferences, particularly relating to masculinity preferences, has been questioned. A recent meta-analysis by Gildersleeve, Haselton, and Fales (2014a) found evidence for robust cyclic shifts specific to women’s preferences for hypothesized cues of male genetic quality, at least when men were evaluated for a short-term (i.e., purely sexual) relationship. However, the high number of studies included in that meta-analysis that produced null findings (60 percent) led others to argue that the few significant findings are research artifacts (i.e., publication bias; Wood, Kressel, Joshi, & Louie, 2014) or the result of “p-hacking,” whereby researchers adjust the parameters of their analysis to produce a desired result (Harris, Pashler, & Mickes, 2014). To address these concerns, Gildersleeve, Haselton, and Fales (2014b) constructed p curves of significant findings from their original meta-analytic data, finding that, consistent with their original report of robust cyclic shifts, the p curves were right-skewed and had a large number of highly significant p values (p < .01), which is characteristic of a nonzero true effect (Simonsohn, Nelson, & Simmons, 2014). More recently, in the largest study to date (n = 584) of cyclic shifts in women’s preferences for masculine male faces, Jones et al. (2018) used a within-subjects design and found no evidence that preferences for facial masculinity are related to cyclic variation in women’s steroid hormone levels. This suggests that menstrual cycle shifts, at least in terms of masculinity preference, are not as robust as previously supposed, but future work of this scale (i.e., large within-subject designs) should be conducted to assess preferences for other physical traits.

Preferences for nonphysical traits, such as dominant and intrasexually competitive behavior (Gangestad, Garver-Apgar, Simpson, & Cousins, 2007; Gangestad, Simpson, Cousins, Garver-Apgar, & Christensen, 2004; Lukaszewski & Roney, 2009), courtship language (Rosen & López, 2009), and creativity/intelligence (Haselton & Miller, 2006; but see Gangestad, Garver-Apgar, Simpson, & Cousins, 2007; Prokosch, Coss, Scheib, & Blozis, 2009), are also higher at peak fertility than at times when conception risk is lower. Guéguen (2009a, 2009b) showed in two field studies that women may be more receptive to courtship behavior at peak fertility. In these two studies, women were more likely to agree to a man’s request to dance or to share phone numbers if they were in the late follicular (fertile) phase of their cycle (the phase directly preceding ovulation), compared with the luteal phase (the phase following ovulation, when fertility is lower) and menses. Other cyclic shifts have also been detected at other points in the cycle. During the luteal phase, preferences increase for traits such as apparent health (Jones,
Investigating the Ovulatory Cycle: An Overview of Research and Methods

Little, et al., 2005; Jones, Perrett, et al., 2005) and self-resemblance (DeBruine, Jones, & Perrett, 2005; Holzleitner et al., 2017), suggesting a motivation to promote affiliation with healthy and related individuals prior to and during pregnancy. Correspondingly, increases in progesterone across the cycle are associated with increased sensitivity to social information (Maner & Miller, 2014) and the salience of emotional displays indicating danger (Conway et al., 2007). However, preferences for putative signals of investment have not been shown to increase during this phase (Gangestad et al., 2007), suggesting that these traits remain consistently attractive.

Sexual Behavior and Motivation

Various cyclical effects are nullified when women judge men for a long-term rather than a short-term relationship (Gangestad et al., 2007; Haselton & Miller, 2006; V. S. Johnston et al., 2001; Jones, Perrett, et al., 2005; Little, Jones, & Burriss, 2007; Pawłowski & Jasienska, 2005; Penton-Voak et al., 1999; Puts, 2005; Rosen & López, 2009). This may be because women trade off good genes against material benefits by securing as long-term partners men who provision, while engaging men whose traits signify good genes as short-term or extra-pair copulation partners (Gangestad & Simpson, 2000). In support of this hypothesis is evidence that women at high conception risk report greater sexual desire (Roney & Simmons, 2013), are more interested in extra-pair men (Gangestad, Thornhill, & Garver, 2002), are less motivated toward sex for the purposes of intimacy (Sheldon, Cooper, Geary, Hoard, & DeSoto, 2006), are more sexually opportunistic (Gangestad, Thornhill, & Garver-Apgar, 2010), selectively flirt more with men who possess markers of genetic fitness (Cantú et al., 2014), are more likely to visit a singles nightclub without their primary partner (Grammer, Jutte, & Fischmann, 1997), and show greater interest in extra-pair men if their partners are less attractive (Haselton & Gangestad, 2006; Larson, Pillsworth, & Haselton, 2012; Pillsworth & Haselton, 2006; see also Meltzer, 2017, for findings related to masculinity and marital satisfaction), are less symmetrical (Gangestad, Thornhill, & Garver-Apgar, 2005), and have MHC alleles that do not complement their own (Garver-Apgar, Gangestad, Thornhill, Miller, & Olp, 2006).

Although there is some evidence that women’s interest in their partners (as indicated by pupil dilation) is greatest at ovulation, a similar pattern is observed for responses to attractive opposite-sex celebrities (Laeng & Falkenberg, 2007). Furthermore, sexual interest and number of sexual encounters also peak during the late follicular phase in lesbians (Burleson, Trevathan, & Gregory, 2002), suggesting that cyclical effects on motivation and behavior may not be related to fear of pregnancy, proximity to men, or a reaction to male sexual advances.

Women may be more competitive at peak fertility (e.g., Pearson & Schipper, 2013). Financially, women are more likely to discount the future by preferring smaller rewards in the present over larger rewards later (Lucas & Koff, 2017), suggesting increased impulsivity, although they do so less after viewing images of attractive males (Kaighobadi & Stevens, 2013). Women are less cooperative and more likely to punish other women when playing an ultimatum game during peak fertility (Eisenbruch & Roney, 2016; Lucas,
Koff, & Skeath, 2007), especially if their opponent is attractive (Lucas & Koff, 2013). Perhaps relatedly, competition with same-sex rivals for access to desirable men also seems to increase with conception risk, with women exhibiting two main strategies for competing with same-sex rivals for male attention: derogation of female competitors and appearance enhancement (Buss & Schmitt, 1996; although it should be noted that men also use these intrasexually competitive strategies, see Fisher & Cox, 2010). Despite rating their own attractiveness more highly (Röder, Brewer, & Fink, 2009; Schwarz & Hassebrauck, 2008), women derogate competitors by judging their facial attractiveness more harshly (Fisher, 2004; see also Piccoli, Foroni, & Carnaghi, 2013; Vukovic et al., 2009; Welling et al., 2007) when maximally fertile. Exposure to this derogation in turn causes men to lower their attractiveness judgments of the derogating females’ rivals (Fisher & Cox, 2009), demonstrating the effectiveness of this tactic. Spending patterns also appear to reflect intrasexual competitiveness: women at peak fertility are more willing to spend money on and choose products that enhance appearance (Durante, Griskevicius, Hill, Perilloux, & Li, 2011; Hill & Durante, 2009), although the total amount of money they spend remains the same (Röder et al., 2009).

Additionally, women’s aversion to behaviors that may be sexually maladaptive or dangerous is greater at peak fertility, with increases in ratings of disgust toward incest (Fessler & Navarrete, 2003) and decreases in behaviors that may increase the risk of rape (Bröder & Hohmann, 2003; but see Fessler, 2003). Attention from unwanted suitors around peak conception risk could be tremendously detrimental because rape at this time would be more likely to lead to conception (i.e., would be more reproductively costly), and so ancestral women may have evolved “rape avoidance” strategies that they use more often during peak fertility than at other times. Indeed, fewer women are raped near ovulation than would be expected by chance and women engage in fewer behaviors that may put them at risk of sexual assault (e.g., walking alone in a dimly lit area, going on a first date, going out to a bar) when they are near ovulation, despite being more physically active in general (see Chavanne & Gallup, 1998). Women are more assertive during times characterized by high estradiol and low progesterone (i.e., the late follicular phase; Blake, Bastian, O’Dean, & Denson, 2017) and perform better on a measure of physical strength after reading a sexual assault scenario near ovulation (Petralia & Gallup, 2002), which could reflect adaptive shifts designed to thwart an attacker when conception is most likely. Furthermore, women show greater bias against out-group men near ovulation (Navarrete, Fessler, Fleischman, & Geyer, 2009), and this bias may depend on the extent to which women associate out-group men with physical formidability (i.e., they report a greater preference for familiar men when fertility is high, particularly if they perceive unfamiliar men as being more physically intimidating; McDonald, Asher, Kerr, & Navarrete, 2011). Correspondingly, female gait has been shown to be more attractive during the luteal phase than during the follicular phase (Provost, Quinsey, & Troje, 2008), which the authors argue may serve to prevent fertility from being widely advertised, as changes in gait are likely to be easier to detect at a distance than are changes in face shape or odor. If gait were to be more attractive at peak fertility, male attention may be drawn indiscriminately, which may bring unwanted male attention. These inhibitory
effects on motivation may help to ensure that women less frequently perform behaviors that can lead to disadvantageous pairings or ill-timed/unwanted pregnancies.

**Cues to Ovulation**

The long-held assumption that human females have lost estrus has been challenged by suggestions that there may be physical and behavioral cues to a woman’s cycle phase (for reviews, see Haselton & Gildersleeve, 2011; Thornhill & Gangestad, 2008). At peak fertility, women are more attractive in terms of facial appearance (Puts et al., 2013; Roberts et al., 2004), odor (Gildersleeve, Haselton, Larson, & Pillsworth, 2012; Havlíček, Dvořáková, Bartoš, & Flegr, 2006; Kuukasjärvi et al., 2004; S. L. Miller & Maner, 2011; Singh & Bronstad, 2001), and voice (Bryant & Haselton, 2009; Pipitone & Gallup, 2008; but see Fischer et al., 2011). In fact, the scent of an ovulating woman has a direct endocrinological influence on men: men who smell T-shirts worn by women near ovulation, but not worn by women at other cycle phases or T-shirts not worn by anyone, experience an increase in their circulating testosterone levels (S. L. Miller & Maner, 2010), which is a response associated with sexual arousal (e.g., Stoleru, Ennaji, Cournot, & Spira, 1993). Women also have greater breast symmetry (Manning, Scutt, Whitehouse, Leinster, & Walton, 1996; Scutt & Manning, 1996), are perceived as more attractive by their romantic partners (Cobey, Buunk, Pollet, Kippling, & Roberts, 2013), and modulate their appearance and clothing to enhance attractiveness (Durante, Li, & Haselton, 2008; Haselton, Mortezaie, Pillsworth, Bleske-Rechek, & Frederick, 2007; Hill & Durante, 2009; Röder, Brewer, & Fink, 2009; Schwarz & Hassebrauck, 2008; see also Meltzer, McNulty, Miller, & Baker, 2015, for findings related to women’s weight loss goals) when they are most fertile. G. Miller, Tyber, and Jordan (2007) investigated earnings through tips among female lap dancers as a function of menstrual cycle and contraceptive status. They found that lap dancers receive larger tips at peak fertility compared to when they were in nonfertile phases (menstruation and the luteal phase) of their cycle, whereas contraceptive pill users showed no such variation. It is not clear which trait(s) drive this pattern. One possibility is that men are responding to dynamic cues, such as change in a woman’s scent (Thornhill et al., 2003), flirtation (Cantú et al., 2014), or increased sexual opportunism (Gangestad et al., 2010), which are known to vary over the cycle (discussed earlier).

Further support for the existence of perceivable cues to ovulation comes from studies of the effects of a woman’s cycle phase on other persons. If selection has acted on women to promote late-follicular-phase (p. 113) copulations with high-quality and compatible men, one may also expect it to have acted on their long-term partners to decrease the likelihood of infidelity and/or poaching. Three studies have suggested that women at peak conception risk perceive their partners to be more proprietary and attentive (Gangestad et al., 2002; Haselton & Gangestad, 2006; Pillsworth & Haselton, 2006), and one has shown that men’s ratings of the dominant appearance of other men increase when their partners are fertile (Burriss & Little, 2006). Moreover, men whose proprietary behaviors increase most during their partners’ fertile phase have partners whose interest in extra-
pair copulation increases most during peak fertility (Gangestad et al., 2002; Haselton & Gangestad, 2006). This suggests that women’s attention to extra-pair men during estrus (and perhaps their tendency toward ornamentation, see Haselton et al., 2007) may drive their partner’s increased attention to them.

In sum, the variety and volume of studies investigating human sexual behavior as a function of cycle status are substantial, and the importance of collecting accurate, comparable ovulatory cycle information is becoming increasingly evident (Gangestad et al., 2016). Next, we review the various methods of ascertaining conception risk, including recently published recommendations for future research (Gangestad et al., 2016; Gonzales & Ferrer, 2016; Jones et al., 2018).

**Ascertaining Conception Risk**

Studies of ovulatory effects differ substantially in their methods, which may explain some of the disparities between studies’ results (e.g., Little, Jones, Burt, & Perrett, 2007, vs. Cárdenas & Harris, 2007). Similar results might suggest that methodological differences between studies are not important, but differences impede comparison of results and some methods may be more reliable than others (Gangestad et al., 2016). At a minimum, studies investigating ovulatory effects must measure an aspect of the phenotype of a woman (or man, e.g., the woman’s primary partner) and relate it to estimates of the woman’s conception risk or hormonal profile. Estimating conception risk requires determining a woman’s position in her cycle (usually the day of her cycle when she participates in a study, with the day of the onset of menstrual bleeding classed as day one, or in relation to the number of days before/after ovulation), or assessing her hormonal profile (by assaying samples of saliva, blood, or urine) and assigning conception risk values or categorical phases according to actuarial data. Researchers might then determine whether the dependent variables of interest vary as a function of cycle phase (e.g., Penton-Voak et al., 1999) or assess the correlation between these dependent variables and assigned conception risk probabilities (e.g., Gangestad & Thornhill, 1998), or both (e.g., Haselton & Miller, 2006). However, although the various methods used to produce estimates of conception risk may differ considerably in accuracy, to suggest that one method is inherently “better” than any other is to overlook the fact that not all are applicable to every situation. For example, it would not be practical to draw blood for hormonal assay or use transvaginal ultrasonography to check follicle growth at every testing session for diary-, field-, or internet-based studies. Thus, our discussion focuses on practical techniques for maximizing accuracy in conception risk estimation for a variety of methods. We begin by outlining the less commonly used methods of transvaginal ultrasonography, basal body temperature, and cervical mucus qualities, and then review the more popular self-report and hormonal assay methods. Lastly, we discuss conception risk values, which can be used in conjunction with other techniques (e.g., self-report).
Transvaginal Ultrasonography, Basal Body Temperature, and Cervical Mucus Qualities

The most accurate method of identifying the timing of ovulation is daily transvaginal ultrasonography. This method involves the insertion of an ultrasound probe into the vaginal canal to examine the ovaries and other reproductive organs. Transvaginal ultrasonography can be used to precisely track follicular growth and release, and to correlate phenotypic changes to these underlying biological events (Cobey et al., 2012, 2013). Although accurate, this method is invasive and requires medical training and expensive equipment. Therefore, transvaginal ultrasonography is unlikely to be considered appropriate for all researchers when the focus is on cyclic shifts in mate preference or behavior.

A simpler method of estimating the timing of ovulation involves tracking a woman’s basal body temperature (BBT), which exhibits a periovulatory increase and remains elevated throughout the luteal phase (Marshall, 1968). The day of ovulation can be approximated using the “three over six rule,” meaning that ovulation typically occurs on the first day of the first time in the cycle when BBT is higher for three consecutive days than the mean of the preceding six days (Marshall, 1968). Exceptions to this rule (p. 114) may be permitted if there is a spike in temperature on only a single day among six lower temperature days, or if a rise in BBT can be accounted for by illness or other disturbances (Colombo & Masarotto, 2000). Participants may be given a chart on which to record their daily temperature readings (and menstruation days), or a commercially available temperature computer may be used (e.g., the OvuTherm “Sophia,” Craig Medical Distribution). Temperature computer devices measure and record daily BBT readings and calculate the probable fertile days of the cycle (Freundl et al., 2003).

The consistency of cervical mucus can also indicate conception risk. Cervical mucus is typically thin, clear, and ductile during the follicular phase. After ovulation, the mucus becomes more viscous and opaque (Colombo & Masarotto, 2000). In addition, follicular phase mucus crystallizes in a “ferning” pattern, which can be seen under a microscope when a drop of mucus is placed on a glass slide and allowed to dry (Freundl et al., 2003). Mucus collected from the vulva can be palpated and examined with the naked eye for opacity and consistency. Because salivary mucus shows similar cyclic changes in ferning patterns, commercial kits may allow for examination of either cervical or salivary mucus (Freundl et al., 2003). Ovulation generally occurs on the last day of the cycle during which cervical mucus is thin, clear, and ductile and exhibits a ferning pattern (Colombo & Masarotto, 2000).

To compare the accuracy of various approaches for detecting ovulation, Freundl et al. (2003) tested eight different methods of cycle monitoring, each on a sample of 14 to 16 women. The methods included hormonal measures, BBT, mucus characteristics, or a combination of these methods. The true day of ovulation was determined by daily urinary luteinizing hormone (LH) measurements and transvaginal ultrasonography. Temperature computer devices exhibited the lowest false-negative rate but high rates of false positives,
Investigating the Ovulatory Cycle: An Overview of Research and Methods

whereas the reverse was true of examining mucus ferning patterns under a microscope. A hormone computer device based on LH and an estrogen metabolite showed intermediate false-negative and false-positive rates. The symptothermal method of natural family planning, in which women chart both daily body temperature and mucus characteristics, produced the greatest overall predictive accuracy. Thus, without transvaginal ultrasonography, the most reliable approach is probably to use two methods: one with a low rate of false positives (mucus characteristics), the other with a low rate of false negatives (BBT). However, hormonal methods offer good overall accuracy (with neither high rates of false positives nor high rates of false negatives), are less invasive, and rely less on (potentially inaccurate) reports by participants. Importantly, although the more valid methods of estimating cycle day involve monitoring physiological variables that change cyclically, including hormone levels, BBT, and the appearance or consistency of salivary and cervical mucus, these methods are often impractical when predicting ovulatory status for scheduling testing sessions. For these purposes, researchers often rely on self-report data.

Self-Report

Location in the cycle can be quickly estimated via self-report, although women are often inaccurate when asked to estimate their cycle length (Jukic et al., 2008; Small, Manatunga, & Marcus, 2007) or last menstrual onset (Waller; Spears, Gu, & Cunningham, 2000; Wegienka & Baird, 2005). Self-report questions often target the date of the onset of menstrual bleeding for the previous or current menses. Providing a calendar or asking participants in advance to track their menstruation using one of the several available mobile phone applications may help participants report more accurately their cycle onset date, which can then be used to establish the cycle day on which the task is administered. The date of the onset of menses is usually classed as day one (i.e., a woman whose period began 10 days ago would currently be at day 11). Conception risk values or categorical phases can then be assigned according to cycle day. This method, usually called the “forward counting” method, is the easiest to use because it requires the participant to provide only one date, usually at the time the task is administered. No further contact with the participant is required and so it is frequently employed in between-subject internet-based studies or when using a between-subject design in general (e.g., Little, Jones, & Burriss, 2007; Little, Jones, & DeBruine, 2008). The “backward counting” method (sometimes called “reverse cycle day” or “backward” method; see Gangestad et al., 2016), which predicts the day of ovulation by counting back from the onset of the menses phase subsequent to testing, is considered more reliable because the follicular phase tends to be more variable in duration than the luteal phase (Baird et al., 1995). The cycle day of interest in relation to ovulation is calculated by assuming that ovulation occurs 14 (Beckmann et al., 1998; Knaus, 1929; Ogino, 1930) or 15 (Dixon, Schlesselman, Ory, & Blye, 1980) days prior to the first day of the next menstruation. Trussell, Rodriguez, and Ellertson (1998) suggest that subtracting 13 days from a woman’s usual cycle length (i.e., 14 days before the next menses begins) provides a less biased estimate. Women vary in the duration of their cycles, so a further optional stage in this
process is to transform women’s estimated cycle days into their expected equivalents in a 28-day cycle, such that the cycle begins on day 1 and ends on day 28, and the actual or estimated day on which ovulation occurs is day 15 (Burriss et al., 2015; Puts, 2006).

There are two general methods of obtaining data that permit the backward-counting method. The first method involves establishing probable cycle duration (by soliciting reports of average duration or having participants specify the dates for more than one previous menstrual onset and calculating the interval). It is then assumed that the next menses will begin one expected cycle length after the previous known date of onset. This method is sometimes referred to as the “forward-backward counting” method and days can be classed relative to estimated ovulation (i.e., number of days before/after). The second method involves having participants report back to the researcher the date of onset of their first menses after testing has ended. This method has the advantage of greater accuracy and of providing data specific to the cycle during which the participant has been tested. However, this method cannot be used to schedule testing during the cycle of interest, and some participants may fail to provide further information after testing has finished, so it is prudent to use both methods in conjunction (e.g., Fales, Gildersleeve, & Haselton, 2014).

Self-report methods have been employed successfully in many studies but are not precise. Ovulation does not follow every menstruation (Wilcox, Weinberg, & Baird, 1995), and the opposite is also true (Check et al., 1989). Moreover, there is evidence that the duration of the luteal phase is not fixed and, as with the duration of the menses and follicular phases, can be highly variable (Stern & McClintock, 1998). Wilcox, Dunson, and Baird (2000) found that only 10 percent of women with 28-day cycles ovulated precisely 14 days before menstruation. The accuracy of self-report methods can be improved by omitting women who report irregular cycles or abnormal cycle lengths. Cutoff points for abnormal cycle length vary across studies, with some authors omitting participants with cycle lengths in excess of 28 days (Guéguen, 2009a, 2009b), in excess of 40 days (DeBruine et al., 2005; Garver-Apgar, Gangestad, & Thornhill, 2008), in excess of 50 days (Gangestad, Simpson, Cousins, Garver-Apgar, & Christensen, 2004), or fewer than 21 days (Bryant & Haselton, 2009; Frost, 1994; Gangestad & Thornhill, 1998), or omitting those who are more than two standard deviations from a mean of 29 days (Puts, 2005).

Although counting methods, and self-report methods in general, are less accurate than other methods (e.g., transvaginal ultrasonography; Cobey et al., 2012, 2013), their relative convenience for both researchers and participants makes their use popular. To address which counting techniques are most accurate, Gangestad et al. (2016) created more than 58,000 simulated cycles using published distributions of the lengths of follicular and luteal phases. Using these data, they assessed the validity of various counting methods and provided suggestions for future research. Their analyses showed modest validity (d range: 0.40 to 0.55) for most counting methods, and that backward-counting with a confirmed next menstrual onset (d = .70) and daily LH testing (d = .85) is most accurate. These authors recommended the use of within-subject designs to ensure sufficient power for analyses, as between-subject designs require substantially larger
sample sizes for equivalent power (see Gangestad et al., 2016, for power analyses). Moreover, researchers should use the backward-counting method with confirmed next menstrual onset and kits to detect the LH surge and/or assay reproductive hormones (Gangestad et al., 2016). In line with these suggestions, Jones et al. (2018) outlined four common issues with cycle work: insufficient power, overuse of biased self-report methods, overuse of between-subject designs, and testing hormones only twice in the cycle (at estimated high and low fertility phases). Taking the aforementioned suggestions and addressing these criticisms of menstrual cycle work may yield important insights into the relationships among conception risk, hormones, and women’s behavior.

**Hormonal Assessment**

To measure hormone levels across the cycle, researchers can have participants visit the laboratory multiple times to provide several samples (i.e., saliva, blood, or urine) for hormonal assay across one or more menstrual cycles (e.g., Jones et al., 2018), or they can use counting methods to estimate the date of ovulation and test participants during the fertile window and other points in the menstrual cycle when fertility is likely to be lower (e.g., Haselton, Mortezaie, Pillsworth, Bleske-Rechek, & Frederick, 2007). Ovulation test kits provide a relatively inexpensive and effective means of monitoring cycling hormones. Most over-the-counter ovulation test kits respond to levels of LH in urine or to both LH and metabolites of estradiol, both of which peak at midcycle. LH levels spike about one or two days before ovulation (Garcia, Jones, & Wright, 1981; World Health Organization, 1980). Urinary tests for the LH surge agree with results from ultrasonography 97 to 100 percent of the time, making this one of the most accurate tests available (Guermandi et al., 2001). When researchers use ovulation test kits, it is typically in conjunction with counting methods; participants are generally scheduled for study sessions according to self-report data and then take LH tests on several consecutive days around the time the surge is expected (e.g., Burriss et al., 2015; Haselton et al., 2007). Once researchers have confirmed an LH surge, they can be reasonably confident that their participant is within the fertile window of their cycle. A further advantage of this method is that the results of LH tests are available immediately, which makes it possible for researchers to reactively schedule testing sessions.

The preovulation rise in LH is also highly correlated with an increase in the ratio of estrogen to progesterone levels (Baird, Weinberg, Wilcox, McConnaughey, & Musey, 1991; Baird, Weinberg, Wilcox, McConnaughey, Musey, & Collins, 1991). Consequently, the ratio of the levels of these hormones can also provide a fairly accurate measure of a woman’s proximity to ovulation (Wilcox et al., 1995), even without the use of ovulation test kits to confirm ovulation. Therefore, the experimenter can collect biological samples from participants, usually saliva (e.g., Welling et al., 2007) or urine (e.g., Venners et al., 2006), at the time(s) of their participation and then store the samples in a noncycling freezer before conducting hormonal assays.
In studies of cyclical phenotypic variation in which users and nonusers of hormonal contraceptives have been separately tested, cyclical effects present in nonusers have consistently not been present in users (e.g., Burriss & Little, 2006; Frost, 1994; Gangestad & Thornhill, 1998; Guéguen, 2009b; Laeng & Falkenberg, 2007; Little et al., 2007; Penton-Voak et al., 1999; Puts, 2005). This suggests that these effects are controlled by fluctuations in hormone levels and are not connected to changes due to menstruation independent of ovulation (Gangestad & Thornhill, 1998). In other words, the underlying hormonal mechanisms, and not fertility per se, are likely responsible for changes in behavior associated with the menstrual cycle. Menstrual cycle effects may be driven by estradiol (Feinberg et al., 2006; Garver-Apgar et al., 2008; Roney & Simmons, 2008; Roney et al., 2011; Rosen & López, 2009; Rupp et al., 2009), progesterone (Garver-Apgar et al., 2008; Jones, Perrett, et al., 2005; Puts, 2006; Rupp et al., 2009), prolactin (Puts, 2006), cortisol (López, Hay, & Conklin, 2009), testosterone (Welling et al., 2007), or a combination of these hormones (e.g., Puts et al., 2013). Also, different hormonal fluctuations can be associated with different phenotypic changes, and it is unlikely that one hormone is singularly responsible for all noted cyclic shifts. Therefore, measuring change in multiple hormones over time—not necessarily in conjunction with high versus low fertility (e.g., across the day)—can yield important information about underlying mechanisms.

Conception Risk Values

A woman’s daily conception risk is an estimate of the average likelihood of her conceiving with a single unprotected copulation. After estimating the day of the cycle on which a participant has been tested, a researcher may determine that the day is inside or outside of the cycle’s fertile window or assign the day a conception risk value taken from one of several studies. Daily conception risk values may be reported by cycle day relative to the onset of the previous menses (e.g., Jochle, 1973; Wilcox, Dunson, Weinberg, Trussell, & Baird, 2001) or relative to the day of ovulation (e.g., Wilcox et al., 1995). Conception risk is strongly determined by the timing of copulation in relation to ovulation (Wilcox et al., 1995), and the number of days since the onset of the previous menstruation may be a poor estimate of a woman’s proximity to ovulation. For example, the 14th day after the onset of menstruation may be very near ovulation for a woman with a 28-day cycle, but probably several days away from ovulation for a woman with a 32-day cycle. Thus, greater precision in estimating the fertile window facilitates detection of significant relationships between estimated conception risk and preferences, behaviors, or attitudes. However, this is not to say that a narrower fertile window (e.g., six days) is necessarily preferable to a broader window (e.g., nine days): When a woman’s actual date of ovulation is unknown, it is advisable to assign a broader fertile window so as to increase the likelihood of including genuinely high fertility days (Gildersleeve, Haselton, & Fales, 2014b). Gangestad et al. (2016) assessed the optimal window size and found that broader windows (eight to nine days) are more likely to be valid, except when the backward-counting method is used and the date of ovulation is known (in which case the chances of classifying a fertile day as a nonfertile day are reduced). Nevertheless,
Estimates of daily conception risk values relative to the day of ovulation are available from several sources (Barrett & Marshall, 1969; Bremme, 1991; Colombo & Masarotto, 2000; Schwartz, MacDonald, & Heuchel, 1980; Schwartz, Mayaux, Martin-Boyce, Czyglik, & David, 1979; Stirnemann, Samson, Bernard, & Thalabard, 2013; Weinberg, Wilcox, Baird, & Gladon, 1998; Wilcox et al., 1995, 2001; Wilcox, Weinberg, & Baird, 1998), although some of these studies report reanalyses of the same data set. The appropriate values for use in menstrual effect studies will depend on the quality of the methods and the sample size used in each conception risk study. The reliability of conception risk values in these studies depends on both the precision in detecting ovulation and the proper attribution of pregnancies to coital acts on the days on which conception occurred. As Colombo and Masarotto (2000) point out, attributing pregnancies to the coital day closest to the presumed day of ovulation (e.g., Bremme, 1991) produces a bias that artificially increases conception risks nearer the day of ovulation. The model of Schwartz et al. (1980) produces the best estimate of the day of conception for cycles in which more than one coital act occurred (Colombo & Masarotto, 2000). Other researchers have used the Schwartz et al. (1980) model and have sufficiently reliable methods for detecting ovulation—using BBT (Schwartz et al., 1980), BBT plus mucus ferning (Colombo & Masarotto, 2000), or the ratio of estrogen to progesterone levels (e.g., Wilcox et al., 1998)—that we may consider their daily conception risk estimates reliable. Wilcox et al. (2001) provide conception risk values for each day of the cycle (i.e., the probability of conception following unprotected sex for each day of the cycle) using data from 696 cycles (213 women), and these values are commonly used by researchers (e.g., Roney & Simmons, 2008). More recently, Stirnemann et al. (2013) calculated the probability of being within a five-day fertile window for each day of the cycle based on ultrasound measurements from 5,830 early pregnancies and determined that day 12 (using the forward-counting method) had the highest conception probability. Although Wilcox et al. (2001) and Stirnemann et al. (2013) defined conception risk differently (i.e., probability of conception on a given day vs. probability of being in the five-day fertile window), validities of estimates using either method are roughly equivalent (Gangestad et al., 2016).

Conclusion

The method researchers use depends largely on their available resources and on their quasi-independent variable of interest. If conception risk is the variable of interest, it makes sense to exclude participants when there is evidence that their cycles are anovulatory. However, if a researcher is interested in underlying hormonal mechanisms, then anovulation matters less because there are still (albeit smaller) hormone fluctuations across anovulatory cycles (e.g., Ellison, Lager, & Caffee, 1987). If a researcher is interested in conception probability or the fertile window in general, it is preferable to estimate the day of a woman’s cycle on which she is tested relative to the day on which...
she ovulates rather than relative to the day on which her previous menses began. The day of ovulation can most accurately be predicted via ovulation test kits or a combination of BBT and mucus methods, with Gangestad et al. (2016) recommending the backward-counting method with participants’ confirmation of next menstrual onset and use of LH test strips to confirm ovulation. Because underlying hormonal mechanisms, rather than fertility per se, likely drive menstrual shifts in phenotype (e.g., Gangestad et al., 2016; Roney & Simmons, 2013), multiple hormonal assays using within-subject designs with sufficiently large samples are preferred (Gangestad et al., 2016; Jones et al., 2018). Furthermore, researchers relying on counting methods are recommended to exclude participants with irregular cycles or abnormal cycle lengths at the recruitment stage (e.g., using a prescreening questionnaire), rather than after data are collected, to avoid the appearance of p-hacking.

Finally, none of the suggestions outlined in this chapter necessarily invalidate findings of previous studies whose methods differ from those recommended here. However, the suggestions do underline the need for replication of underpowered studies and those studies using less than ideal methods, as well as for preregistering methods to limit the opportunity for post hoc flexibility (Harris et al., 2014). Our recommendations, like those outlined by Gangestad et al. (2016) and Jones et al. (2018), are intended to serve as guidelines for future research on menstrual effects and on the evolution of human behavior in general.

(p. 118) As reviewed in this chapter, many menstrual cycle effects have been documented in the literature. These shifts in behavior, preferences, and appearance serve putatively adaptive functions that likely increase the chances of reproducing, especially with optimal partners. Menstrual cycle effects illustrate how underlying mechanisms can inform ultimate explanations for human behavior. Despite recent controversies (e.g., Gildersleeve et al., 2014a, 2014b; Harris et al., 2014; Wood et al., 2014) and inconsistencies (see Jones et al., 2018), this line of research remains a fruitful area for further investigation.

References


Investigating the Ovulatory Cycle: An Overview of Research and Methods


Investigating the Ovulatory Cycle: An Overview of Research and Methods


Fessler, D. M. T. (2003). Rape is not less frequent during the ovulatory phase of the menstrual cycle. *Sexualities, Evolution & Gender, 5*(3), 127-147. doi: 10.1080/146166041001662361


Investigating the Ovulatory Cycle: An Overview of Research and Methods


Investigating the Ovulatory Cycle: An Overview of Research and Methods


Investigating the Ovulatory Cycle: An Overview of Research and Methods


Investigating the Ovulatory Cycle: An Overview of Research and Methods


Investigating the Ovulatory Cycle: An Overview of Research and Methods


Lisa L. M. Welling
Oakland University, Rochester, MI, USA

Robert P. Burriss
University of Basel, Basel, Switzerland