Received: 11 January 2021 | Revised: 22 July 2021 | Accepted: 23 July 2021

DOI: 10.1002/ajhb.23663

FEATURE ARTICLE

American Journal of Human Biology WILEY

Recognizing normal reproductive biology: A comparative analysis of variability in menstrual cycle biomarkers in German and Bolivian women

Virginia J. Vitzthum[1] | **Jonathan Thornburg**[2] | **Hilde Spielvogel**[3] | **Tobias Deschner**[4]

[1]Evolutionary Anthropology Laboratory, Department of Anthropology, & The Kinsey Institute, Indiana University, Bloomington, Indiana, USA

[2]Department of Astronomy, & IUCSS, Indiana University, Bloomington, Indiana, USA

[3]Instituto Boliviano de Biología de Altura, La Paz, Bolivia

[4]Department of Primatology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Correspondence
Tobias Deschner, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
Email: deschner@eva.mpg.de

Funding information
Fulbright Association; Indiana University Bloomington; Max-Planck-Institut für Evolutionäre Anthropologie; University of California, Riverside; National Science Foundation; School for Advanced Research

Abstract

The idealized "normal" menstrual cycle typically comprises a coordinated ebb and flow of hormones over a 28-day span with ovulation invariably shown at the midpoint. It's a pretty picture—but rare. Systematic studies have debunked the myth that cycles occur regularly about every 28 days. However, assumptions persist regarding the extent and normalcy of variation in other cycle biomarkers. The processes of judging which phenotypic variants are "normal" is context dependent. In everyday life, normal is that which is most commonly seen. In biomedicine normal is often defined as an arbitrarily bounded portion of the phenotype's distribution about its statistical mean. Standards thus defined in one population are problematic when applied to other populations; population specific standards may also be suspect. Rather, recognizing normal female reproductive biology in diverse human populations requires specific knowledge of proximate mechanisms and functional context. Such efforts should be grounded in an empirical assessment of phenotypic variability. We tested hypotheses regarding cycle biomarker variability in women from a wealthy industrialized population (Germany) and a resource-limited rural agropastoral population (Bolivia). Ovulatory cycles in both samples displayed marked but nonetheless comparable variability in all cycle biomarkers and similar means/medians for cycle and phase lengths. Notably, cycle and phase lengths are poor predictors of mid-luteal progesterone concentrations. These patterns suggest that global and local statistical criteria for "normal" cycles would be difficult to define. A more productive approach involves elucidating the causes of natural variation in ovarian cycling and its consequences for reproductive success and women's health.

1 | INTRODUCTION

Two major reasons for investigating the extent and causes of variation in human biology are *How do we identify dysfunction?* and *How do we interpret biological differences within and between populations?*

The first question is more often the focus of biomedicine, reflecting its core mission to improve human health. The second is a foundational motivator for bioanthropology, which seeks to understand the evolution and adaptations of our highly polytypic species. Although somewhat divergent in aims and methods, both intellectual arenas grapple with the general question of *What is normal?* and apply their answers to better the lives of humans in the present and to better understand the lives of those in the past.

To achieve such goals, Western scientific tradition often dichotomizes otherwise continuous phenotypic variability into normal versus abnormal. This approach is historically rooted in Descartes' (1637) influential conceptualization of the body as a well-tuned machine, necessarily invariant in the form and coordination of its components. Variants not meeting the criteria for "normal" are classified as pathologies requiring medical intervention. Such criteria for hormones and other biomarkers typically involve the designation of upper and/or lower thresholds outside of which the biomarker is considered abnormal (see Vitzthum, 2020; Wiley, 2021 for further discussion).

The inherent challenges, even arbitrariness, of dichotomizing continuous variation in a biomarker is illustrated by the continuing debates over defining biomarker thresholds for identifying luteal phase deficiency/defect (LPD) and whether LPD even exists as a pathology warranting medical intervention (Abdulla et al., 2018; Chung et al., 2018; Mesen & Young, 2015). LPD, defined by Georgeanna Jones (1949) based on several biomarkers, is characterized as insufficient progesterone production for endometrial development and successful implantation. Over the course of many decades several studies have concluded that criteria for LPD are inadequate in one way or another for definitively identifying LPD cases. Collectively, these findings prompted The Practice Committee of the American Society for Reproductive Medicine (2012, 2015) to state, "No minimum serum progesterone concentration defines 'fertile' luteal function" and "LPD, as an independent entity causing infertility, has not been proven." Nevertheless, noting the importance of progesterone in successful pregnancy, there remain many thoughtful dissenters from this opinion; progesterone supplementation has long been and still is a treatment for women given a diagnosis of LPD (Mesen & Young, 2015).

Normalizing the familiar, and hence pathologizing (or at least "othering") the unfamiliar, is arguably a common process across human cultures. In her consideration of what "normal" means, Mead (1947) proposed that "... the statistically usual is identified with the basically human...", that is, the most commonly observed phenotypes in one's own population are equated with "normal" human bodies and behaviors. Wiley and Cullin (2020) have elaborated on this point, developing a novel framework (Biological Normalcy) for investigating the interplay between normative (sociocultural) views of normal bodies and statistical distributions of biological traits (see the contributions to this special issue).

Biomedical and bioanthropological research on menstrual cycling overlap and diverge in interesting ways that exemplify how normative views influence our understanding of the human female body. For more than 130 years, physicians and gynecological researchers, almost all in North America and Western Europe, expended considerable effort on documenting and debating the normal duration and predictability ("regularity") of menstrual cycling in their fellow country women (Vitzthum, 2009). For much of that time, it was the implicit assumption of such work that the physiology of these women is necessarily that of the entire species.

In contrast, with Washburn's (1951) call for the integration of evolutionary theory into the study of human biology, bioanthropology began to shift its mission from merely measuring and classifying humans to interpreting biological variation as potential adaptations to local environments. Studies of human biological, behavioral, and cultural adaptations, undertaken in diverse populations, revealed marked variation in numerous traits. Any attempts to define normal took a backseat to understanding how a phenotype contributes to an individual's ability to "surmount the challenges to life" (Lasker, 1969; Mazess, 1975; Thomas et al., 1989); also see (Leslie & Little, 2003; Vitzthum, 2008). The coin of this realm is a specific phenotype's fitness (lifetime reproductive success [LRS]), not where it falls in either a local or global statistical distribution of the trait.

Nonetheless, it was not until the final decades of the twentieth century that substantial data on ovarian functioning in women from nonindustrialized, natural fertility (i.e., noncontracepting) populations became available (Ellison et al., 1986; Jasienska & Ellison, 1998; Johnson et al., 1987; Konner & Worthman, 1980; Leslie et al., 1993; Leslie et al., 1996; Leslie & Fry, 1989; Panter-Brick et al., 1993; Strassmann, 1996a, 1997; van der Walt et al., 1977; Vitzthum, 2001a; Vitzthum et al., 1994; Worthman et al., 1993).

The delay is attributable, in part, to the technological and logistical challenges of investigating the reproductive

physiology of populations in remote locales (Leslie & Little, 2003; Vitzthum, 2021). Furthermore, in the natural fertility populations in which anthropologists typically work, a large segment of the fecund women of reproductive age are either pregnant or breastfeeding, and thus not experiencing menstrual cycles. This reproductive reality creates a significant selection bias. Not only are there fewer women who can provide data on menstrual cycling, these women are more likely to be the youngest, the oldest, and the least fertile.[1] Hence menstrual cycling, and the associated biomarkers, in the available cycling women may not reflect the breadth of cycle biology in the study population.

Depending on the research question, one solution to this challenge is to collect data from each willing woman, recruited regardless of her current cycling status. This strategy was successfully used to test hypotheses regarding menstrual hut visits and menstrual cycling in the Dogon (Strassmann, 1996a, 1997) and to investigate rates of pregnancy loss in Turkana (Leslie et al., 1993, 1996) and Bangladeshi populations (Holman, 1996; Holman & Wood, 2001).

If the goal is unbiased determination of ovarian cycle hormone concentrations, then each premenopausal woman, recruited regardless of cycling status, must be followed until she has completed one or more cycles during which there is collection of hormonal (e.g., progesterone) and nonhormonal (e.g., menses onset) biomarkers. By adding a biomarker of conception (a urine test for hCG), it is also possible to ascertain whether and how conception cycles differ from nonconception cycles. Although there have been such longitudinal studies in industrialized populations (beginning with Baird et al., 1991), to the best of our knowledge only a single study has collected such longitudinal data from a non-industrialized population, Project REPA (Reproduction and Ecology in Provincía Aroma). Over the course of 2 years, 316 rural Bolivian women participated for up to 8 cycles of biospecimen sampling to measure progesterone and detect conceptions, which were then followed through to either natural pregnancy loss or full-term birth.

The data from Project REPA demonstrated that these Bolivian women successfully conceived at progesterone levels only about two-thirds as high as the progesterone levels of US women (Vitzthum et al., 2004). In other words, contrary to an appearance of subfecundity in comparison to US women, the progesterone levels in this population are clearly normal in terms of adequate function and reproductive fitness. This finding refutes the assumption that the relatively high levels of progesterone characteristic of US women are necessary for the successful production of a human offspring and are therefore an appropriate standard against which to evaluate ovarian steroid concentrations in other populations.

1.1 | Defining "normal" menstrual cycles

More generally, the results from Project REPA and other studies prompt the question, *How do we recognize normal female reproductive biology in diverse human populations?*

The specific context motivating this question guides the possible answers. Health care providers need to recognize pathology and to treat it, and human biologists need to differentiate adaptive phenotypes and correctly characterize the nature of the evolutionary processes that have shaped our species. In everyday life the most commonly seen phenotypes are understood to be normal (Mead 1947). In biomedicine normal is often equated with healthy, which the World Health Organization defines as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." It is difficult, arguably impossible, to either achieve or operationalize this ideal. Bodies malfunction for various reasons. No complex organism is likely to function at peak performance (however operationalized) throughout an entire lifespan, even in the very best of conditions. It's neither useful nor informative to restrict "normal" to only those phenotypic variants achieving peak performance and to label all other variants as "not healthy" or "abnormal."

In contemporary evolutionary sciences, the concept of normal is informed by the theories and evidence of adaptation. Adaptive phenotypic variants are those that have the relatively highest LRS in the local environment. Comparing LRS of phenotypic variants is one option for judging what is normal. There are, however, some challenges in this approach.

Higher LRS need not be associated with the healthiest persons in a population. Investment in offspring typically comes at a cost to self. Any organism is a compromise, the result of evolutionary processes and proximate factors acting on the array of morphological, physiological and other traits composing the individual. It is thus impossible to maximize the performance of all these traits. Trade-offs in performance are inevitable; the dynamic interaction of these trade-offs, for the most part unknown to us at this time, contribute to an individual's LRS. Furthermore, measuring LRS is difficult, especially in a species that lives for many decades during which a given phenotypic state may appear to be nonadaptive in the present moment but is, in fact, adaptive by virtue of having contributed to relatively higher *lifetime* RS. For example, although anovulation is an outcome of some

pathologies, anovulation when breastfeeding (lactational amenorrhea) is a mechanism that increases LRS and is therefore rightly considered adaptive (and normal). Thus, whether anovulation itself is normal cannot be judged independently of a functional context. Although physiological phenotypes are especially susceptible to misinterpretation when attempting to distinguish adaptive/normal from nonadaptive/abnormal, such a distinction is a pervasive problem for most phenotypes. Nonetheless, even if its measurement is difficult, LRS is a plausible criterion for identifying normal phenotypes, with the caveat that relying on LRS may not serve biomedicine's mission to identify and eliminate pathology.

Another approach for recognizing normal phenotypic variation is to focus on the specific proximate mechanisms and functions associated with a given phenotype (Vitzthum, 2020). The yardstick for normal is thus shifted from concerns with statistical distributions and debates over appropriate standards to ascertaining the various pathways, signals, and responses that are determinants of how the phenotype operates in the body of which it is a part.

It is worth reminding ourselves that a "phenotype" is defined by investigators. Our own predilections for what to measure bias our observations and our conclusions. The study of luteal-phase progesterone concentrations is telling on this point. Given the role of progesterone in pregnancy, it is reasonable to posit the hypothesis that the amount of progesterone produced during a cycle is an important determinant of fertility. If so, it follows that a cycle with low progesterone levels may be less fertile than those with higher levels; the same may be said for populations with lower average progesterone levels. However, if empirical evidence that such cycles and populations are not in fact less fertile, then the hypotheses must be re-examined.

Rather than the absolute quantity of progesterone, perhaps it is the *change* in progesterone concentration that is the biologically salient phenotype (Vitzthum, 2020). This hypothesis is consistent with the evidence that prompted The Practice Committee of the American Society for Reproductive Medicine (2012, 2015) to conclude that LPD as defined was not a cause of infertility and the finding that rural Bolivians have a high total fertility rate despite having some of the lowest progesterone levels. Given the limits of our knowledge of the mechanisms underpinning human reproductive functioning, it is likely that various commonly measured biomarkers of the menstrual cycle are red herrings—we measure them because we can rather than because of any known mechanistic or adaptive function.

Although determining LRS or untangling proximate mechanisms can be formidable, these are more productive approaches than comparing diverse populations to the statistical distribution of a phenotype in some designated population, particularly if chosen for reasons of familiarity and/or the accidents of history. Taking the tack that each population has its own normal (i.e., distribution) and then, as before, binning continuous variation into categories based on arbitrary thresholds carries the limitations noted previously for this approach and does not necessarily address the questions of interest. The prevalence of pathology and the tradeoffs among functions likely do differ across populations. Assuming that the statistical mean and an arbitrarily selected distribution about the mean is "normal" may fail to recognize these patterns and processes. Recognizing normal by investigating LRS and/or proximate mechanisms is also better aligned with the push for evidence based medicine in the health fields and for discarding "just so stories" in the evolutionary sciences in favor of testable hypotheses (Bateson & Laland, 2013; Tinbergen, 1963; Vitzthum, 2020; Williams, 1966) .

1.2 | Investigating menstrual cycle variability

Studies of LRS and of proximate mechanisms are ideally grounded in an empirical assessment of variability in a specified phenotype. Without such data there is a tendency in actual practice to fall back on the everyday concept of normal—that which is commonly seen and measured—with all its attendant limitations. Identifying and quantifying the patterns of phenotypic variation within and between populations is key to pinpointing the causes and consequences of this variation.

To this end, we have tested hypotheses regarding variability in menstrual cycling with data from two very different populations. The cycle biomarkers selected for our analyses have been widely used to characterize ovarian cycles and, in some instances, to distinguish purportedly abnormal/dysfunctional cycles. The hypotheses that we test reflect some commonly held implicit and explicit beliefs of "normal" menstrual cycling that have influenced interpretations of within and between population variation in cycle phenotypes.

Virtually every depiction of the menstrual cycle, whether in medical texts and adorning the walls of health care facilities or in less technical writings and websites, shares an idealized vision of the ebb and flow of "female" hormones over a 28-day span. Ovulation is invariably at the midpoint, neatly halving a 4-week cycle into follicular (preovulatory) and luteal (postovulatory) phases. It's a pretty picture—symmetrical, rhythmic, predictable—and rare.

This flawed portrayal is the referent for millions of girls and women, serving in effect as the most commonly

observed menstrual phenotype in those populations in which biomedicine is widely practiced. Unlike readily visible phenotypes, populational variability in menstrual cycling is mostly hidden from view in everyday life. Because otherwise invisible hormones are made manifest in these drawings, a veil appears lifted. But lacking any indicators of biological or statistical variability, the depicted statistical norm (mean, median) for features of the cycle becomes *the* normative (the culturally acceptable phenotype). Since most women do not always have 28-day cycles, curiosity and concern may prompt questions. "Am I normal?" "How abnormal am I?" When normal is narrowed to the statistical mean, the judged "other" is likely to include the self. Likewise, focusing on differences in population means for a phenotype without considering the overlaps in the distributions lends itself to questioning the normalcy of populations that differ from an assumed standard.

The veracity or not of such assessments requires documenting and comparing variation in cycle biomarkers across samples representing various populations and subpopulations.

Cycle length (duration in days) is the most studied feature of the menstrual cycle. Bookended by the first day of menstrual bleeding, menses can be recorded on a calendar or other device and the days counted (Vitzthum, 2021). Several studies (Creinin et al., 2004; Jukic et al., 2008) have reported various biases in recalled cycle length and variability (also referred to as predictability or regularity) including a preference for reporting 28 or 30 day lengths, which would tend to give the impression of less cycle variability than actually exists. Nonetheless, more than a century of recorded data collection and debate has produced the consensus opinion that marked variability in cycle length is, in fact, common and normal (i.e., not pathological). For example, the International Federation of Gynecology and Obstetrics (FIGO) classified cycles 24 to 38 days long as "normal" (Fraser et al., 2011) and, based on their analyses of a multidecade longitudinal data set from US women, Harlow et al. (2000) recommended defining "standard" cycles as those from 18 to 40 days long.

Cycle length is properly seen as a biomarker rather than as an intrinsic attribute of ovarian functioning. Cycle length itself does not have an inherent biological significance; rather, length in days is a perceived feature of the cycle that can be defined and measured. Identical cycle lengths can result from a wide combination of follicular and luteal phase lengths. By inference, it is likely that identical cycle lengths are associated with marked variation in reproductive physiology including ovarian hormone levels.

Intriguingly, the recognition of marked natural (and normal) variation in cycle length has not been accorded to the evident variation in cycle hormones. This difference in perspective is likely due, in part, to the recency of our ability to measure hormones in biosamples (blood, saliva, and urine) collected frequently over the course of a cycle (Vitzthum, 2021). With continuing technological advancements comes the acquisition of more data and the growing recognition that there is much more hormonal variation than previously appreciated. It must be kept in mind, however, that although hormonal measurements are more direct biomarkers than cycle length of the underlying ovarian physiology, these hormonal measurements are nonetheless still proxies for physiological processes, the full complexity of which we are only beginning to grasp (Prior, 2020).

Although groundbreaking, early studies of ovarian steroid hormones (progesterone, estrogens) relied on small samples and/or on assaying only a few blood samples from each participant, in some cases collecting samples without regard to cycle day or only timed to the preceding first day of menstrual bleeding. There was at that time less appreciation for the full extent of cycle variability and, in any case, serial daily blood sampling for a full cycle is prohibitively expensive and demanding for researchers and study participants alike.

Pioneering cross-population comparisons of ovarian steroids sought to evaluate hypothesized links between these hormones and the risks for breast and other cancers. In general, these studies found lower mean concentrations of ovarian steroids in Asian compared to US and U.K. "white" populations (Dickinson et al., 1974, MacMahon et al., 1974, Trichopoulos et al., 1984, Bernstein et al., 1990, Key et al., 1990, Shimizu et al., 1990, Wang et al., 1991). There was little, if any, suggestion in the published literature from these epidemiological studies that the relatively lower hormone levels observed in the Asian samples were anything other than "normal" (e.g., that the hormone levels might be dysfunctional or could cause population differences in fertility).

The first anthropological study of ovarian hormones in African populations also observed lower concentrations compared to those observed in European and US samples (van der Walt et al. 1977). However, these investigators explicitly argued that the lower concentrations were indicative of lower fecundity and were perhaps evolutionary adaptations to energetic stress. Their interpretation reflects both a biomedical lens that equates low ovarian steroid levels with dysfunction/pathology and an evolutionary lens that argues that even seeming impairment (e.g., lower fecundity) may be adaptive.

Community-based studies of complete cycles in large samples became feasible with the development of assays

for the measurement of unbound steroids in saliva samples and steroid metabolites in urine samples. Measurements in urine and saliva are reasonably reliable biomarkers for luteal-phase progesterone concentrations in serum (considered the gold standard for endogenous hormone levels) (reviewed in Vitzthum, 2021). Because of differences in assay protocols and laboratory conditions, there are likely to be differences across laboratories in their measurement of the absolute concentration of a given analyte in a specific sample even though each assay is a demonstrably reliable biomarker (Dabbs Jr et al., 1995). Nonetheless, different assays can be expected to show comparable *variability* in progesterone concentration across a longitudinal series of samples from a single individual because each biomarker is co-varying with the same changes in endogenous physiology. In our analyses we are thus able to take advantage of these relationships to evaluate similarities and differences in variability in progesterone biomarkers in our German and Bolivian study samples.

2 | STUDY HYPOTHESES

Our study hypotheses test the common (but generally unstated) assumptions that a "normal" menstrual cycle occurs on a reasonably predictable schedule with ovulation at the midpoint of the cycle and hormonal changes occurring in a predictable rhythm in concert with the timing of menses and ovulation.

As discussed earlier, one feature (cycle length) of this idealized version of the cycle has been extensively studied and shown to be far more variable than is widely appreciated. Nonetheless, the myths regarding cycle length persist and these myths, in fact, underpin a substantial amount of research in many fields including biomedicine and evolutionary sciences.

To test our hypotheses, we used data from our studies of German and rural Bolivian women, chosen because Germany is a wealthy industrialized country with universal health care and Bolivia's rural population is resource poor with, at the time of the study, among the highest fertility and mortality rates in the Western hemisphere. Also, as noted earlier, the longitudinal study design in Project REPA reduced selection bias, thereby producing a study sample of cycles representative of this population.

The selected biomarkers included cycle length, follicular and luteal phase lengths, progesterone concentration during its luteal peak, and late luteal duration (biomarkers are defined under Methods below).

Our first null hypothesis is that, absent strong evidence to the contrary, within population *variability* in a phenotype (in the present case, a cycle biomarker) is about the same across populations. This expectation is predicated on a century's worth of data that variation within human populations is often greater than variation between populations (that is, most of the total variation is within rather than between populations, a reflection of our species' shared biology). This pattern is demonstrable for numerous traits spanning the molecular to the morphological, and has been observed in an impressive variety of organisms (e.g., Hunley et al., 2016; Lewontin, 1972; Relethford, 2002).

Our second null hypothesis is that the covariance of an index of mid-level progesterone concentration to the length (in days) of the entire cycle, the follicular and luteal phases, and the late luteal duration is about the same in both populations. This hypothesis derives from the finding that the cycles of Bolivian women are fecund even though the progesterone levels are lower than those observed in a sample of US women. Since the observed progesterone levels in both samples are functionally sufficient for reproduction, the covariance of progesterone levels to other cycle biomarkers is not expected to differ between the Bolivians and Germans.

Our third null hypothesis is that cycle length is an equally poor predictor of mid-luteal progesterone concentration in both study populations. As previously discussed, cycle length itself does not have an inherent biological significance. Identical cycle lengths can result from a wide combination of follicular and luteal phase lengths. By inference, it is likely that identical cycle lengths are associated with marked variation in reproductive physiology including progesterone concentrations (i.e., the covariance is poor). This pattern is expected to be comparable across populations.

It is not our intention, nor would it be wise, to examine every possible cycle biomarker that could be defined from the observed data. Multiple statistical testing runs the risk of producing spurious significant outcomes by chance alone. Therefore, our selection of biomarkers for the present analyses is not exhaustive but rather is intended to reflect a few of the more common implicit assumptions about cycle physiology, particularly as regards variation in cycle and phase lengths, and the relationships between these biomarkers and progesterone concentrations (see [Harris & Vitzthum, 2013; Vitzthum, 2009] for additional discussion).

3 | METHODS

3.1 | Analytical variables

Following common practice, *cycle length* was defined as the total number of days from first day of menstrual

bleeding up to and including the day before the first day of bleeding of the subsequent cycle. *Follicular phase* (i.e., time to ovulation) was defined as the first day of menstrual bleeding up to and including the day before ovulation. *Luteal phase* was defined as the day of ovulation up to and including the cycle's last day. *Late luteal duration* was defined as the number of days from observed progesterone peak up to and including the last day of the cycle.

For each cycle, serial hormone concentrations were aligned on the first day of the subsequent cycle (=reverse-day$_0$) and cycle days were numbered in reverse order (-1, -2, -3 ...) until the cycle's first day of menstrual bleeding. Hormone indices (H, referring to salivary progesterone (P4) in the Bolivian sample and urinary pregnanediol glucuronide (PdG) in the German sample) were defined as ($\int$ of H from day$_x$ to day$_y$)/(day$_y$ − day$_x$), where x to y is any span of days and H at any time is defined by linear interpolation of the observed H measurements. For these analyses, the defined index for luteal phase H was *mean-peak-H* (x = day of peak luteal H − 2.5, y = day of peak luteal H + 2.5) (Vitzthum, 2021; Vitzthum et al., 2004).

Crude ovulation rate (ovulatory cycles/total cycles) was calculated for the Bolivian and German samples. In each sample, total cycles include all cycles for which a determination of ovulatory or anovulatory could be made according to the criteria given below. Cycles in which a conception was detected are included in the count of ovulatory cycles for calculation of ovulation rate. Conception cycles are not included in the analyses for other biomarkers because conception causes changes to the hormonal profiles distinct from those in nonconception ovulatory cycles. Because up to 8 cycles were collected from each Bolivian woman, there is some selection bias towards a lower estimate of the crude ovulation rate (i.e., women who are more likely, for whatever reason, to experience anovulation will not become pregnant and therefore will disproportionately contribute more cycles to the sample).

3.2 | German study

All study protocols were approved by the Institutional Review Board at Indiana University, and all participants gave informed consent. Healthy participants were recruited from Leipzig, Göttingen, Potsdam, and Hannover, Germany, during 2008 through posted notices and by word of mouth. The study sample ($n = 63$) comprised nonpregnant premenopausal German adults who had not used hormonal contraception for at least 3 months prior to study, had never had any hormonal medical treatments, and were not following any special dietary practices or physical training regimens. Beginning the day after menstrual bleeding started and continuing through at least the first day of menstrual bleeding of the subsequent cycle, each woman self-collected and froze ($\leq -5°C$) daily first-morning urine samples stored in polypropylene tubes. Samples were then transported with ice packs to the laboratory where they were stored at $-80°C$ until assayed for pregnanediol glucuronide (PdG, the principal urinary metabolite of progesterone) with a direct microtiter plate enzyme immunoassay (Heistermann et al., 1993; Meyer et al., 1990). PdG concentrations were standardized by creatinine [Cr], expressed as ng PdG/mg Cr. Inter-assay coefficients of variation were 11.7 and 11.4% for high- and low-value quality controls respectively. Intra-assay coefficients of variation were 10.1 and 7.8% for high- and low value quality controls, respectively (see Milich et al. (2015) for additional details).

The occurrence and timing of ovulation was determined by a sustained rise in PdG of 2 standard deviations above the mean of the previous three to five values (following Deschner et al. (2003)). Cycles in which (mean-luteal-PdG) < 2(mean-follicular-PdG) were designated anovulatory and excluded (n = 5); 1 ovulatory cycle in which luteal phase sample collection was inadvertently truncated was also excluded. Other than the calculation for ovulation rate, the final sample for the analyses presented here is 57 complete ovulatory cycles from 57 German women aged 22–41 years; mean (SD) = 30.0 (4.8) years; median = 29 years.

3.3 | Bolivian study

All study protocols were approved by the Institutional Review Board at the University of California, Riverside, and all the participants gave informed consent. The data for the present analyses are from Project REPA (Vitzthum et al., 2004; Vitzthum et al., 2006). The study sample comprised premenopausal Bolivian adults who had never used hormonal contraception nor any hormonal medical treatments, and were not following any special dietary practices or physical training regimens. All of the women lived in rural communities primarily dependent on agropastoralism; seasonal variation in arduous physical labor and food supplies is common (Vitzthum et al., 2009). To avoid the potential selection that can occur in natural fertility populations if only those who are currently menstruating were to participate, women were recruited without regard to breastfeeding or menstrual status. Breastfeeding/nonmenstruating women contributed infant feeding and other data. With the initiation of the first postpartum menses, they then also followed the same data collection protocols as all other menstruating study participants (Vitzthum et al., 2000;

Vitzthum et al., 2001). Of 316 study participants (approximately 80% of the eligible residents in the study region), 191 contributed data on their menstrual cycles.

Beginning shortly after menstrual bleeding started and continuing through to at least the first day of menstrual bleeding of the subsequent cycle, saliva samples were collected from a woman in polypropylene tubes (containing sodium azide as a preservative) thrice weekly by a research assistant during a home visit. Samples were subsequently shipped to a US laboratory where they were stored at −20°C until assayed for salivary progesterone (P4) by direct radioimmunoassay; intraassay and interassay CVs were 9.9% and 12.0%, respectively (see Lu et al. (1999); Lu, Chatterton, Vogelsong, & May (1999); and Vitzthum et al., (2008) for additional details). A study of paired saliva-serum samples synchronously collected from Bolivian women demonstrated that the ratio of salivary P4 to serum P4 is similar to that found in other populations (Thornburg & Vitzthum, 2009).

The calculation for ovulation rate is described above. For all other analyses, the occurrence of ovulation was based on several criteria (Vitzthum et al., 2004). In brief, cycles lacking a sustained rise in P4, or in which mean-peak-luteal-P4 < 110 pmol/liter, or in which mean-peak-luteal-P4 ≤ mean-follicular-P4 were designated anovulatory and excluded. Cycles in which a conception was detected and cycles following an early pregnancy loss were also excluded. Only those cycles for which we could confidently estimate the day of ovulation based on a sustained rise in P4 were included in our analyses. Based on this set of criteria, the final sample for the analyses presented below is 223 completed ovulatory cycles from 102 Bolivian women aged 21–38 years; mean (*SD*) = 29.0 (4.9); median = 30.

The Bolivian and German samples do not differ in their age structure.

3.4 | Analyses

Analyses were done using custom-written Perl programs, the Statistics::Regression Perl package (version 0.53), and SPSS (v27.0). To mitigate the risk of spurious associations (Holländer et al., 2004), our analyses did not presume arbitrarily selected bins (thresholds, cutoffs) for any continuous variables.

Analyses included ovulation rates and descriptive statistics of hormone indices, cycle length, follicular and luteal phase lengths, and late luteal duration (the time in days from peak-PdG or peak-P4 to the cycle's end). To evaluate hypothesized associations between the peak progesterone concentration indices and the other biomarkers, we calculated r^2 based on Pearson correlation analyses. We also evaluated whether age was a significant covariate in any of these analyses.

4 | RESULTS

Immediately below we present the outcomes from each of our analyses. In the subsequent section we discuss these findings.

Crude ovulation rate is 388/526 = 74% (95% CI for the population rate, assuming our sample is representative: 71%–78%) for the Bolivian sample and 58/63 = 92% (95% CI for the population rate, assuming our sample is representative: 82%–97%) for the German sample.

Descriptive statistics for the cycle biomarkers in these two samples of ovulatory cycles are given in Table 1, and r^2 from correlation analyses of the cycle biomarkers are given in Table 2. Scatterplots, histograms, and cumulative distributions of the data are presented in Figures 1–3.

TABLE 1 Descriptive statistics of cycle biomarkers

	Units	Median	Mean	*SEM*	*SD*	CV	range	5th per.	95th per.
Germany (n = 57 cycles, 57 women)									
Cycle length	days	29.0	29.2	0.56	4.22	14.5	21–47	25	36
Follicular-phase length	days	15.0	15.8	0.52	3.92	24.8	8–30	11	22
Luteal-phase length	days	14.0	13.4	0.26	1.95	14.6	9–17	10	17
Late luteal duration	days	6.0	6.5	0.27	2.05	31.5	3–11	3	10
Mean-peak-PdG	ng/mg Cr	4983	5013	244.6	1848	36.9	1500–9928	2185	7991
Bolivia (n = 223 cycles, 102 women)									
Cycle length	days	28.0	28.4	0.25	3.75	13.2	19–50	24	35
Follicular-phase length	days	15.0	15.6	0.31	4.62	29.6	5–38	10	24
Luteal-phase length	days	13.0	12.8	0.20	3.03	23.7	5–22	7	18
Late luteal duration	days	7.0	7.1	0.16	2.32	32.7	3–12	3	11
Mean-peak-P4	pmol/L	224	242	6.87	103	42.6	111–658	121	471

TABLE 2 r2 from correlation analyses

	Number of ovulatory cycles	Cycle length	Follicular-phase length	Luteal-phase length	Late luteal duration
Germany: Mean-peak-PdG	57	0.0034	0.0082	0.0950	0.0017
Bolivia: Mean-peak-P4	223	0.00058	0.0005	0.000025	0.0098

The measures of central tendency (median and mean) for cycle length and both phase lengths are very similar in the Bolivian and German samples (Table 1). The median and mean of late luteal duration in the Germans is only slightly shorter than median and mean of late luteal duration in the Bolivians (6.0 days and 6.5 days compared to 7.0 days and 7.1 days).

Of greater relevance in the present analysis is the magnitude of variability, measured here by the ratio of the 95th to the 5th percentile (see Table 1) for each biomarker in the two samples. Cycle length variability is very similar in the two samples, and both phase lengths are only modestly more variable in the Bolivian than in the German sample (which likely reflects the "noise" inherent in the every-other day sampling in the Bolivian study versus the daily sampling in the German study). Furthermore, variability in the progesterone biomarker indices is also very similar in the two samples.

Specifically, in the German women, between the 5th to 95th percentile for each variable, cycle length varies by about 45%, follicular-phase length varies about 2-fold, luteal phase length varies about 1.7-fold, late luteal duration varies 3.33-fold, and mean-peak-PdG varies 3.65-fold.

In the Bolivian women, between the 5th to 95th percentile for each variable, cycle length varies by about 45%, follicular-phase length varies about 2.4-fold, luteal phase length varies about 2.5-fold, late luteal duration varies 3.67-fold, and mean-peak-P4 varies 3.9-fold.

Within each study sample, correlations of peak progesterone indices with cycle length, phase lengths, and late luteal duration were all very low (Table 2). In Germans, r^2 ranged from 0.0017 to 0.095, and in Bolivians r^2 ranged from 0.000025 to 0.0098. The lack of correlations among pairs of these variables is evident in the scatter plots in Figures 1-3.

There was no evidence of significant linear or nonlinear age-, parity- or height-associated variation in any of the examined cycle variables.

5 | DISCUSSION AND CONCLUSIONS

Although the ovary is often spoken of as if it were an autonomous entity, cycle physiology is not isolated from the rest of the body. Advancements on the many aspects of women's health requires recognition of the ubiquity of cycle variability and the integration of this reality into all levels of research from bench to community (Alliende, 2002, 2013; Alvergne & Högqvist Tabor, 2018; Prior, 2020; Shea & Vitzthum, 2020). Likewise, evolutionary models that neglect variability in physiological mechanisms or, worse yet, ignore physiology in favor of simpler models risk promulgating just-so stories and traveling down blind alleys (Bachofner & Lobmaier, 2018; Harris & Vitzthum, 2013; Lobmaier & Bachofner, 2018; Vitzthum, 2020).

The present comparative study of biomarkers of ovarian physiology in two very different populations contradicts assumptions that the "normal" menstrual cycle is necessarily predictable and that the features of such normal cycles (e.g., phase lengths, hormone levels) are highly correlated.

Rather, ovulatory cycles in the Bolivians and Germans displayed marked but nonetheless comparable variability in cycle length (varying about 45% between the 5th and 95th percentiles) and mean-peak-luteal progesterone concentration (varying 3.72-fold and 3.65 fold, respectively, between the 5th and 95th percentiles). Phase lengths and late luteal duration varied modestly more in the Bolivians than in the Germans, perhaps due to the different sampling schedule (thrice-weekly in Bolivians versus daily in Germans).

Reflecting their more arduous living conditions, ovulation rate is lower in the Bolivians than in the Germans (74% versus 92%). This observation highlights the importance of distinguishing ovulatory and anovulatory cycles in analyses of biomarkers lest the very different features characterizing anovulatory cycles confound the analyses of natural variation in ovulatory cycles. In an ovulatory cycle, the ovum is released from the follicle, which then transforms into the corpus luteum that produces progesterone and estradiol. Anovulatory cycles do not, obviously, have a luteal (i.e., postovulatory) phase and, as such, are typically characterized by low (even flat) progesterone and estradiol profiles during the later portion of the anovulatory cycle. Menstrual bleeding may indicate the end of an anovulatory cycle and the beginning of a new follicular phase or may not be evident until the end of a subsequent cycle (long "cycles" defined by the appearance of menstrual bleeding may, in some

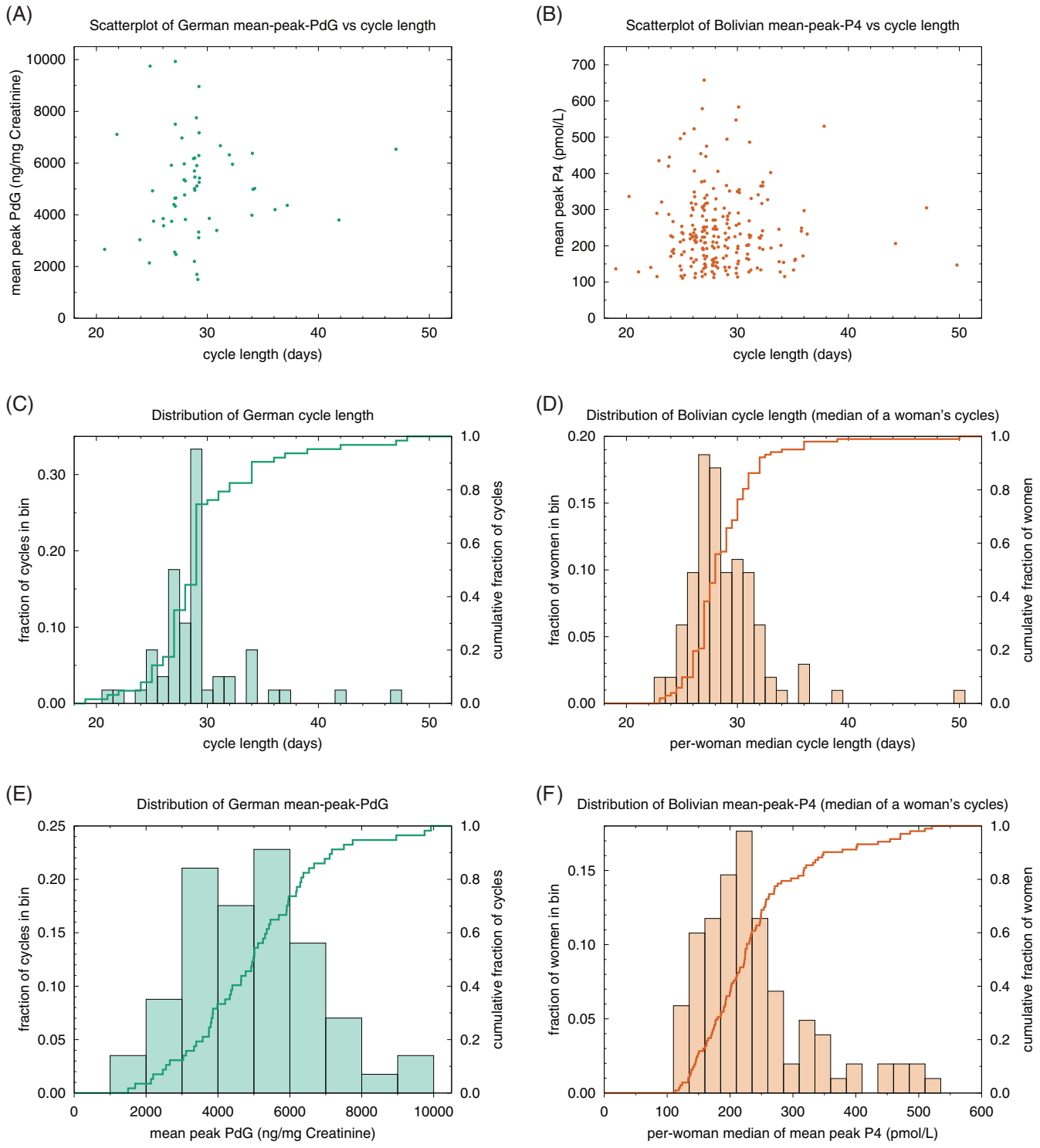

FIGURE 1 Distributions of cycle length and indices of peak-luteal-progesterone biomarkers. Parts (A) and (B) are scatterplots of a peak-luteal-progesterone biomarker versus cycle length, for the German and Bolivian samples respectively; the points are randomly jittered horizontally by up to ±0.3 days to avoid overlapping points. Parts (C) and (D) are the distribution of cycle length, and parts (E) and (F) are the distribution of a peak-luteal-progesterone index, for the German and Bolivian samples respectively. In each of parts (C)–(F) the histogram is plotted using the left vertical scale, and the cumulative distribution is plotted (as the "stepped" curve) using the right vertical scale. Note that in part (D) the distribution includes half-integer days because the median over a single woman's cycles may be a half-integer

cases, comprise an anovulatory cycle followed by the follicular and luteal phases of an ovulatory cycle).

Combining anovulatory and ovulatory cycles will yield average steroid profiles that are easily misinterpreted as a dampening of these hormones in every cycle. Whether a sample consists of ovulatory cycles having high ovarian steroid levels plus a few anovulatory cycles, or several cycles all having reduced ovarian steroid levels, leads to quite

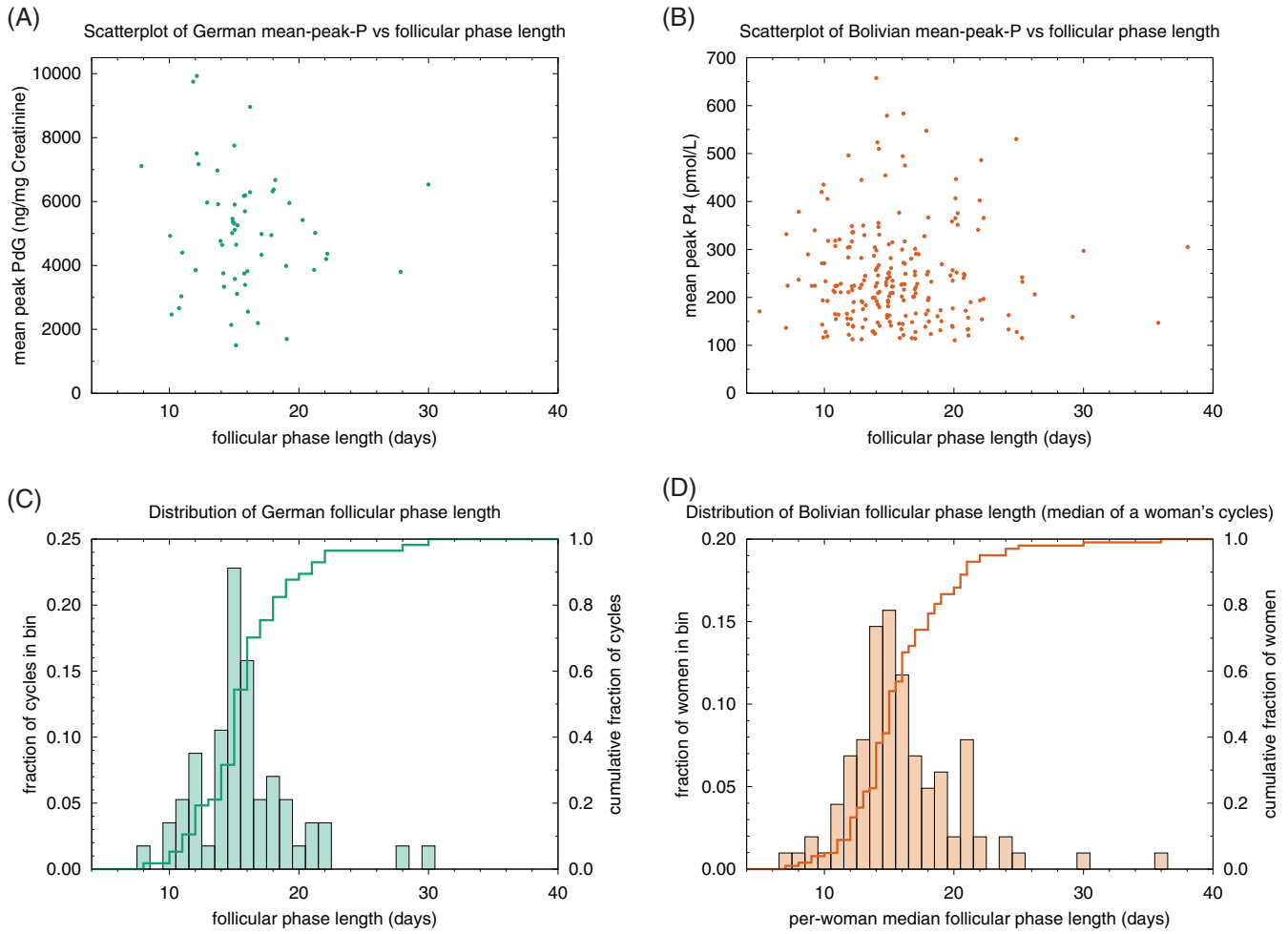

FIGURE 2 Distributions of follicular-phase length and indices of peak-luteal-progesterone biomarkers. Parts (A) and (B) are scatterplots of a peak-luteal-progesterone biomarker versus follicular-phase length, for the German and Bolivian samples respectively; the points are randomly jittered horizontally by up to ±0.3 days to avoid overlapping points. Parts (C) and (D) are the distribution of follicular-phase length for the German and Bolivian samples respectively. In each of parts (C) and (D) the histogram is plotted using the left vertical scale, and the cumulative distribution is plotted (as the "stepped" curve) using the right vertical scale. Note that in part (D) the distribution includes half-integer days because the median over a single woman's cycles may be a half-integer

different conclusions regarding ovarian functioning and the theories that seek to explain that functioning (see Figure 4). Neglecting to distinguish ovulatory and anovulatory segments seriously confounds cross-populational comparisons of hormone levels (Vitzthum, 2009) and, in the case of an individual's cycles, could prompt a mis-diagnosis of luteal phase deficiency. If there is a need for a single index for a sample of ovulatory and anovulatory cycles, this can readily be calculated from the weighted average of the reported indices for each class of cycles. If only the index for the combined sample is reported, then subsequent scholars cannot ascertain separate indices for anovulatory and ovulatory samples. It is, therefore, best practice to publish separate indices for ovulatory and anovulatory cycles if one is also publishing an index for an analytical sample comprising both cycle types.

Age-associated variation in cycle biomarkers has been reported for some, but not all, studies. Such age dependency reflects, in part, the inclusion in some studies of teenagers, peri-menopausal women, and cycles of unknown ovulation status. Because the samples for our study comprised only adults during their peak reproductive years (21–41 years old) and only ovulatory cycles were included in analyses, the absence in both our study samples of a significant association between age and cycle biomarkers is not surprising.

Measures of central tendency (medians, means) in ovulatory cycles in the Bolivian and German samples were similar for all biomarkers excepting, of course, the indices of peak-luteal progesterone concentration. (Recall that differences in the biomarkers and laboratory assays preclude comparison of the absolute hormone concentrations

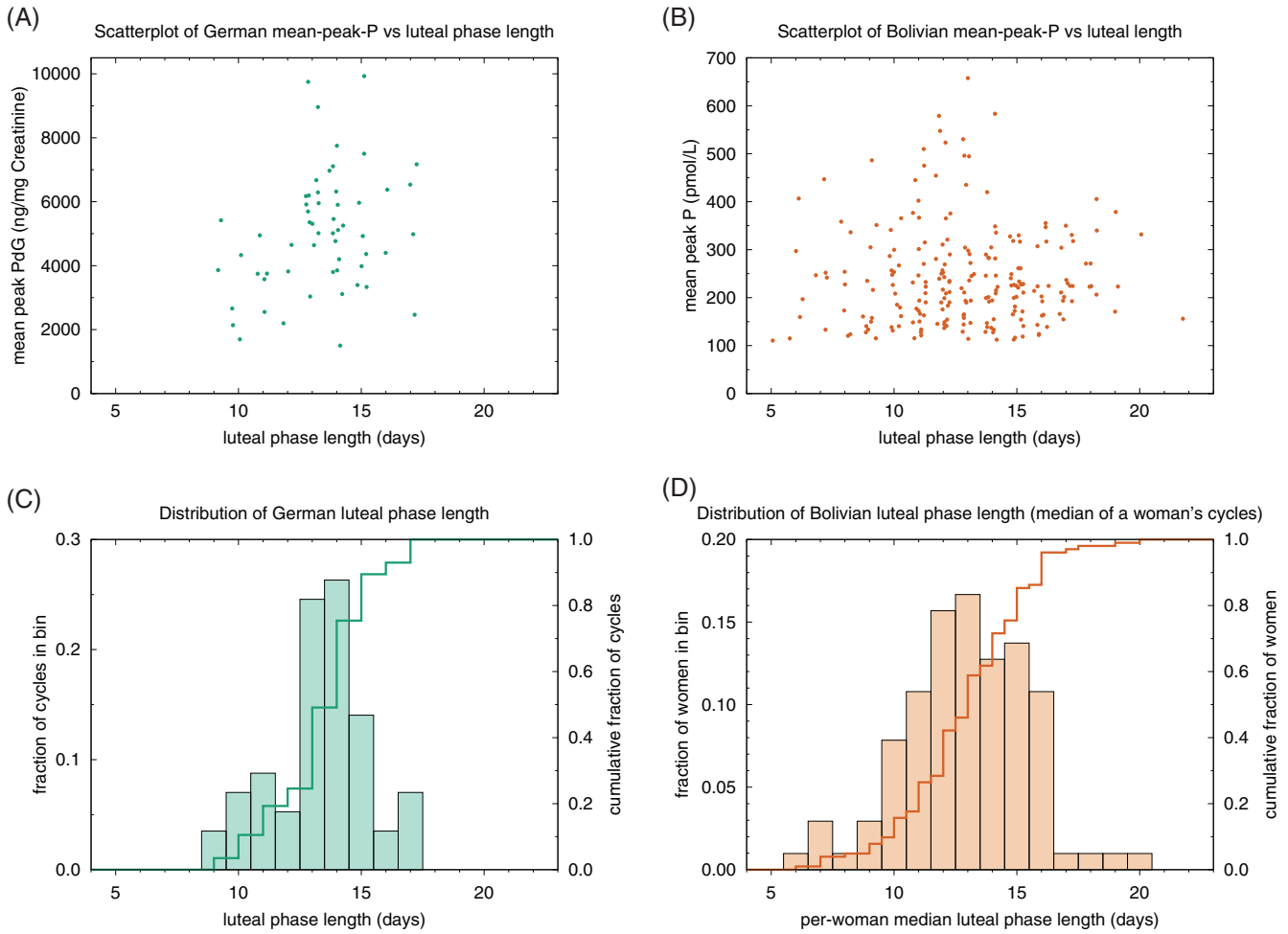

FIGURE 3 Distributions of luteal-phase length and indices of peak-luteal-progesterone biomarkers. Parts (A) and (B) are scatterplots of a peak-luteal-progesterone biomarker versus luteal-phase length, for the German and Bolivian samples respectively; the points are randomly jittered horizontally by up to ±0.3 days to avoid overlapping points. Parts (C) and (D) are the distribution of luteal-phase length for the German and Bolivian samples respectively. In each of parts (C) and (D) the histogram is plotted using the left vertical scale, and the cumulative distribution is plotted (as the "stepped" curve) using the right vertical scale. Note that in part (D) the distribution includes half-integer days because the median over a single woman's cycles may be a half-integer

measured in these two studies, but do not prevent evaluation and comparison of relative variability in the hormone concentrations).

Consistent with our second hypothesis, relative variability in the mean-peak-luteal progesterone indices was similar in the Germans and Bolivians even though mean salivary P4 level in this same Bolivian sample was low relative to mean salivary P4 level in a US sample (Vitzthum et al., 2004).

The patterns of variability observed in our two samples strongly suggests that globally applicable statistical criteria for "normal" menstrual cycles would be very difficult to define and likely be of little practical use.

Cycle length, the most commonly used criterion for evaluating normalcy, is a poor predictor of variation in the underlying physiology (at least as regards the progesterone indices examined here). This lack of much association between these progesterone indices and cycle length suggests that neither is a principal determinant of the other. The same can be said for the other non-hormonal cycle biomarkers observed in this study. A study of reproductive aging in US women similarly concluded that cycle length is too variable to predict hormone levels (Ferrell et al., 2005).

The similarity in reported median and mean cycle lengths (circa 28–29 days) in our Bolivian and German samples should not be taken to be a species norm. The Dogon and several other populations have median/mean cycle lengths of 28–29 days, however, median/mean cycle length was 31 days or more in samples from Japan, India and Papua New Guinea (reviewed in Vitzthum, 2009). Furthermore, predictable 28 or 29 day cycle lengths are not a reliable indicator of nonpathological functioning. Reported cycle lengths for the Lese are 28–29 days

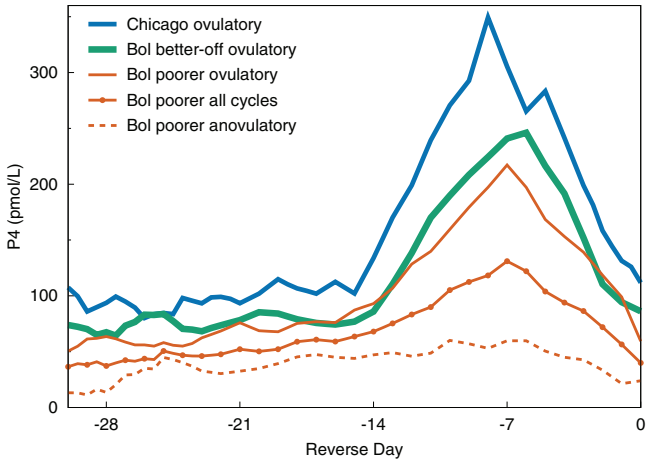

FIGURE 4 Salivary progesterone profiles in Chicago, better-off Bolivian, and poorer Bolivian samples. Cycles are aligned on the first day of the subsequent cycle, and days are numbered backward from that point (reverse day). The poorer Bolivian sample comprising all cycles appears to have very suppressed progesterone levels, but is, in fact, a mixture of two types of cycles: Anovulatory (with flat progesterone profiles) and ovulatory (with a substantially higher progesterone profile than the samples of all cycles). (Modified from Vitzthum et al., 2002)

(Bentley et al., 1990), but many were sterile due to endemic venereal disease (Strassmann, 1997). In the Dogon, "regular menses were a sign of sterility, not fecundity" (Strassmann, 1997).

Our findings present a significant and unsettling challenge for researchers and clinical practice in a variety of fields. It's entirely reasonable that health care providers look to a set of standards for evaluating pathology and that researchers rely on current understandings regarding study variables to streamline their study designs. For these reasons and others, cycle length has been very widely used as a proxy for underlying physiological processes including the timing of ovulation and the rises and falls in hormone levels. For example, in evolutionary psychology some commonly used study designs are predicated on the assumption that ovulation, accompanied by a predicably timed suite of hormonal changes, occurs midway through the cycle. On this basis, claims have been made regarding, for example, an hypothesized human female "estrus" and peri-ovulatory preferences for certain attributes in mates. Not surprisingly (given what we have demonstrated here about the poor covariance between cycle length and hormonal indices), collectively the findings from such studies are inconsistent and disputed (Bachofner & Lobmaier, 2018; Harris & Vitzthum, 2013; Lobmaier & Bachofner, 2018; Strassmann, 2013). Likewise, as discussed earlier, the debates surrounding luteal phase deficiency remain unresolved. There are many other examples where a reliance on assumptions regarding the variability and covariance of cycle biomarkers has muddied the waters rather than leading to insight, discovery, and effective diagnosis.

As discussed earlier, the normalcy of many phenotypic variants cannot be judged independently of a functional context. Anovulation during breastfeeding is now near universally recognized as an adaptive mechanism that directs investment to the current, rather than future, offspring. Notably, our current understanding of lactational amenorrhea came to light only during the final decades of the 20th Century. The initial hypothesis that nursing suppressed ovarian function was contrary to the biomedical position that had long prevailed, but which was eventually changed under the weight of evidence. Of particular importance was the discovery, gained through meticulous research, of the behavioral-physiological mechanisms that linked infant suckling to anovulation (e.g., Konner & Worthman, 1980).

Likewise, anovulation during an illness may be a outcome of a reproductive system unable to function or it may be an adaptive response that favors increased investment in one's own body and temporarily reduced investment in the production of a new offspring. On the flip side, in a healthy person there can be a temporary reduction of investment in immunity during ovulation so as to permit the entry of sperm and investment in reproduction (Abrams & Miller, 2011; Lorenz et al., 2015). Such strategic trade-offs of resources (e.g., nutrients, energy, time) into different life-supporting and offspring-producing functions are a common feature of adaptive life history strategies.

Unlike many morphological phenotypes (especially in adulthood), physiological processes characteristically have the necessary flexibility to adjust functioning (even dramatically) in the face of changing environmental conditions. Normal physiological functioning is not necessarily static. Much like many behavioral phenotypes, the adaptiveness of a physiological state depends on the context. For example, there is no single (or narrow range for) "normal" blood pressure; rather, variation in blood pressure is part of an adaptive response to internal and external conditions that makes it possible for the organism to successfully meet the inevitable challenges of life (James, 2019). Likewise, Bolivian women successfully conceived at relatively low progesterone levels and had one of the highest fertility rates in the Americas.

In the present analyses, nonhormonal biomarkers in the Bolivian and German women are similar in both variability and magnitude. In sum, the lower mean progesterone observed in Bolivians (and other nonindustrialized populations) cannot rightly be interpreted as an indication of impaired or suppressed hormonal production simply because the mean P4 level is numerically similar to

the lower end of the distribution of progesterone levels in wealthy industrialized populations. The progesterone levels in these Bolivian women, and arguably other non-industrialized populations, are clearly functionally normal.

A robust assessment of whether a specific phenotypic variant is a pathology or an adaptive (normal) response is more likely to be realized if grounded in empirical evidence of the underlying mechanisms.

That said, we are left to wonder at the causes of so much variability in the features of the menstrual cycle evaluated here. There are many hypotheses regarding the determinants of variation in human female reproductive biology. Some models focus on the roles of energy intake, expenditure, and stores (body fat) in modulating menstrual cycling (Ellison, 1990, 1994; Frisch & Revelle, 1970; Prior, 1985a; Prior, 1985b; Prior, 1987; Strassmann, 1996b). Others draw on life history theory to explain strategically timing reproductive investment given adverse environments (Chisholm, 1993; Peacock, 1990, 1991; Vitzthum, 1990, 1997, 2001b; Wasser & Barash, 1983). Genetic, dietary, immunological, psychosocial, behavioral and numerous other factors have all been investigated. There is evidence in support of many hypotheses, leaving the conundrum of discerning which are the major determinants of the observed variation in a cycle phenotype.

The answers will likely rely on gaining a more detailed understanding of the specific mechanisms by which these factors may affect reproductive functioning. Due consideration should be given to whether a defined phenotype is a convenient construct (e.g., cycle length) or is a functionally meaningful trait, to phenotypic variation in the population (i.e., the distribution as well as the central tendency) and to physiological flexibility in an individual, to the functional context in which a phenotypic variant is evaluated and to the functional sufficiency and adaptive fitness of a variant in that context.

Although humans are a highly polytypic species, our shared biology makes it likely that what is learned of the operation of mechanisms in one population will be applicable more widely.

Investigating the origins of natural variation in ovarian cycles, and the consequences of such variation for reproductive success and women's health, is likely to be a daunting but nonetheless productive path forward to bettering the lives of humans in the present and to better understanding the lives of those in the past.

ACKNOWLEDGMENTS

We thank the Human Biology Association for their support for the presentation of an earlier version of this paper at the Plenary Session of the HBA 2021 Annual Meetings (organized by Jennifer Cullin and Virginia Vitzthum). We thank Andrea Wiley and Jennifer Cullin for their organization of, and invitation to, the seminar on "Biological Normalcy," hosted in 2019 by the School for Advanced Research (SAR), Santa Fe, New Mexico, with additional financial support from the US National Science Foundation (grant #073285). We thank all the seminar participants for lively and enlightening discussions.

Bolivian Study: We thank E. Caceres for assistance in the communities and the laboratory, and the Instituto Boliviano de Biología de Altura (La Paz) for facilities and logistical support. Financial support to Vitzthum for the collection of Bolivian data, and subsequent data preparation and analyses, was provided by the University of California (Riverside), the State University of New York (Binghamton), Indiana University (Bloomington), and National Science Foundation Grant SBR 9506107. Funding for progesterone assays was provided to G. Bentley and R. Chatterton (Northwestern University, Illinois) by National Science Foundation Grant SBR 9507320.

German Study: We thank C. Deimel and F. Schaebs for assistance in data collection and performing laboratory assays, and the Max Planck Institute for Evolutionary Anthropology (Leipzig) for facilities and logistical support. Financial support was provided to Vitzthum by the Fulbright Senior Fellow Research Award Program and Indiana University (Bloomington), and provided to Deschner by the Max Planck Gesellschaft.

We especially thank the participants of both studies for their patience, contributions, and good humor. Open access funding enabled and organized by Projekt DEAL.

AUTHOR CONTRIBUTIONS

Virginia J. Vitzthum: Conceptualization; formal analysis; funding acquisition; investigation; methodology; resources; writing - original draft. **Jonathan Thornburg:** Formal analysis; software; visualization; writing - original draft; writing-review & editing. **Hilde Spielvogel:** Investigation; methodology; resources; writing-review & editing. **Tobias Deschner:** Investigation; methodology; project administration; resources; writing-review & editing.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT

Research data are not shared at this time.

ORCID

Virginia J. Vitzthum https://orcid.org/0000-0002-6765-4557

Jonathan Thornburg https://orcid.org/0000-0001-6096-8265

Tobias Deschner https://orcid.org/0000-0002-9873-316X

ENDNOTE

[1] An extensive literature, much of it by demographers, on the composition and attributes of "natural fertility" populations has amassed since Henry's (1961) introduction of the concept. Likewise, the study of subfecundity at older and younger ages (i.e., menopause and adolescent subfecundity) has a long history (e.g., Short, 1976). The challenges, including selection biases, inherent in the study of natural fertility populations are widely recognized and debated; a discussion of these is beyond the scope of the present paper.

REFERENCES

Abdulla, S. H., Bouchard, T. P., Leiva, R. A., Boyle, P., Iwaz, J., & Ecochard, R. (2018). Hormonal predictors of abnormal luteal phases in normally cycling women. *Frontiers in Public Health*, *6*, 144. https://doi.org/10.3389/fpubh.2018.00144

Abrams, E. T., & Miller, E. M. (2011). The roles of the immune system in women's reproduction: Evolutionary constraints and life history trade-offs. *Yearbook of Physical Anthropology*, *54*, 134–154.

Alliende, M. E. (2002). Mean versus individual hormonal profiles in the menstrual cycle. *Fertility and Sterility*, *78*, 90–95.

Alliende, M. E. (2013). Luteinizing hormone surge in normally ovulating women. *Fertility and Sterility*, *99*(5), e14–e15. https://doi.org/10.1016/j.fertnstert.2013.01.122

Alvergne, A., & Högqvist Tabor, V. (2018). Is female health cyclical? Evolutionary perspectives on menstruation. *Trends in Ecology & Evolution*, *33*(6), 399–414. https://doi.org/10.1016/j.tree.2018.03.006

Bachofner, L. M., & Lobmaier, J. S. (2018). Towards a more holistic view of fertility: The need to consider biological underpinnings rather than only data. *Hormones and Behavior*, *106*, A10–A11. https://doi.org/10.1016/j.yhbeh.2018.07.004

Baird, D. D., Weinberg, C. R., Wilcox, A. J., McConnaughey, D. R., Musey, P. I., & Collins, D. C. (1991). Hormonal profiles of natural conception cycles ending in early, unrecognized pregnancy loss. *Journal of Clinical Endocrinology & Metabolism*, *72*, 793–800. https://doi.org/10.1210/jcem-72-4-793

Bateson, P., & Laland, K. N. (2013). Tinbergen's four questions: An appreciation and an update. *Trends in Ecology & Evolution*, *28*(12), 712–718. https://doi.org/10.1016/j.tree.2013.09.013

Bentley, G. R., Harrigan, A. M., & Ellison, P. T. (1990). Ovarian cycle length and days of menstruation of Lese horticulturalists. *American Journal of Physical Anthropology*, *81*, 193–194.

Bernstein, L., Yuan, J. M., Ross, R. K., Pike, M. C., Hanisch, R., Lobo, R., Stanczyk, F., Gao, Y. T., & Henderson, B. E. (1990). Serum hormone levels in pre-menopausal Chinese women in Shanghai and white women in Los Angeles: Results from two breast cancer case-control studies. *Cancer Causes Control*, *1*, 51–58.

Chisholm, J. S. (1993). Death, hope, and sex: Life-history theory and the development of reproductive strategies [and comments and reply]. *Current Anthropology*, *34*(1), 1–24. https://doi.org/10.1086/204131

Chung, M. S., Nemer, L. B., & Carr, B. R. (2018). Does luteal phase deficiency exist and what is its association with infertility? *Reproductive Medicine International*, *1*(1), 002. https://doi.org/10.23937/rmi-2017/1710002

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

Dabbs, J., Jr., Campbell, B., Gladue, B., Midgley, A., Navarro, M., Read, G., Susman, E., Swinkels, L., & Worthman, C. M. (1995). Reliability of salivary testosterone measurements, a multicenter evaluation. *Clinical Chemistry*, *41*, 1581–1584.

Descartes, R. (1637). *Discourse on the method of rightly conducting one's reason and seeking the truth in the sciences*. Thomas Newcombe. https://www.gutenberg.org/files/25830/25830-h/25830-h.htm

Deschner, T., Heistermann, M., Hodges, K., & Boesch, C. (2003). Timing and probability of ovulation in relation to sex skin swelling in wild West African chimpanzees, pan troglodytes verus. *Animal Behaviour*, *66*(3), 551–560.

Dickinson, L. E., MacMahon, B., Cole, P., & Brown, J. B. (1974). Estrogen profiles of Oriental and Caucasian women in Hawaii. *New England Journal of Medicine*, *291*, 1211–1213.

Ellison, P. T. (1990). Human ovarian function and reproductive ecology, new hypotheses. *American Anthropologist*, *92*, 933–952.

Ellison, P. T. (1994). Advances in human reproductive ecology. *Annual Review of Anthropology*, *23*, 255–275.

Ellison, P. T., Peacock, N. R., & Lager, C. (1986). Salivary progesterone and luteal function in two low-fertility populations of Northeast Zaire. *Human Biology*, *58*, 473–483.

Ferrell, R. J., O'Connor, K. A., Rodríguez, G., Gorrindo, T., Holman, D. J., Brindle, E., Miller, R. C., Schechter, D. E., Korshalla, L., Simon, J. A., Mansfield, P. K., Wood, J. W., & Weinstein, M. (2005). Monitoring reproductive aging in a 5-year prospective study: Aggregate and individual changes in steroid hormones and menstrual cycle lengths with age. *Menopause*, *12*(5), 567–577. https://doi.org/10.1097/01.gme.0000172265.40196.86

Fraser, I., Critchley, H., Broder, M., & Munro, M. (2011). The FIGO recommendations on terminologies and definitions for Normal and abnormal uterine bleeding. *Seminars in Reproductive Medicine*, *29*, 383–390. https://doi.org/10.1055/s-0031-1287662

Frisch, R., & Revelle, R. (1970). Height and weight at menarche and a hypothesis of critical body weights and adolescent events. *Science*, *169*, 397–399.

Harlow, S. D., Lin, X., & Ho, M. J. (2000). Analysis of menstrual diary data across the reproductive life span applicability of the bipartite model approach and the importance of within-woman variance. *Journal of Clinical Epidemiology*, *53*(7), 722–733. https://doi.org/10.1016/S0895-4356

Harris, A. L., & Vitzthum, V. J. (2013). Darwin's legacy: An evolutionary view of women's reproductive and sexual functioning. *Journal of Sex Research*, *50*(3–4), 207–246. https://doi.org/10.1080/00224499.2012.763085

Heistermann, M., Tari, S., & Hodges, J. K. (1993). Measurement of faecal steroids for monitoring ovarian function in New World primates, Callitrichidae. *Reproduction*, *99*(1), 243–251. https://doi.org/10.1530/jrf.0.0990243

Henry, L. (1961). Some data on natural fertility. *Eugeneics Quarterly*, *18*, 81–91.

Holländer, N., Sauerbrei, W., & Schumacher, M. (2004). Confidence intervals for the effect of a prognostic factor after selection of an 'optimal' cutpoint. *Statistics in Medicine*, *23*(11), 1701–1713. https://doi.org/10.1002/sim.1611

Holman, D. (1996). *Total Fecundability and Pregnancy Loss in Rural Bangladesh (PhD Thesis)*. Pennsylvania State University.

Holman, D., & Wood, J. W. (2001). Pregnancy loss and fecundability in women. In P. T. Ellison (Ed.), *Reproductive ecology and human evolution* (pp. 15–38). Aldine de Gruyter.

Hunley, K. L., Cabana, G. S., & Long, J. C. (2016). The apportionment of human diversity revisited. *American Journal of Physical Anthropology*, *160*(4), 561–569. https://doi.org/10.1002/ajpa.22899

James, G. D. (2019). The adaptive value and clinical significance of allostatic blood pressure variation. *Current Hypertension Reviews*, *15*(2), 93–104. https://doi.org/10.2174/1573402115666190301144316

Jasienska, G., & Ellison, P. T. (1998). Physical work causes suppression of ovarian function in women. *Proceedings of the Biological Sciences*, *265*, 1847–1851.

Johnson, P. L., Wood, J. W., Campbell, K. L., & Maslar, I. A. (1987). Long ovarian cycles in women of highland New Guinea. *Human Biology*, *59*(5), 837–845.

Jones, G. E. S. (1949). Some newer aspects of management of infertility. *JAMA*, *141*, 1123–1129.

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. *American Journal of Epidemiology*, *167*(1), 25–33. https://doi.org/10.1093/aje/kwm265

Key, T. J., Chen, J., Wang, D. Y., Pike, M. C., & Boreham, J. (1990). Sex hormones in women in rural China and in Britain. *British Journal Cancer*, *62*, 631–636.

Konner, M., & Worthman, C. M. (1980). Nursing frequency, gonadal function, and birth spacing among !Kung hunter-gatherers. *Science*, *207*, 788–791.

Lasker, G. W. (1969). Human biological adaptability. *Science*, *166*, 1480–1486.

Leslie, P. W., Campbell, K. L., & Little, M. A. (1993). Pregnancy loss in nomadic and settled women in Turkana, Kenya: A prospective study. *Human Biology*, *65*(2), 237–254.

Leslie, P. W., Campbell, K. L., Little, M. A., & Kigondu, C. S. (1996). Evaluation of reproductive function in Turkana women with enzyme immunoassays of urinary hormones in the field. *Human Biology*, *68*(1), 95–117.

Leslie, P. W., & Fry, P. H. (1989). Extreme seasonality of births among nomadic Turkana pastoralists. *American Journal of Physical Anthropology*, *79*(1), 103–115. https://doi.org/10.1002/ajpa.1330790111

Leslie, P. W., & Little, M. A. (2003). Human biology and ecology: variation in nature and the nature of variation. *American Anthropologist*, *105*, 28–37.

Lewontin, R. C. (1972). The apportionment of human diversity. In T. Dobzhansky, M. K. Hecht, & W. C. Steere (Eds.), *Evolutionary biology* (pp. 381–398). Springer US. https://doi.org/10.1007/978-1-4684-9063-3_14

Lobmaier, J. S., & Bachofner, L. M. (2018). Timing is crucial: Some critical thoughts on using LH tests to determine women's current fertility. *Hormones and Behavior*, *106*, A2–A3. https://doi.org/10.1016/j.yhbeh.2018.07.005

Lorenz, T. K., Worthman, C. M., & Vitzthum, V. J. (2015). Links among inflammation, sexual activity and ovulation: Evolutionary trade-offs and clinical implications. *Evolution, Medicine, and Public Health*, *2015*(1), 304–324. https://doi.org/10.1093/emph/eov029

Lu, Y., Bentley, G. R., Gann, P. H., Hodges, K. R., & Chatterton, R. T. (1999). Salivary estradiol and progesterone levels in conception and nonconception cycles in women: Evaluation of a new assay for salivary estradiol. *Fertility and Sterility*, *71*(5), 863–868. https://doi.org/10.1016/S0015-0282

MacMahon, B., Cole, P., Brown, J. B., Aoki, K., Lin, T. M., Morgan, R. W., & Woo, N. C. (1974). Urine oestrogen profiles of Asian and North American women. *International Journal of Cancer*, *14*, 161–167.

Mazess, R. B. (1975). Biological adaptation: Aptitudes and acclimatization. In E. S. Watts, F. E. Johnston, & G. W. Lasker (Eds.), *Biosocial interrelationships in population adaptation* (pp. 9–18). The Hague: Mouton.

Mead, M. (1947). The concept of culture and the psychosomatic approach. *Psychiatry*, *10*, 57–76.

Mesen, T. B., & Young, S. L. (2015). Progesterone and the luteal phase. *Obstetrics and Gynecology Clinics of North America*, *42*(1), 135–151. https://doi.org/10.1016/j.ogc.2014.10.003

Meyer, H. H. D., Sauerwein, H., & Mutayoba, B. M. (1990). Immunoaffinity chromatography and a biotin-streptavidin amplified enzyme immunoassay for sensitive and specific estimation of estradiol-17β. *The Journal of Steroid Biochemistry and Molecular Biology*, *35*, 263–269.

Milich, K. M., Deimel, C., Schaebs, F. S., Thornburg, J., Deschner, T., & Vitzthum, V. J. (2015). Links between breast cancer and birth weight: An empirical test of the hypothesized association between size at birth and premenopausal adult progesterone concentrations. *Hormones and Cancer*, *6*(4), 182–188. https://doi.org/10.1007/s12672-015-0221-6

Panter-Brick, C., Lotstein, D. S., & Ellison, P. T. (1993). Seasonality of reproductive function and weight loss in rural Nepali women. *Human Reproduction*, *8*, 684–690.

Peacock, N. R. (1990). Comparative and cross-cultural approaches to the study of human female reproductive failure. In J. DeRousseau (Ed.), *Primate life history and evolution* (pp. 195–220). Wiley-Liss.

Peacock, N. R. (1991). An evolutionary perspective on the patterning of maternal investment in pregnancy. *Human Nature*, *2*(4), 351–385. https://doi.org/10.1007/BF02692197

Practice Committee of the American Society for Reproductive Medicine. (2012). The clinical relevance of luteal phase deficiency: A committee opinion. *Fertility and Sterility*, *98*(5), 1112–1117. https://doi.org/10.1016/j.fertnstert.2012.06.050

Practice Committee of the American Society for Reproductive Medicine. (2015). Current clinical irrelevance of luteal phase deficiency: A committee opinion. *Fertility and Sterility*, *103*(4), e27–e32. https://doi.org/10.1016/j.fertnstert.2014.12.128

Prior, J. C. (1985a). Hormonal mechanisms of reproductive function and hypothalamic adaptation to endurance training. In J. Puhl (Ed.), *The menstrual cycle and physical activity* (pp. 63–75). Human Kinetics.

Prior, J. C. (1985b). Luteal phase defects and anovulation: Adaptive alterations occurring with conditioning exercise. *Semester Reproductive Endocrinology*, *3*, 27–33.

Prior, J. C. (1987). Physical exercise and the neuroendocrine control of reproduction. *Baillière's Clinical Endocrinology and Metabolism*, *1*, 299–317.

Prior, J. C. (2020). Women's reproductive system as balanced estradiol and progesterone actions—A revolutionary, paradigm-shifting concept in women's health. *Drug Discovery Today: Disease Models*, *32*, 31–40. https://doi.org/10.1016/j.ddmod.2020.11.005

Relethford, J. H. (2002). Apportionment of global human genetic diversity based on craniometrics and skin color. *American Journal of Physical Anthropology*, *118*(4), 393–398. https://doi.org/10.1002/ajpa.10079

Shea, A. A., & Vitzthum, V. J. (2020). The extent and causes of natural variation in menstrual cycles: Integrating empirically-based models of ovarian cycling into research on women's health. *Drug Discovery Today: Disease Models*, *32*, 41–49. https://doi.org/10.1016/j.ddmod.2020.11.002

Shimizu, H., Ross, R.K., Bernstein, L., Pike, M. C., & Henderson, B. E. (1990). Serum oestrogen levels in postmenopausal women, comparison of American whites and Japanese in Japan. *British Journal of Cancer*, *62*, 451–453.

Strassmann, B. I. (1996a). Menstrual hut visits by Dogon women: A hormonal test distinguishes deceit from honest signaling. *Behavioral Ecology*, *7*(3), 304–315. https://doi.org/10.1093/beheco/7.3.304

Strassmann, B. I. (1996b). The evolution of endometrial cycles and menstruation. *The Quarterly Review of Biology*, *71*(2), 181–220.

Strassmann, B. I. (1997). The biology of menstruation in *Homo sapiens*: Total lifetime menses, fecundity, and non-synchrony in a natural-fertility population. *Current Anthropology*, *38*(1), 123–129. https://doi.org/10.1086/204592

Strassmann, B. I. (2013). Concealed ovulation in humans: Further evidence. In K. Summers & B. Crespi (Eds.), *Foundations of human behavioral evolution in the works of R. D. Alexander* (pp. 139–151). Oxford University Press.

Thomas, R. B., Gage, T. B., & Little, M. A. (1989). Reflections on adaptive and ecological models. In M. Little & J. D. Haas (Eds.), *Human Population Biology* (pp. 296–319). Oxford University Press.

Thornburg, J., & Vitzthum, V. J. (2009). Quantifying selection bias in cross-sectional studies of ovarian hormones. *American Journal of Human Biology*, *21*, 271. https://scholarworks.iu.edu/dspace/handle/2022/25473

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschr Tierpsychol*, *20*, 410–433.

Trichopoulos, D., Yen, S., Brown, J., Cole, P., & MacMahon, B. (1984). The effect of Westernization on urine estrogens, frequency of ovulation, and breast cancer risk. A study of ethnic Chinese women in the Orient and the USA. *Cancer*, *53*, 187–192.

van der Walt, L. A., Wilmsen, E. N., Levin, J., & Jenkins, T. (1977). Endocrine studies on the San ('Bushmen') of Botswana. *South African Medical Journal*, *52*, 230–232.

Vitzthum, V. J. (1990). An Adaptational Model of Ovarian Function (Report 90-200). In *Research reports, Population Studies Center* (pp. 90–200). University of Michigan.

Vitzthum, V. J. (1997). Flexibility and paradox: The nature of adaptation in human reproduction. In M. E. Morbeck, A. Galloway, & A. Zihlman (Eds.), *The evolving female: A life history perspective*. Princeton University Press.

Vitzthum, V. J. (2001a). The home team advantage: Reproduction in women indigenous to high altitude. *The Journal of Experimental Biology*, *204*, 3141–3150.

Vitzthum, V. J. (2001b). Why not so great is still good enough: Flexible responsiveness in human reproductive functioning. In P. T. Ellison (Ed.), *Reproductive ecology and human evolution* (pp. 179–202). Aldine de Gruyter.

Vitzthum, V. J. (2008). Evolutionary models of Women's reproductive functioning. *Annual Review of Anthropology*, *37*, 53–73.

Vitzthum, V. J. (2009). The ecology and evolutionary endocrinology of reproduction in the human female. *Yearbook of Physical Anthropology*, *52*, 95–136.

Vitzthum, V. J. (2020). How it works: The biological mechanisms that generate demographic diversity. In O. Burger, R. Lee, & R. Sear (Eds.), *Human Evolutionary Demography*. Open Book Publishers. https://osf.io/p59eu/; https://hdl.handle.net/2022/25797.

Vitzthum, V. J. (2021). Field methods and strategies for assessing female reproductive functioning. *American Journal of Human Biology*, *33*, e23513. https://doi.org/10.1002/ajhb.23513

Vitzthum, V. J., Ellison, P. T., & Sukalich, S. (1994). Salivary progesterone profiles of indigenous Andean women. *American Journal of Physical Anthropology*, *18*, 201–202.

Vitzthum, V. J., Spielvogel, H., Caceres, E., & Gaines, J. (2000). Menstrual patterns and fecundity among non-lactating and lactating cycling women in rural highland Bolivia: Implications for contraceptive choice. *Contraception*, *62*(4), 181–187. https://doi.org/10.1016/S0010-7824

Vitzthum, V. J., Spielvogel, H., Caceres, E., & Miller, A. (2001). Vaginal bleeding patterns among rural highland Bolivian women: Relationship to fecundity and fetal loss. *Contraception*, *64*(5), 319–325. https://doi.org/10.1016/S0010-7824

Vitzthum, V. J., Spielvogel, H., & Thornburg, J. (2004). Interpopulational differences in progesterone levels during conception and implantation in humans. *Proceedings of the National Academy of Sciences*, *101*(6), 1443–1448. https://doi.org/10.1073/pnas.0302640101

Vitzthum, V. J., Spielvogel, H., Thornburg, J., & West, B. (2006). A prospective study of early pregnancy loss in humans. *Fertility and Sterility*, *86*(2), 373–379. https://doi.org/10.1016/j.fertnstert.2006.01.021

Vitzthum, V. J., Thornburg, J., & Spielvogel, H. (2009). Seasonal modulation of reproductive effort during early pregnancy in humans. *American Journal of Human Biology*, *21*, 548–558.

Wang, D. Y., Key, T. J., Pike, M. C., Boreham, J., & Chen, J. (1991). Serum hormone levels in British and rural Chinese females. *Breast Cancer Research and Treatment*, *18*(Suppl. 1), 41–45.

Washburn, S. L. (1951). The new physical anthropology. *Transactions of the New York Academy of Sciences*, *13*(7), 298–304. https://doi.org/10.1111/j.2164-0947.1951.tb01033.x

Wasser, S. K., & Barash, D. P. (1983). Reproductive suppression among female mammals: Implications for biomedicine and sexual selection theory. *The Quarterly Review of Biology*, *58*(4), 513–538.

Wiley, A. S. (2021). Pearl Lecture: Biological normalcy: A new framework for biocultural analysis of human population variation. *American Journal of Human Biology*, *33*(5), e23563.

Wiley, A. S., & Cullin, J. M. (2020). Biological normalcy. *Evolution, Medicine, and Public Health*, *2020*(1), 1–1. https://doi.org/10.1093/emph/eoz035

Williams, G. C. (1966). *Adaptation and natural selection: A critique of some current evolutionary thought*. Princeton University Press.

Worthman, C. M., Jenkins, C. L., Stallings, J. F., & Lai, D. (1993). Attenuation of nursing-related ovarian suppression and high fertility in well-nourished intensively breast-feeding Amele women of lowland Papua New Guinea. *Journal of Biosocial Science*, *25*, 425–443.

How to cite this article: Vitzthum, V. J., Thornburg, J., Spielvogel, H., & Deschner, T. (2021). Recognizing normal reproductive biology: A comparative analysis of variability in menstrual cycle biomarkers in German and Bolivian women. *American Journal of Human Biology*, e23663. https://doi.org/10.1002/ajhb.23663