Quantifying Selection Bias in Cross-Sectional Studies of Ovarian Hormones

Jonathan Thornburg[1] and Virginia J. Vitzthum[2,3]
[1] Department of Astronomy, Indiana University, Bloomington, Indiana, USA
[2] Anthropology Department, Indiana University, Bloomington, Indiana, USA
[2] Kinsey Institute for Research in Sex, Gender, & Reproduction, Indiana University, Bloomington, Indiana, USA

Abstract

Most studies of ovarian hormones in adult women collect data from a cross-sectional sample of participants meeting various selection criteria including not having been pregnant or breastfeeding for several months. Although this approach is intended to eliminate the effects of these factors on hormonal variation, it introduces a selection bias of unknown magnitude: in a non-contracepting population, those women with the highest fecundity are more likely to be either pregnant or lactating, and so not included in a study sample. Thus a cross-sectional sample disproportionately represents women with the lowest fecundity (and potentially the lowest hormone levels). Here we present a preliminary evaluation of the magnitude of this selection bias, focusing on progesterone (P) levels near the luteal peak. We use data from Project REPA, a longitudinal study of reproductive functioning in rural Bolivians, recruited without regard to reproductive status (Vitzthum *et al.*, 2004). Drawing from 542 non-conception cycles in 144 women, we construct simulated cross-sectional samples meeting various inclusion criteria and compare their anovulation rates and progesterone levels.

Introduction

For reasons of logistics and cost, most field studies of ovarian hormones in free-living women use a cross-sectional design, where participants are recruited based on various selection criteria, and then each participant is observed for a relatively short period of time (often only a single ovarian cycle). Typical selection criteria include not having been pregnant or breastfeeding (BF) for several months. While such a study design appears to eliminate the effects of these factors on hormonal variation, it introduces a selection bias: in a non-contracepting population, those women with the highest fecundity are more likely to be either pregnant or lactating, and hence not included in the study population. Thus such a cross-sectional sample disproportionately represents women with the lowest fecundity (and potentially the lowest hormone levels).

There is clear evidence of within-population variation in fecundity. The apparent fecundability (defined here as the number of observed conceptions divided by the total number of observed cycles) was more than twice as high in BF than in non-BF women in our sample (Vitzthum *et al.*, 2000). European data show a four-fold difference in fecundability among non-contracepting same-aged women (Dunson *et al.*, 2002). This argues that selection bias may be a serious problem when natural-fertility populations are studied.

Here we present what appears to be the first quantitative assessment of the magnitude of the selection bias in cross-sectional studies of ovarian hormones. We use data from Project REPA (Reproduction and Ecology in Provincá Aroma), a longitudinal study of reproductive functioning in rural Bolivians, recruited without regard to reproductive status (Vitzthum *et al.*, 2004). We construct simulated cross-sectional samples meeting various selection criteria, and compare their progesterone (P) levels and the fraction of their cycles which are anovulatory.[1]

The Reproductive Cycle. Figure 1 gives a schematic view of part of an adult (post-menarcheal, pre-menopausal) woman's reproductive life cycle, in which she progresses from pregnancy, through birth, BF without cycling, BF while cycling, and finally cyling without BF. (Of course, any actual woman need not go through all these phases in succession, in particular she may conceive at any time after she resumes ovulation.) Here we only consider the last two of these phases, i.e., we focus on non-conception cycles starting with a woman's first post-partum menses and continuing until her next conception. Our sample is quite close to unbiased within these phases of the reproductive cycle, although there remains some residual selection bias against earlier phases (a woman who conceives at her first post-partum ovulation would not be represented in our sample).

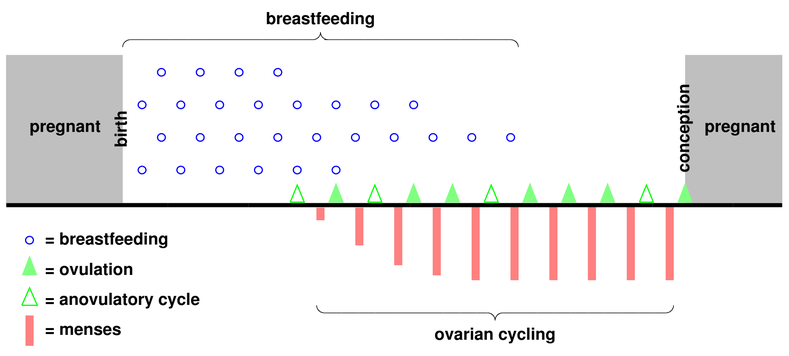

Figure 1: Schematic view of part of an adult woman's reproductive life cycle in between two pregnancies. Time runs from left to right.

Materials and Methods

Population and Samples.. Data collection was conducted within the framework of Project REPA, a multidisciplinary longitudinal study of reproductive functioning and health among rural Aymara families indigenous to the Bolivian altiplano. Preliminary work began in 1989, followed by more than 2 years of continuous field work in 1995–1997. All study protocols were approved by the Institutional Review Board of the University of California, Riverside.

Volunteers, recruited during 12 months beginning in November 1995, represented 80% of the eligible women (aged 19-40 years, currently in stable sexual unions, and not using contraception) in 30 communities scattered over 200 km^2 situated about midway between La Paz and Oruro. The basic data for this study comprise serial saliva samples, collected approximately every 2–3 days beginning at the first post-partum menses (or at recruitment if the woman was already cycling when recruited) and later assayed for P. Conceptions were detected using urine samples (collected starting at day 24–25 of each cycle), tested for human chorionic gonadatrophin (hCG) using field pregnancy-test kits. Additionally, women's BF status was recorded for each cycle, and reproductive-history questionnaires provided information on BF and menstruation status prior to recruitment.

Here we use a sample of 542 non-conception cycles contributed by 144 of the 316 adult female participants. Participants contributed 1–7 completed cycles each, with a median of 4. Of these 144 women, 68 contributed at least one cycle while BF, 81 contributed at least one cycle while not BF, and 33 contributed both BF and non-BF cycles.

Progesterone Indices and Ovulation Status. For each cycle (considered to start on the first day of the menses), we computed P indices and classified the cycle as ovulatory or anovulatory in the manner described in Vitzthum *et al.* (2004). Briefly, we defined mean peak-luteal P (P_{mpl}) as the average P in a $\pm$2.5-day range centered on the highest observed P value in the luteal period, and mean follicular P (P_{mf}) as the average P in a roughly 10-day period in the follicular phase; we classified a cycle as ovulatory if $P_{\text{mpl}} > 110$ pmol/liter and $P_{\text{mpl}} > P_{\text{mf}}$. (We have previously shown (Vitzthum *et al.*, 2004) that this anovulation criterion shows excellent agreement with visual assessments based on inspection of the serial P values.)

For each woman contributing one or more ovulatory cycles to a given cross-sectional sample, we define $\langle\langle P_{\text{mpl}} \rangle\rangle$ to be the "per-woman" mean of P_{mpl} for all of her ovulatory cycles in the sample. Our analyses then consider the distribution of $\langle\langle P_{\text{mpl}} \rangle\rangle$ among all women in the sample. (Note that this definition weights each woman in the sample equally, regardless of how many cycles she contributed to the sample.)

Cross-Sectional Samples. We consider three (simulated) cross-sectional samples:

early BF
This sample contains all cycles where the woman is BF and the cycle start date is < 90 days after the start of her first post-partum menses. This sample (74 cycles from 32 women) contains only the first few post-partum BF cycles from each woman.

later BF
This sample contains all cycles where the woman is BF and the cycle start date is ≥ 90 days after the start of her first post-partum menses. This sample (174 cycles from 52 women) contains BF cycles from each women after she had been cycling for several months.

later non-BF
This sample contains all cycles where the women is not BF, and where the cycle start date is ≥ 180 days after the start date of her last BF cycle. This sample (73 cycles from 24 women) also included 4 cycles from one woman who never breastfed her children. The selection criteria for this sample are typical for cross-sectional samples.

In many cases the available data only allowed the timing of a cycle relative to the first post-partum menses or the last BF cycle to be determined to within some range of dates. Cycles were only included in the cross-sectional samples if all possible dates within the range satisfied the selection criteria.

Anovulation

When a woman first resumes cycling after giving birth, the full resumption of normal ovarian function may require several cycles. A consequence of this is that her first few post-partum cycles are more likely to be anovulatory than her later cycles. One might also expect that cross-sectional samples of women who have cycled for some time without conceiving (and are thus selected to be relatively less fecund) might show relatively high fractions of anovulatory cycles.

Figure 2 shows the anovulation rates (number of anovulatory cycles divided by total number of cycles) for our cross-sectional samples. The anovulation rate for the early BF sample (i.e., the first few post-partum cycles) is almost twice than of the later BF sample. However, contrary to expectation, the anovulation rate of the later non-BF sample is very similar to that of the later BF sample. In fact, essentially *any* cross-sectional sample of our data which begins at least 2 months after the first post-partum menses shows a similar anovulation rate ($\approx 25\%$).

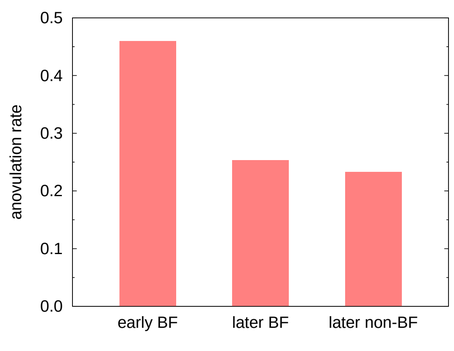

Figure 2: Anovulation rates of various cross-sectional samples.

Progesterone Indices

Figure 3 shows the distributions of $\langle\langle P_{\text{mpl}} \rangle\rangle$ for each of our cross-sectional samples. $\langle\langle P_{\text{mpl}} \rangle\rangle$ is relatively low for the early BF sample, higher for the later BF sample, and lower again for the later non-BF sample. Compared to the later BF sample, the mean $\langle\langle P_{\text{mpl}} \rangle\rangle$ is approximately 15% smaller in each of the other two samples. Moreover, very high $\langle\langle P_{\text{mpl}} \rangle\rangle$ values ($\geq$ 335 pmol/liter) make up fully $\frac{1}{6}$ of the later BF sample, but are completely absent in the other two samples. The biological significance of these differences is also magnified by the limited range over which $\langle\langle P_{\text{mpl}} \rangle\rangle$ varies in ovulatory cycles: As shown in Figure 3, if an individual cycle has $P_{\text{mpl}} < 110$ pmol/liter we consider the cycle to be anovulatory, so a decrease from 254 down to 215 pmol/liter actually represents a 27% drop in the amount ($\langle\langle P_{\text{mpl}} \rangle\rangle - 110$ pmol/liter) by which $\langle\langle P_{\text{mpl}} \rangle\rangle$ exceeds the anovulation threshold.

The difference in mean $\langle\langle P_{\text{mpl}} \rangle\rangle$ between the later BF and later non-BF samples is statistically significant ($t = 2.19$, 1-tailed $p = 0.016$).[2]

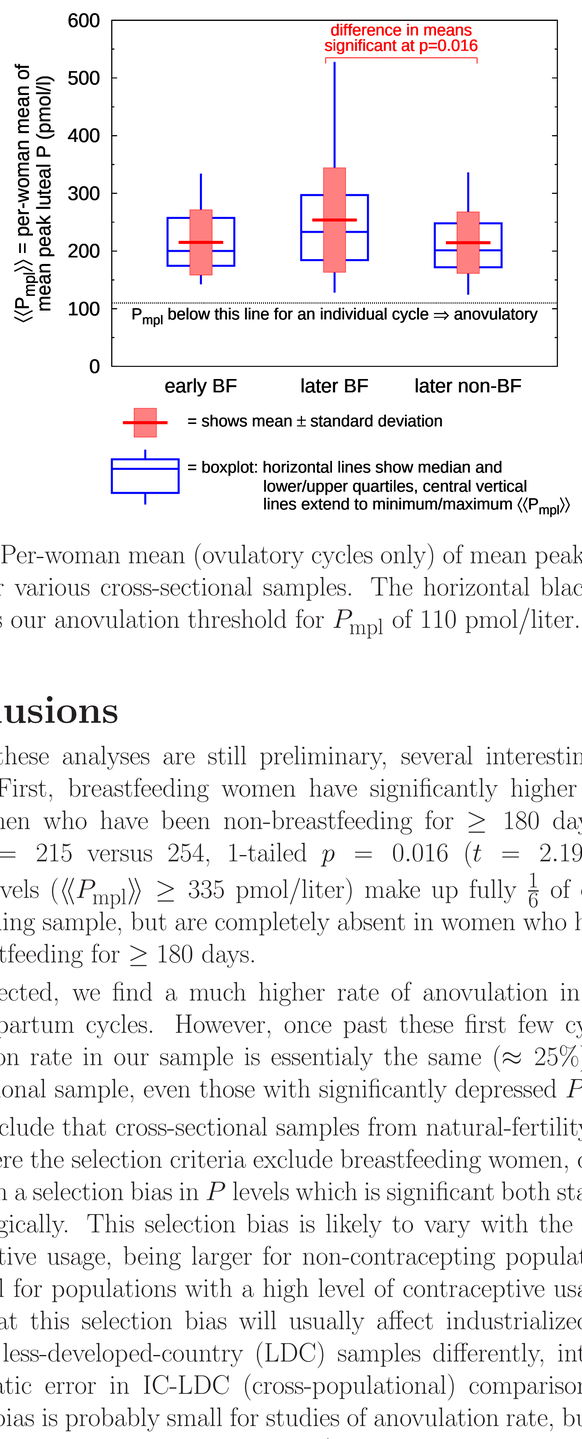

Figure 3: Per-woman mean (ovulatory cycles only) of mean peak luteal P (P_{mpl}) for various cross-sectional samples. The horizontal black dashed line shows our anovulation threshold for P_{mpl} of 110 pmol/liter.

Conclusions

While these analyses are still preliminary, several interesting results emerge: First, breastfeeding women have significantly higher P levels than women who have been non-breastfeeding for $\geq$ 180 days (mean $\langle\langle P_{\text{mpl}} \rangle\rangle = 215$ versus 254, 1-tailed $p = 0.016$ ($t = 2.19$). Very high P levels ($\langle\langle P_{\text{mpl}} \rangle\rangle \geq 335$ pmol/liter) make up fully $\frac{1}{6}$ of our main breastfeeding sample, but are completely absent in women who have been non-breastfeeding for $\geq$ 180 days.

As expected, we find a much higher rate of anovulation in the first few post-partum cycles. However, once past these first few cycles, the anovulation rate in our sample is essentialy the same ($\approx 25\%$) for *any* cross-sectional sample, even those with significantly depressed P levels.

We conclude that cross-sectional samples from natural-fertility populations, where the selection criteria exclude breastfeeding women, do indeed suffer from a selection bias in P levels which is significant both statistically and biologically. This selection bias is likely to vary with the extent of contraceptive usage, being larger for non-contracepting populations and very small for populations with a high level of contraceptive usage. This means that this selection bias will usually affect industrialized-country (IC) and less-developed-country (LDC) samples differently, introducing a systematic error in IC-LDC (cross-populational) comparisons. This selection bias is probably small for studies of anovulation rate, but further research is needed to determine if and/or how our results generalize to ovarian hormones other than P, to fecundability, or to other measures of ovarian functioning.

Acknowledgements

We thank E. Caceres and H. Spielvogel for invaluable assistance with Bolivian data collection, and the Instituto Boliviano de Biología de Altura for facilities and logistical support. Financial support was provided by the U.S. NSF, the University of California, Binghamton University, and Indiana University.

References

Dunson, D. B., B. Colombo, and D. D. Baird, 2002, Human Reproduction **17**(5), 1399.

Vitzthum, V. J., H. Spielvogel, E. Caceresc, and J. Gaines, 2000, Contraception **62**, 181.

Vitzthum, V. J., H. Spielvogel, and J. Thornburg, 2004, Procedings of the U.S. National Academy of Sciences **101**(6), 1443.

[1]These results are preliminary, and should not be cited without permission of the authors.

[2]Many women contributing to the early BF sample also contribute to the later BF sample, so (since different cycles' P_{mpl} from the same woman are correlated) these samples are not independent. This makes many standard statistical tests (for the significance of the $\langle\langle P_{\text{mpl}} \rangle\rangle$ difference) invalid, although a bootstrap method might be applicable. No woman contributed a cycle to both the later BF and later non-BF samples, so a t test *is* valid between these samples.