Human Reproduction, Vol.35, No.3, pp. 488–489, 2020
Advance Access Publication on February 12, 2020 doi:10.1093/humrep/deaa006

human reproduction

INVITED COMMENTARY

Fertility, season and a legitimate role for *P*-values

C.R. Weinberg*

Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences

*Correspondence address. MD A3-03, National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, NC 27709, USA. E-mail: weinberg@niehs.nih.gov

Submitted on December 16, 2019; resubmitted on December 16, 2019; editorial decision on January 14, 2020

Some mammals are seasonal breeders, with fertility depending on the light–dark cycle, but humans are considered to be continuous breeders. Whether human fecundability nonetheless varies with season has not been well studied. In this journal issue, Wesselink *et al.* (2019) report intriguing findings based on a large cohort of women attempting to conceive without fertility treatment in North America (the PRESTO study) and Denmark (the Snart Gravid study). Their analysis carefully adjusted for the seasonal patterns of the initiations of attempts to conceive, an important adjustment (Basso *et al.*, 1995), which earlier work had neglected. Based on the extensive data from the two cohorts, they report that fecundability was evidently higher in the fall and lower in the spring. Interestingly, the estimated amplitude of the effect was greater at lower latitudes.

The authors base their inference on the estimated peak-to-trough ratio of the seasonal shifted sine wave as their primary parameter, and they provide confidence intervals for that ratio rather than reporting classical *P*-values or significance tests. Generally, I sympathize with the rising preference for estimation (Amrhein *et al.*, 2019) because estimation can be both more meaningful and less prone to misinterpretation than significance tests. However, for studies of seasonality, I believe that *P*-values are needed.

First, as the authors acknowledge, their chosen parameter has an irritating feature: because the estimated peak is by definition necessarily higher than the estimated trough, the estimated peak-to-trough ratio is going to be greater than 1, even if there is no effect of season whatsoever. The simulations provided for the ratio estimates under a no-seasonality scenario partially address that issue but do little for inference under alternatives to the null. A joint confidence region for the two coefficients (of the sine and the cosine) in their model could have provided an informative alternative approach, would not have been subject to any boundary constraint, and would have also enabled the calculation of a confidence arc (days of the year being on a circle) for the location of the peak. However, both of those approaches (the peak-to-trough confidence interval and the two-parameter confidence region) require the strong assumption that the pattern is in fact a shifted sinusoidal, an assumption that was not assessed. If that implicit assumption is violated, then we cannot have confidence in confidence intervals.

In particular, the evidence from Denmark is obscured by the peak-to-trough analytic approach. One could easily compute a 2-degree-of-freedom chi-squared statistic (accompanied by the corresponding *P*-value) based on the improvement in fit provided by a model with the shifted sine wave compared with a null model representing the total absence of a seasonal effect (both coefficients equal to 0). That goodness-of-fit statistic would have provided a valid way to assess to what extent the Danish data were consistent with no seasonal effect. A point that is sometimes missed is that, unlike confidence intervals or confidence regions for specified parameters, goodness-of-fit statistics only require (for their validity) that we know the right model under the no-effect scenario. Figure 2 of the paper suggests that fecundability does seem to vary sinusoidally with the day of the year in North America, while it could be seasonal but not sinusoidal in Denmark.

These statistical issues raise legitimate concerns about the analyses that the authors present. Nevertheless, the observation that the estimated peaks occurred at similar times of the year in Denmark and in North America and that the least pronounced effects in North America were in the upper latitudes (closer to Danish latitudes) lends credibility to the findings and suggests to me that the data from the two regions tell compatible stories about fecundability and season.

Suppose then that there is a seasonal pattern to fecundability in the northern hemisphere. Whether that seasonal pattern is an artefact secondary to factors that influence both fertility and the timing of initiation of attempts (Basso *et al.*, 1995) or whether it is a real biologic effect of the light–dark cycle or of behaviours and exposures (such as diet) that influence fecundability (and vary with season) remains to be explored.

Funding

Federal government to C.R.W., NIH salary support.

Conflict of interest

C.R.W. has no conflicts of interest to report.

References

Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature* 2019;**567**:305–307.

Basso O, Olsen J, Bisanti L, Juul S, Bolden S. Are seasonal preferences in pregnancy planning a source of bias in studies of seasonal variation in reproductive outcomes? The European Study Group on Infertility and Subfecundity. *Epidemiology* 1995;**6**:520–524.

Wesselink A, Wise LA, Hatch EE, Mikkelsen EM, Sorensen HT, Riis AH, Mckinnon CJ, Rothman KJ. Seasonal patterns of fecundability in North America and Denmark: a preconception cohort study. *Hum Reprod* 2020; doi: https://doi.org/10.1093/humrep/dez265.