

Day-specific probabilities of clinical pregnancy based on two studies with imperfect measures of ovulation

D.B.Dunson^{1,3}, D.D.Baird², A.J.Wilcox² and C.R.Weinberg¹

¹Biostatistics Branch, National Institute of Environmental Health Sciences, MD A3-03, Research Triangle Park, NC 27709, and

²Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA

³To whom correspondence should be addressed

Two studies have related the timing of sexual intercourse (relative to ovulation) to day-specific fecundability. The first was a study of Catholic couples practising natural family planning in London in the 1950s and 1960s and the second was of North Carolina couples attempting to become pregnant in the early 1980s. The former identified ovulation based on the ovulatory shift in the basal body temperature, while the latter used urinary assays of hormones. We use a statistical model to correct for error in identifying ovulation and to re-estimate the length of the fertile window and day-specific fecundabilities. We estimate the same 6-day fertile interval in both studies after controlling for error. After adjusting for error both data sets showed the highest estimate of the probability of pregnancy on the day prior to ovulation and both fell close to zero after ovulation. Given that the fertile interval is before ovulation, methods that anticipate ovulation by several days (such as the assessment of cervical mucus) would be particularly useful for couples who want to time their intercourse either to avoid or facilitate conception.

Key words: basal body temperature/fecundability/fertile interval/ovulation/urinary metabolites

Introduction

Two large prospective studies provide data for estimating the probability of clinically detectable pregnancy with intercourse on particular days of the menstrual cycle relative to ovulation. The first study enrolled married British couples in the 1950s and 1960s who used the basal body temperature (BBT) method of natural family planning (Barrett and Marshall, 1969). Data were collected on dates of intercourse, and the day of ovulation was assumed to be the last day of hypothermia (estimated using the coverline rule applied to daily BBT measurements) (Barrett and Marshall, 1969). A total of 241 couples provided usable data.

The second study was done in the early 1980s with 221 healthy North Carolina couples who were attempting to become pregnant and were enrolled when they discontinued their birth control (Wilcox *et al.*, 1988). Each day women recorded

whether or not they had intercourse and collected a first morning urine specimen. The day of ovulation was estimated from the rapid decline in the ratio of oestrogen to progesterone that accompanies luteinization of the ovarian follicle, based on urinary hormone metabolites (Baird *et al.*, 1991). This steroid-based estimate of ovulation date is designated 'day of luteal transition' (DLT).

Data from these studies have been used to estimate the day-specific probabilities of clinical pregnancy and the length of the fertile interval. Day-specific pregnancy probabilities (Royston, 1982) were reported, based on the Barrett and Marshall data (Barrett and Marshall, 1969), using a previous model (Schwartz *et al.*, 1980). The estimated single-day probability increases to a peak of 0.36, 2 days prior to the last day of hypothermia. Intercourse as early as 8 days prior to the last day of hypothermia, and as late as 3 days afterwards apparently resulted in pregnancy. A similar pattern, but with a shorter interval and lower estimates was reported (Wilcox *et al.*, 1998). The estimated single-day probabilities of pregnancy peak 2 days prior to the estimated day of ovulation. The apparent fertile interval extends from ~5 days before the DLT to the DLT.

These estimates are sensitive to errors in identifying the ovulation date (Bongaarts, 1983). To illustrate this, imagine that pregnancy is possible only with intercourse on the day of ovulation, and with zero probability on all other days. If there is any error in estimating the day of ovulation, then the estimated day will be shifted by ≥ 1 days from the true day for some proportion of cycles. Some pregnancies will appear to result from intercourse before or after ovulation. The apparent pattern is consequently smeared, causing the estimated fertile interval to be artefactually extended. If such error could be corrected, estimates of day-specific probabilities would be made more accurate and studies using different markers of ovulation could be compared more meaningfully.

Dunson and Weinberg have extended the standard fertility model to allow for measurement error in identifying the day of ovulation (Dunson and Weinberg, 1999a). They propose a semiparametric Bayesian mixture model that can estimate the distribution of measurement errors and correct the estimates of fertility parameters for such errors. The purpose of this paper is to apply this approach to an analysis of the two fertility studies in order to: (i) compare the performance of the BBT and DLT measures of ovulation; (ii) estimate the day-specific probabilities of pregnancy and identify the fertile window, controlling for error in measuring ovulation; and (iii) compare the two patterns of day-specific probabilities of pregnancy.

Table I. Characteristics of study populations.

Characteristic	Barrett and Marshall	Wilcox <i>et al.</i> (1988)
Ovulation indicator	rise in BBT	DLT
No. of women	241	221
Percentage with previous pregnancy	100	64
Percentage >30 years of age	55	30
No. of cycles total	*	740
No. of cycles in analysis	2192	674
No. of clinical pregnancies	103	151

*Total no. cycles unknown.

BBT = basal body temperature; DLT = day of luteal transition.

Materials and methods

Description of study populations and cycle selection

Characteristics of the two study populations used in this analysis are summarized in Table I. The Barrett and Marshall study sample consisted of British married couples who had at least one child upon entering the study (Barrett and Marshall, 1969). Of the women, 90% were aged 20–39 years, with the rest aged 40–50 years. The couples were recruited upon seeking advice about natural family planning from the Catholic Marriage Advisory Council. Most were trying to avoid pregnancy at the start of follow-up. An unknown number of women who regularly produced temperature charts that were difficult to interpret were excluded from the study, as were individual cycles with no identifiable day of ovulation. The useable data consisted of 2192 menstrual cycles from 241 women. Pregnancy was reported in 103 cycles.

The Wilcox study sample (Wilcox *et al.*, 1988) consisted of North Carolina women who were planning to become pregnant and had no history of serious chronic illness or fertility problems. The majority of the women were college educated (71%) and white (96%). One third were nulliparous and 80% were aged 26–35. Only one was aged >40 years. The data consisted of 740 menstrual cycles from 221 women. Pregnancy was detected chemically in 199 of these cycles. Of the pregnancies, 48 were defined as early losses, since they ended within 6 weeks of the last menstrual period. The remaining 151 pregnancies survived long enough that they would likely have been detected by the methods used by Barrett and Marshall. These are designated clinical pregnancies. We restricted the analysis of the North Carolina study to these 151 clinical pregnancies (early losses were treated as non-conception cycles) in order to make the two studies comparable. We further restricted the analysis to menstrual cycles for which a day of ovulation could be identified and there were no relevant missing data on timing of intercourse. This left 674 out of the original 740 cycles (91%), and 141 of the 151 clinical pregnancies (93%).

Analytical method: modelling probability of pregnancy

Spermatozoa can remain viable in the female reproductive tract for several days or more (Perloff and Steinberger, 1964). Therefore, if there is intercourse on multiple days in a menstrual cycle where pregnancy occurs, the specific day of intercourse responsible for that pregnancy cannot be determined with certainty.

A method of estimating the daily probabilities of clinical pregnancy based on the assumption that batches of sperm introduced into the reproductive tract on different days mingle and compete independently has been proposed (Barrett and Marshall, 1969). Under this model the probability of a pregnancy in a given cycle is:

$$\Pr(\text{Pregnancy in menstrual cycle } j | X_{jk}) = 1 - \prod_k (1 - p_k)^{X_{jk}}$$

where X_{jk} is an indicator of intercourse on day k of cycle j , $j = 1, \dots, J$, and p_k is interpretable as the probability that pregnancy would occur with intercourse only on day k .

The Barrett and Marshall model only allows for timing of intercourse effects. This model was extended (Schwartz *et al.*, 1980) to allow the probability of clinical pregnancy to also depend on factors unrelated to timing of intercourse. These factors are summarized in a parameter (A) referred to as the ‘cycle viability’ probability, which is the probability that the aggregate of all factors not related to timing of intercourse are favourable to clinical pregnancy.

A complication in these studies is that most couples contribute more than one menstrual cycle to the data set and there is evidence of heterogeneity among couples in that some couples have a higher probability of cycle viability. This produces statistical dependency in the data. Also, less fertile couples contribute more cycles to the data set and therefore distort estimates of the mean fecundability. A random-effects model was proposed (Zhou *et al.*, 1996) that accounts for within-couple dependency in cycle viability. A similar model will be incorporated into the estimation in this paper.

Correcting for errors in estimating the day of ovulation

Most models implicitly assume that the day of ovulation is measured without error. When markers for ovulation are error-prone, the time index ‘ k ’ (denoting the day relative to ovulation) is not known precisely. One consequence is that studies with different methods for estimating ovulation are not estimating equivalent ‘ p_k ’ parameters, limiting comparability across studies. In a cycle where day of ovulation has been estimated incorrectly, the time between the true and assigned day of ovulation will be one or more days. The Zhou *et al.* (1996) model was extended (Dunson and Weinberg, 1999a) to allow for these errors by including the parameters π_l , denoting the probability of a shift of l days in the assigned day of ovulation relative to the true day of ovulation. We explain this model in greater detail in Appendix I.

Ideally, ‘day 0’ would be interpretable as the true day of ovulation after adjusting for measurement error. This would be the case if the assigned day of ovulation based on the marker does not systematically deviate from the true day of ovulation. There is evidence to suggest that the urinary luteinizing hormone (LH) peak (Collins *et al.*, 1983; France *et al.*, 1992) and the last day of hypothermia (France *et al.*, 1992) both occur close to ovulation on average. The DLT was identified based on an algorithm that was designed to be concordant with the day of the urinary LH peak (Baird *et al.*, 1991). Thus, on average both the DLT and the last day of hypothermia should approximate the true day of ovulation with little systematic bias.

Combining the two study populations

Once the intercourse indicators from both studies have been indexed to the corresponding estimated day of ovulation, a combined analysis of the two data sets can be carried out. We must also allow, however, for the possibility that the fecundability of the couples differs between the samples.

We begin with an analysis of each data set separately, comparing the cycle viability parameters (A) and the single-day pregnancy probabilities. In order to pursue the statistical comparison of results from the two studies, we made further simplifying assumptions. Based on the results of separate analyses of each data set, we can set up a parsimonious combined analysis by constraining a subset of the parameters to be equivalent in both studies while allowing for specific differences between the two cohorts. Each cohort is permitted its own distribution of errors. The performance of the two measures of

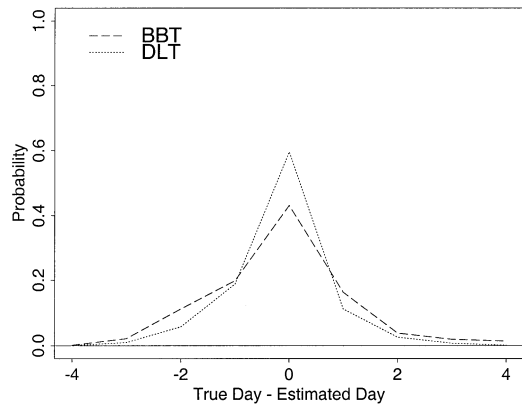


Figure 1. Estimated distribution of error in two markers of ovulation. The dashed line represents error in the last day of hypothermia, and the dotted line represents the error in the day of luteal transition (as estimated from urinary hormones). BBT = basal body temperature; DLT = day of luteal transition.

ovulation can be compared, by testing for a difference in the estimated proportion of cycles where ovulation has been assigned without error.

We first analyse each data set separately using the algorithm proposed by Dunson and Weinberg (1999a). We constrain the probability of pregnancy due to intercourse outside of a wide potential fertile window to be zero. We choose the potential fertile window based on the maximum likelihood estimates from the Schwartz model which does not adjust for measurement error (Schwartz *et al.*, 1980), presuming that the true window should be contained within the apparent window. All days with estimated (Schwartz model) single-day pregnancy probabilities (Ap_k) >0.01 are included in the window.

Based on this criterion the potential fertile window for the Barrett and Marshall cohort spans the 9-day interval from 7 days before to 1 day after the last day of hypothermia. The window is 6 days in the Wilcox *et al.* study, ranging from 5 days before to the day of the DLT.

The potential fertile window for the combined analysis is also identified based on estimates for the single day probabilities of clinical pregnancy (i.e. Ap_k). Since the model assumes that the day-specific probabilities are >0 , we must define a cut-off to constrain the width of the fertile interval. Days are included in the fertile window if the lower confidence bound for the probability of clinical pregnancy is >0.01 or the point estimate is >0.035 . After comparing the results based on separate analyses of the two cohorts, we adopt a more parsimonious model for a joint analysis: This model assumes that the day-specific p_k parameters are equal for the two cohorts, but allows the cohorts to have separate cycle viability parameters. Each of the two methods for assigning ovulation is allowed its own error distribution.

Results

Using the methods described above, we estimated the measurement error distributions corresponding to both the BBT-based marker of ovulation and the hormone-based marker of ovulation. The estimated error distributions are plotted in Figure 1. It appears that the hormone-based measure has less error than the BBT-based measure. According to these estimates, 60% of the DLT-estimated days of ovulation are correct, compared with 43% of the BBT-estimated days.

We use these error estimates to correct the day-specific pregnancy probabilities for error in identifying ovulation. In both studies, the maximum probability of pregnancy occurs

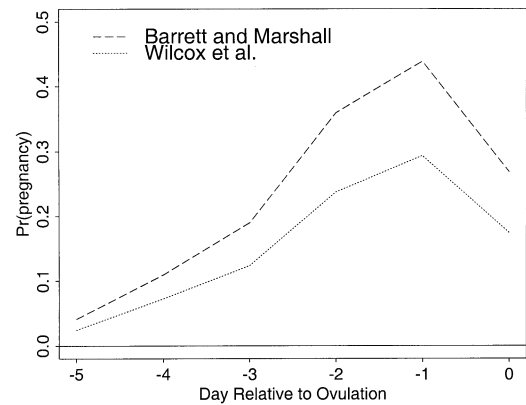


Figure 2. Estimated probability of achieving a clinical pregnancy based on a single act of intercourse in each study. The dashed line represents the estimates from the Barrett and Marshall cohort, and the dotted lines represent the estimates from the Wilcox *et al.* cohort.

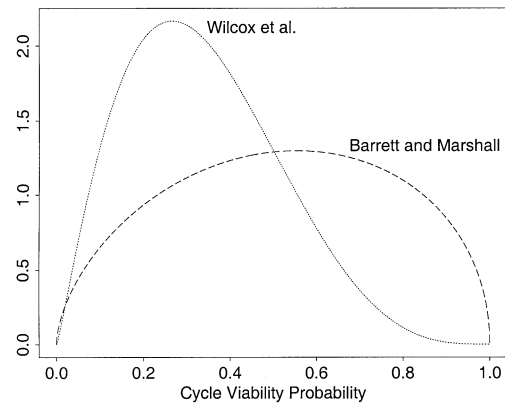


Figure 3. Estimated density function of cycle viabilities for couples in the Barrett and Marshall study (dashed line) and couples in the Wilcox *et al.* study (dotted line).

with intercourse one day prior to the estimated day of ovulation. The adjusted fertile interval for both studies starts ~ 5 days prior to ovulation and ends on the day of ovulation. The difference in the day-specific p_k parameters between the two cohorts is small. However, the mean cycle viability probability is substantially lower in the Wilcox *et al.* cohort (0.35 compared with 0.51).

Figure 2 shows the error-corrected day-specific probabilities of pregnancy for the Barrett and Marshall and Wilcox *et al.* cohorts based on the parsimonious pooled model described above. The cycle viability probability is significantly lower for couples in the Wilcox *et al.* cohort ($P < 0.01$). The distribution of cycle viabilities for couples in each study are shown in Figure 3. It appears that the heterogeneity among couples in fecundability is higher in the Barrett and Marshall cohort than in the Wilcox *et al.* cohort.

Discussion

We have analysed data from two prospective human fertility studies to compare the performance of two methods of estimating ovulation, to describe the day-specific pattern of pregnancy probabilities, and to improve the estimate of the fertile interval.

It appears that the DLT measure of ovulation is less error-prone than the BBT-based measure. The actual error in using the rise in BBT may be greater than we estimate: Barrett and Marshall discarded an unknown number of cycles because the temperature charts were considered uninterpretable. BBT has commonly been found to identify ovulatory cycles as anovulatory (Kesner *et al.*, 1992) and it was found that the variance of a BBT-based marker relative to a urinary LH-based marker was greater than that for a hormonal measure based on the ratio of oestrogen and progesterone (Royston, 1991). Therefore, it is not surprising that measures of ovulation based on urinary metabolites show more reliability than the measure based on basal body temperature (Vermesh *et al.*, 1987; Kesner *et al.*, 1992).

Errors in measuring ovulation distort estimates of the day-specific probabilities of pregnancy and extend the apparent length of the fertile interval. Controlling for measurement error, our analysis suggests the fertile interval starts ~5 days prior to ovulation and ends on the day of ovulation (although we cannot rule out small probabilities beyond these limits). This 6-day interval is the same as the uncorrected estimate from the North Carolina study (Wilcox *et al.*, 1998), but is much shorter than the nine days reported (Royston, 1982) for the Barrett and Marshall data. The two studies are in good agreement with regard to both the length and location of the fertile interval. Our estimate of the fertile interval coincides with the absence of contraceptive Glycodelin A (GdA) in the uterus (Mandelin *et al.*, 1997; Seppala *et al.*, 1998), suggesting that GdA may play a fundamental role in regulating the fertile interval.

The estimated probability of clinical pregnancy is highest on the day prior to ovulation. The correction for ovulation measurement error in the Barrett and Marshall data reduced the estimated probability of pregnancy to near zero after the day of ovulation, consistent with the result previously reported with the (uncorrected) analysis of the Wilcox data (Wilcox *et al.*, 1995, 1998). This suggests that the oocyte has a very short viability after ovulation and/or that spermatozoa deposited in the reproductive tract after ovulation are unable to reach the oocyte.

The finding that the estimated peak of fecundability is on the day before ovulation differs from results previously reported (Wilcox *et al.*, 1995) showing fecundability peaking on the day of ovulation. The earlier analysis included both early losses and clinical pregnancies, while we use only clinical pregnancies. If intercourse occurs on the day of ovulation then the egg may have aged at the time of fertilization. This has been suggested as an explanation for the apparently high probability of early loss found for conceptions resulting from intercourse on the day of ovulation (Wilcox *et al.*, 1998), a possibility that could explain the difference between the reported patterns.

Couples having difficulty conceiving often try to time their intercourse to optimize their chances. Given that the highest conception rates occur on the 2 days prior to ovulation, it is important to use a signal that allows couples to time intercourse for the several days of fertility before ovulation. The basal body temperature shift comes too late. Urinary LH kits only

identify the short time from the start of the urinary LH surge to ovulation (Collins *et al.*, 1983). Cervical mucus change provides an earlier and more useful cue. Mucus receptivity begins several days before ovulation (Katz *et al.*, 1997) so couples who have frequent intercourse after this cue will tend to have intercourse on those days with the highest probabilities of clinical pregnancy.

Day-specific estimates of fecundability were significantly lower in the Wilcox data than in the Barrett and Marshall data. There are several possible explanations. It is possible that this reflects differences in the spermatozoa between males in the two populations. A more likely possibility is that the selection of cycles for analysis may have distorted the apparent fecundability in the two cohorts. In both studies, some cycles were excluded from the analysis. In the Barrett and Marshall study, an unknown (but possibly large) number of temperature charts were discarded because they were difficult to interpret. If those discarded cycles were more likely to come from non-pregnancy cycles (e.g. cycles with erratic temperature charts tend to be less fertile), then the estimated fertility based on the non-discarded cycles would be biased upward. Only a small number of the discarded cycles from the Wilcox *et al.* study were anovulatory or hormonally abnormal cycles. The majority of the excluded cycles were discarded because of days with missing coital records (that is, the woman did not mark either 'yes' or 'no' for intercourse on a relevant day). The Barrett and Marshall data are even less informative in this way, since women marked only the days on which they had intercourse, leaving no way to distinguish 'no' from missing data. The possibility that some acts of intercourse were not recorded produces another potential source of upward bias in estimates of the daily probabilities based on the British data (Dunson and Weinberg, 1999b).

It is also possible that couples in the Barrett and Marshall cohort that had intercourse during the fertile interval were more fecund than couples who only had intercourse outside the interval. Since most of the couples in the British study were trying to avoid pregnancy, couples that had intercourse during the fertile interval may have been unable to abstain for a long enough number of days. If these high libido couples are more fertile, then this self-selection to high risk behaviour would create an upward bias in estimates of the daily pregnancy probabilities based on couples attempting to use abstinence to avoid conception.

Other factors related to fecundability also differ between the two study groups. The British couples had all been pregnant before, whereas about a third of the North Carolina couples were attempting pregnancy for the first time so they were of unproven fertility. The North Carolina couples were all attempting to conceive, while the British groups included couples having accidental pregnancies and these are more likely to occur to the more fecund couples.

In summary, the methods applied in this paper can be used to correct for bias in estimating the fertile interval and day-specific pregnancy probabilities, to compare the fecundability in multiple populations, and to compare the performance of available measures of ovulation. If error in determining the day of ovulation is not accounted for, estimates of the fertile

interval and the day-specific pregnancy probabilities will be dependent on the method of assessing ovulation, e.g. different methods of estimating ovulation will often yield different conclusions. A large European study now underway collects data on both basal body temperature and self-assessed changes in cervical mucus. Using the last day of hypothermia based on BBT measurements as the marker, preliminary estimates of the day-specific pregnancy probabilities for the ongoing study are as high as 0.04 across the interval from 8 days before to 2 days after the estimate of ovulation (Masarotto and Romualdi, 1997). It is likely that this apparent 11-day window would shrink drastically if measurement error were accounted for. Future analyses correcting for errors in identifying ovulation could compare fecundabilities across countries in this multinational effort, compare alternative ovulation detection methods to DLT and rise in BBT, as well as compare the fertility parameters of this new cohort to those of the cohorts described here.

Acknowledgements

The authors would like to thank Dr Glinda Cooper and Dr Haibo Zhou for their careful reading of the manuscript.

References

- Baird, D.D., Weinberg, C.R., Wilcox, A.J. *et al.* (1991) Using the ratio of urinary oestrogen and progesterone metabolites to estimate day of ovulation. *Stat. Med.*, **10**, 255–266.
- Barrett, J. C. and Marshall, J. (1969) The risk of conception on different days of the menstrual cycle. *Pop. Studies*, **23**, 455–461.
- Bongaarts, J. (1983) The proximate determinants of natural marital fertility. In Bulatao, R.A., Lee, R.D., Hollerbach, P.E. and Bongaarts, J. (eds), *Determinants of Fertility in Developing Countries*. Vol. 1. Academic Press, New York, USA, pp. 103–138.
- Collins, W.P., Branch, C.M., Collins, P.O. and Sallam, H.M. (1983) Biochemical indices of the fertile period in women. *Int. J. Fertil.*, **26**, 196.
- Dunson, D.B. and Weinberg, C.R. (1999a) Modeling human fertility in the presence of measurement error. *Biometrics*, in press.
- Dunson, D.B. and Weinberg, C.R. (1999b) Accounting for unreported and missing intercourse in human fertility studies. *Stat. Med.*, in press.
- France, J.T., Graham, F.M., Gosling, L. *et al.* (1992) Characteristics of natural conception cycles occurring in a prospective study of sex preselection: fertility awareness symptoms, hormone levels, sperm survival, and pregnancy outcome. *Int. J. Fertil.*, **37**, 244–255.
- Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. (eds), *Bayesian Statistics*. Vol. 4. Clarendon Press, Oxford, UK, pp. 169–193.
- Katz, D.F., Slade, D.A., and Nakajima, S.T. (1997) Analysis of pre-ovulatory changes in cervical mucus hydration and sperm penetrability. *Adv. Contracept.*, **13**, 143–151.
- Kesner, J.S., Wright, D.M., Schrader, S.M. *et al.* (1992) Methods of monitoring menstrual function in field studies: efficacy of methods. *Reprod. Toxicol.*, **6**, 385–400.
- Mandelin, E., Koistinen, H., Koistinen, R. *et al.* (1997) Levonorgestrel-releasing intrauterine device-wearing women express contraceptive glycodeclin A in endometrium during midcycle: another contraceptive mechanism? *Hum. Reprod.*, **12**, 2671–2675.
- Masarotto, G. and Romualdi, C. (1997) Probability of conception on different days of the menstrual cycle: an ongoing exercise. *Adv. Contracept.*, **13**, 105–115.
- Perloff, W.H. and Steinberger, E. (1964) *In vivo* survival of spermatozoa in cervical mucus. *Am. J. Obstet. Gynecol.*, **88**, 439–442.
- Royston, J.P. (1982) Basal body temperature, ovulation, and the risk of conception, with special reference to the lifetimes of sperm and egg. *Biometrics*, **38**, 397–406.
- Royston, P. (1991) Identifying the fertile phase of the human menstrual cycle. *Stat. Med.*, **10**, 221–240.
- Schwartz, D., MacDonald, P.D.M. and Heuchel, V. (1980) Fecundability, coital frequency, and the viability of ova. *Pop. Studies*, **23**, 455–461.
- Seppala, M., Koistinen, H., Mandelin, E. *et al.* (1998) Glycodeclins: role in regulation, potential for contraceptive development and diagnosis of male infertility. *Hum. Reprod.*, **13** (Suppl. 3), 262–269.
- Vermesh, M., Kletzky, O.A., Davajan, V., Israel, R. (1987) Monitoring techniques to predict and detect ovulation. *Fertil. Steril.*, **47**, 259–264.
- Wilcox, A.J., Weinberg, C.R., O'Connor, J.F. *et al.* (1988) Incidence of early loss of pregnancy. *N. Engl. J. Med.*, **319**, 189–194.
- Wilcox, A.J., Weinberg, C.R., and Baird, D.D. (1995) Timing of sexual intercourse in relation to ovulation. *N. Engl. J. Med.*, **333**, 1517–1521.
- Wilcox, A. J., Weinberg, C. R., and Baird, D. D. (1998) Post-ovulatory ageing of the human oocyte and embryo failure. *Hum. Reprod.*, **13**, 394–397.
- Zhou, H., Weinberg, C.R., Wilcox, A.J., and Baird, D.D. (1996) A random-effects model for cycle viability in fertility studies. *J. Am. Stat. Assoc.*, **91**, 1413–1422.

Received on September 8, 1998; accepted on March 30, 1999

Appendix I. Accounting for Errors in Ovulation

Methods

Under the Schwartz *et al.* (1980) model, the probability of pregnancy for cycle j conditional on a shift of l days is

$$A \left[1 - \prod_k (1 - p_k)^{X_{j,k+l}} \right]$$

With incorporation of errors, as proposed by Dunson and Weinberg (1999a) the observed data likelihood is:

$$\sum_l \left\{ \prod_j \left[A \left(1 - \prod_k (1 - p_k)^{X_{j,k+l}} \right) \right]^{Y_j} \times \left[1 - A + A \left(\prod_k (1 - p_k)^{X_{j,k+l}} \right) \right]^{1 - Y_j} \times \pi_l \right\},$$

where Y_j is 1 if pregnancy occurred in cycle j and 0 otherwise and π_l is the probability that the identified day of ovulation is l days before the true day of ovulation.

We make several simplifying assumptions. First, we assume that the day-specific probabilities of pregnancy are 0 outside of a fertile window. Then we assume that, within the fertile window, the probabilities increase to a peak and then decrease. The error probabilities, π_l are also assumed to be 0 outside of a window. They are constrained to decrease away from a peak at $l = 0$. In order for the p_k parameters to be interpretable as probabilities relative to the true day of ovulation, it is necessary to assume that the most likely difference between the estimated day of ovulation and the true day of ovulation is known. This difference can hypothetically be verified using data from validation studies that record both the day of follicular rupture and the day estimated using the marker. The estimated p_k parameters and fertile interval are valid even if this difference is misspecified. However, the k subscripts will be systematically shifted. Within-couple correlation is accounted for using a beta-binomial random-effects model (Lee and Sabavala, 1987; Zhou *et al.*, 1996).

Analysis

The Markov Chain Monte Carlo (MCMC) algorithm proposed in Dunson and Weinberg (1999a) can be applied directly with the addition on a Metropolis step to estimate β . We assign β a diffuse prior distribution. The algorithm is iterated 120 000 times and the first 10 000 samples are discarded. Convergence is verified using Geweke's diagnostic (Geweke, 1992).