



Bayesian Inference on Predictors of Sex of the Baby

Bruno Scarpa*

Department of Statistical Sciences, University of Padua, Padua, Italy

It is well known that the sex ratio at birth is a biological constant, being about 106 boys to 100 girls. However couples have always wanted to know and decide in advance the sex of a newborn. For example, a large number of papers appeared connecting biometrical variables, such as length of follicular phase in the woman menstrual cycle or timing of intercourse acts to the sex of new baby. In this paper, we propose a Bayesian model to validate some of these theories by using an independent database. Results show that we could not show an effect of the follicular length on the sex of the baby. We also obtain a slightly larger probability, although not significant, of conceiving a female just after the mucus peak day.

Keywords: sex of the baby, Bayesian hierarchical model, aggregated Bernoulli, human fertility

OPEN ACCESS

Edited by:

Rafael T. Mikolajczyk,
Helmholtz Center for Infection
Research, Germany

Reviewed by:

Oscar Gonzalez-Recio,
Department of Environment and
Primary Industries, Australia
Antonia Zapf,
University Medical Center Göttingen,
Germany

*Correspondence:

Bruno Scarpa
scarpa@stat.unipd.it

Specialty section:

This article was submitted
to Population, Reproductive
and Sexual Health,
a section of the journal
Frontiers in Public Health

Received: 27 October 2015

Accepted: 03 May 2016

Published: 24 May 2016

Citation:

Scarpa B (2016) Bayesian Inference
on Predictors of Sex of the Baby.
Front. Public Health 4:102.
doi: 10.3389/fpubh.2016.00102

1. INTRODUCTION

A number of researchers have been investigating if the biological system determining the sex of a newborn is completely random or if it could be predetermined. The starting point is clearly the fact that, although the proportions of sperm conducting chromosomes X and Y are the same, the number of male births is slightly higher than the number of females (1).

Many studies analyzed the relation between sex of the newborn and biological and social factors, such as length of follicular phase (2), pattern of intercourse (3–7), diet of the woman (8, 9), behavior during coital act (10), parents age (11, 12), historical events (11), atmospheric temperature (13), environmental pollution (14), and so on. Despite this very large set of researches in this field, we observed a lack of validation studies based on independent data. Most of these studies are based on really small samples and are difficult to replicate, by requiring specific and expensive epidemiological studies to be run.

In this paper, we are interested in evaluating the effect of two biometrical variables on the sex of the baby, which have been studied extensively in the past: the length of the follicular phase and the pattern of intercourse. Weinberg et al. (2) found that cycles with shorter follicular phase lengths are slightly more likely to result in male babies, while cycles with longer follicular phases are more likely to result in female babies. This theory, however, has been discussed by Gray et al. (15), concluding that there is no association between follicular phase length and sex ratio.

Even more debated is the other hypothesis we are analyzing. Kleegman (3) first supposed that if conception occurs in the day of ovulation the chances that the baby is a male are about 80%, while if the intercourse occurs 2 days earlier the same chances are given to a female baby. Some years later, in three different studies, Guerrero (4–6) shows that the probability to have a male is higher if intercourse occurs far from the ovulation. Billings and Westmore (7), supported by a study by McSweeney (16), instead, state that if intercourse occurs the day of ovulation or the following ones, the probability to have a male is higher, while if intercourse occurs early in the fertile window there are more chances to get a female. Other authors' contribution to this debate were Shettles (10)

and Perez et al. (17) each of them with his own position. From this short review, it is clear that different studies give completely different results, and we need an independent dataset to test such theories.

The validation or confutation of these theories is made difficult by the fact that, typically, multiple intercourse acts occur during a cycle, and it is not known which is the one responsible for the conception. As discussed in many papers [e.g., Ref. (18–21)], in such a case, a statistical model is needed to analyze available data, and exploit all cycles and not only the very small number of one-intercourse cycles.

The sex of the baby can be represented by the aggregation across Bernoulli trials for each intercourse day (22). Because of this aggregated Bernoulli data structure and the need to account for dependency among the multiple cycles from a woman, statistical analyses can be challenging to implement.

In this paper, we propose the use of a Bayesian model for the probability to conceive a female (or a male) and we evaluate the effect of day of intercourse and length of the follicular phase on the sex of the baby by using a large dataset obtained by the union of two fecundability studies led by Colombo (19, 20). We obtain posterior distributions of the parameters, by modifying the Gibbs sampling algorithm proposed by Dunson and Stanford (21) to model conception probabilities.

In Section 2, we present the dataset obtained from the two fecundability studies, Section 3 proposes the model and priors, Section 4 applies the model to data and Section 5 discusses the results. All the analyses in this paper have been implemented in R, a language and environment for statistical computing (23); code is available upon request to the author.

2. MATERIALS AND METHODS

2.1. Two Studies of Daily Fecundability

In order to develop and fit a model relating biometrical variables and sex of the baby, it is necessary to have very complete data on biomarkers and intercourse days for a large number of menstrual cycles that ended up with a conception. Unfortunately, such data are difficult to obtain, and only a small number of available studies include detailed and reliable information on couples' behaviors. Also, the majority of cycles collected in these studies do not end up with a conception, so the number of useful cases considered to evaluate sex of the baby in each study is quite small. For this reason, we focus on the union of data collected from two different studies, having quite a similar design and coordinated by the same researcher.

2.1.1. European Fecundability Study

The European Study of Daily Fecundability (19) collected data on 782 couples, which were married or in a stable relation, from seven European centers (Milan, Verona, Lugano, Dusseldorf, Paris, London, and Brussels) providing services on fertility awareness and natural family planning by using symptothermal methods. Women enrolled were between 18 and 40 years of age, were not taking hormonal medications or drugs affecting fertility, had no known impairment of fecundity and had at least one

menses after cessation of breastfeeding or after delivery and were experienced in the use of one method of natural family planning. The participants were followed prospectively during one or more menstrual cycles, as they recorded daily basal body temperatures, vaginal observations from cervical mucus, and the days during which intercourse and menstrual bleeding occurred. The women had received training at the study centers on how to identify different types of sensation and mucus. Teachers classified each day of the cycle according to a four-point scale according to the type of mucus described by women.

Additional 99 subjects from a different study (24, 25), carried out in Auckland (New Zealand), were included retrospectively in view of their relevance to the aims of the study. As discussed by Colombo and Masarotto (19) recruitment and instruction to the women was quite similar to the one for the European Study. The main specificity in this case was that study design restricted the couples to only one act of intercourse during the fertile phase of the cycle that is a suitable window of days around ovulation.

Data on a total of 7017 (6724 from the European centers and 293 from the New Zealand study) menstrual cycles have been collected, with 453 of these cycles (375 cycles from European centers and 78 from the New Zealand study) ending up in a conception. The average age in years of the women was 29.4 (SD = 4.15).

By using the available data, cervical mucus peak may be obtained as proxy for the ovulation day. The cervical mucus peak day was defined as the last day with best quality mucus by sensation or appearance, in a specific cycle of the woman, known retrospectively. This peak day was taken as "Mucus reference day" and identified as day 0. The number of days between the first day of menstruation and the mucus reference day for each cycle is a proxy for the length of the follicular phase. By considering this indicator, the average length of the follicular phase was, restricting the dataset to cycles conceiving a baby, 17.0 days (SD = 5.6), while the same quantity for cycles conceiving a male was 16.98 days (SD = 5.81) and for cycles conceiving a female was 17.01 days (SD = 5.40).

2.1.2. Italian Fecundability Study

A second prospective study of fecundability (20) was led by the same researcher and carried out in four Italian centers providing services on fertility awareness and natural family planning by using the Billings ovulation method (26) and was concentrated in estimating the effect of cervical mucus symptoms on the probabilities of conception. This study enrolled 191 women providing data on 2536 menstrual cycles, with 141 of these cycles ending in a conception. Women enrolled were between 18 and 40 years of age, were married or in stable relationship, had at least one menses after cessation of breastfeeding or after delivery, were not taking hormonal medications or drugs affecting fertility, and were experienced in the use of the Billings ovulation method of natural family planning. The average age in years of the women was 29.96 (SD = 4.18). Also in this case, the participants were trained at the study centers on how to identify different types of sensation and mucus and collected detailed daily records of vulvar observations of the cervical mucus symptom, and recorded the days during which intercourse and menstrual bleeding occurred.

Teachers classified each day of the cycle according to a five-point scale according to the type of mucus symptom described by women and identified the Billings mucus peak defined as “the last day of the cycle during which at least one characteristic of high fertility in mucus type is observed or felt; this day must also be preceded by an adequate increase in sensations and the appearance of mucus characteristics, and followed by an abrupt change” (20). Ovulation is expected approximately within 2 days of the peak, and this day was used as a reference to determine the end of the fertile phase. The average length of the follicular phase was, restricting to cycles conceiving a baby, 17.24 days (SD = 6.49) and the same quantities for cycles conceiving a male was 17.20 days (SD = 7.21) and for cycles conceiving a female was 17.28 days (SD = 5.69).

The study designs of both, the European and the Italian Fecundability, studies are very similar and some analysis (not presented here) have shown that women enrolled in these different studies have reasonably similar characteristics. Since we are interested in identifying the fertile window by using the same criteria, the main problem in joining data from the two studies is related to the indicator for the day of ovulation used. Although in both studies an indicator based on cervical mucus is used, its definition is slightly different between the two studies, in particular because the classification of mucus (in the European study) and mucus symptoms (in the Italian study) are different according to the different methods of natural family planning. However, we analyzed and compared the distributions of the mucus peak, and consequently of the follicular phase, as estimated in the two studies by using in both cases mucus peaks, and we did not found significant differences. For example, the qqplot comparing the distributions of the cervical mucus peak and of Billings mucus peak collected by the two studies

is plotted in **Figure 1** and shows that the two distributions are not different. In a qqplot, the quantiles of two distributions are plotted: if these points are on the diagonal, this means that the two distributions are equal; the points out of the diagonal show some difference between the distributions. We concluded that the two indicators are similar enough to be considered as one. Note that similar analysis comparing the cervical mucus peak and other ovulation indicators, such as the rise in basal body temperature, shows significant differences (27, 28). We, therefore, considered as mucus peak (MP) the cervical mucus peak for the European study and the Billings mucus peak for the Italian study.

By joining the two studies and selecting only the cycles with conceptions for which the MP was observed, we obtained a total of 521 menstrual cycles that we analyzed in this paper. **Table 1** shows the number of cycles with intercourse for each day in the fertile window for the joint dataset and its frequency over the total number of available cycles. Clearly, a cycle may show an intercourse act in more than 1 day.

In addition to the day of intercourse, we are interested in analyzing the length of follicular phase as predictor of the sex of the baby, by testing the hypothesis discussed by Weinberg et al. (2). These authors show that the relation between sex ratio and follicular phase length is not linear and likely not even monotone. By using their results and to allow our model to include possible non-linearities, we analyze this relationship by considering four classes of follicular phase lengths: shorter of 13, between 13 and 16, between 16 and 19, and longer than 19 days. **Table 2** shows the frequency distribution of the categorized follicular phase length.

Both of the data sets are available for researches complying with the objectives of the study design, which are stated in the

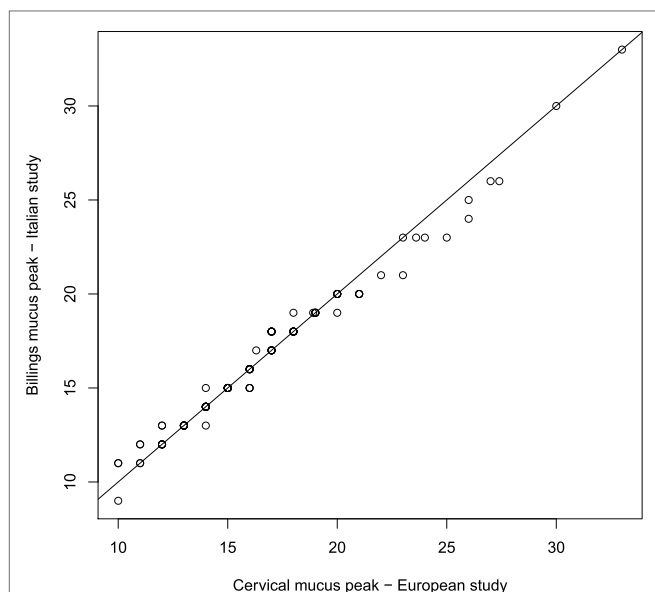


FIGURE 1 | qq-plot comparing the distributions of cervical mucus peak collected in the European study and the Billings mucus peak collected in the Italian study.

TABLE 1 | Number of cycle with intercourse for each day in the fertile window and its relative frequency over the total number of available cycles (521).

Intercourse day	Absolute frequency	Relative frequency (%)
-8	106	20.34
-7	92	17.66
-6	124	23.80
-5	133	25.53
-4	153	29.37
-3	120	23.03
-2	165	31.67
-1	154	29.56
0	203	38.96
1	128	24.57
2	100	19.19
3	97	18.62

TABLE 2 | Observed distribution, absolute and relative, of categorized follicular length.

Follicular length (days)	Frequency	Percentage (%)
≤13	147	28.21
14–16	155	29.75
17–19	119	22.84
>19	100	19.19

website <http://www.stat.unipd.it/en/fare-ricerca/basi-di-dati> where, also, rules to have access to the data are explained.

2.2. The Bayesian Model and Priors

An intercourse act can result in conception only if it occurs in a mid-cycle window (29). Therefore, only intercourse acts during the fertile interval may impacts on the sex of the baby. Moreover, there may be multiple intercourse acts that occur during the potentially fertile phase of the cycle and we need a statistical model to allocate day-specific probabilities of conceiving a male or a female. In addition, we would like to evaluate the effect of covariates, such as the length of the follicular phase.

In order to predict conception probabilities, by considering what is typically done to predict probability of conception [e.g., Ref. (18–21)], we propose a model to predict the probabilities of conceiving a male or a female, which only considers cycles with conception. Let Y_i be an indicator of the sex of the newborn (0 = male and 1 = female) in conception cycle i ($i = 1, \dots, n$), $X_i = [X_{i1}, \dots, X_{iK}]^T$ a vector of intercourse indicators for days 1, ..., K (sometimes it may be useful to index days relatively to the day of ovulation), and $U_i = [u_{i1}, \dots, u_{iH}]^T$ a covariate H -vector for cycle i , in our case we consider the length of the follicular phase discretized in classes. The total number of parameters involved in the model is, therefore, the sum of K parameters related to the days of intercourse and H parameters related to other covariates possibly available (in our case three parameterizing the four classes of follicular length).

Following Barrett and Marshall (18), we assume independence of intercourse acts on different days and we replace their day-specific probability of conception with the day-specific probability of conceiving a male, only considering conceiving cycles. The probability of having a male for a given conceiving cycle is given by the probability of not conceiving a female. The probability of conceiving a female is the product of the probabilities of not conceiving a male in each day when an intercourse act was observed. Our model, therefore, is

$$P(Y_i = 0 | X_i, u_i) = 1 - \prod_{k=1}^K (1 - q_{ik})^{X_{ik}} \tag{1}$$

where q_{ik} is the day-specific probability of conceiving a male in cycle i given intercourse only on day k .

Following Dunson and Stanford (21), the relation between q_{ik} and available covariates U_i is modeled by

$$q_{ik} = 1 - \exp(-\exp(v_{ik}^T \beta)) \tag{2}$$

where $\beta = [\beta_1, \dots, \beta_K, \beta_{K+1}, \dots, \beta_{K+H}]^T$ is a vector of regression coefficients and $V_i = [v_{i1}, \dots, v_{ik}, v_{ik+1}, \dots, v_{iK+H}]^T$ is a $(K + H)$ -vector related to cycle i . We follow the same authors by considering the first K elements of vector V_i as indicators of the day of intercourse observed in cycle i . Therefore, in this case, we define $\lambda_k = \exp(\beta_k)$ for $k = 1, \dots, K$, as baseline parameters indicating the day-specific probabilities of conceiving a female in each day of the fertile interval with an intercourse. The remaining H elements of V_i are the covariates included in U_i and $\gamma_h = \exp(\beta_{K+h})$, $h = 1, \dots, H$, are the

parameters characterizing the changes across levels of categorical covariates. In particular, if we consider as a covariate the length of follicular phase in classes, which is an ordered categorical predictor $z_i = 1, \dots, d$ (in our case $d = 4$), the parameters $\lambda_1, \dots, \lambda_K$ are baseline parameters characterizing the distribution of Y_i for subjects with $z_i = 0$ for days $k = 1, \dots, K$, while γ_h characterizes the multiplicative change in the probability to conceive a female attributable to increasing z_i from h to $h + 1$. We, therefore, may re-write (2) as

$$p_{ik} = 1 - q_{ik} = \lambda_k \prod_{h=1}^{z_i-1} \gamma_h,$$

where here p_{ik} is the day-specific probability of conceiving a female in cycle i given intercourse only on day k .

Note that in this case, λ_k is a parameter related to day-specific covariates, while γ_h is a parameter of cycle-specific covariates, outlining that, differently from the other proposed models (1, 2) for sex of the baby, our model may include both day-specific and cycle-specific covariates, by also allowing cycles for multiple intercourse days.

The Bayesian approach for the estimate of the coefficients λ_k and γ_h require the specification of prior distributions for the parameters included in the model. As done by Dunson and Stanford (21) in a similar context, we chose gamma priors, which are conditionally conjugate for each of the parameters λ_k 's and one-inflated gamma distributions as priors for the γ_h 's, allowing for a positive probability of no effect of the cycle-specific covariates. The joint posterior distribution for all the parameters is summarized as follows

$$\left\{ \prod_{k=1}^K \mathcal{G}(\lambda_k; a_{0k}, b_{0k}) \right\} \left\{ \prod_{h=1}^{d-1} (I_1 - \mathcal{G}(\gamma_h; \pi_h, a_{0h}, b_{0h})) \right\} \prod_{i=1}^n \left(\prod_{k=1}^K \left(\lambda_k \prod_{h=1}^{z_i-1} \gamma_h \right)^{X_{ik}} \right)$$

where $\mathcal{G}(a, b)$ denotes the gamma density with mean a/b and variance a/b^2 and $I_1 - \mathcal{G}(\gamma_h; \pi_h, a_{0h}, b_{0h})$ the one-inflated gamma density consisting of the mixture of a point mass at one (with probability π_h) and a $\mathcal{G}(a_{0h}, b_{0h})$ density. Posterior computation is based on the introduction of an auxiliary vector of independent Poisson latent variables following log-linear gamma frailty models, as already done in the algorithm proposed by Dunson and Stanford (21), and proceed with a Gibbs sampling algorithm (see Appendix). The specification of the covariate and of the priors for the prediction of the sex of the baby will be discussed in the next section.

3. RESULTS

We used the model described in expression (1) to relate the length of follicular phase and the timing of intercourse in the fertile interval to the probability of conceiving a female. As fertile interval, following Colombo and Masarotto (19), we consider the 12 days window included between 8 days before and 3 days after MP. Therefore, we let

$$U_{ik} = [I(k=1), \dots, I(k=12), I(13 < w_i \leq 16), I(16 < w_i \leq 19), I(w_i > 19)]^T$$

where k indexes the day in the fertile interval, with $k = 1$ eight days prior to ovulation, identified by MP, and $k = 12$ three days after the identified day of ovulation. Here w_i is the length of the follicular phase of the i th cycle and we considered changes across the four classes of follicular length, cycles shorter or equal to 13 days, between 14 and 16, between 17 and 19, and larger than 19. However, following Dunson and Stanford (21), we parameterize the follicular length classes by considering the increasing effect of passing from one length class to the following one. The regression coefficients are divided into two subvectors $\beta = (\beta_1, \beta_2)$, with β_1 characterizing the baseline changes in the probability of conception according to timing in the fertile interval, and β_2 , which includes three parameters characterizing the changes across the four classes of follicular length. For the values of $\lambda_k = \exp(\beta_{1,k})$, for $k = 1, \dots, 12$, we chose diffuse priors by letting $a_{0k} = b_{0k} = 1$ for $k = 1, \dots, 12$ as hyper-parameters of the gamma distribution.

The exponentiated regression coefficients, $\gamma_h = \exp(\beta_{2,h})$ for $h = 1, 2, 3$, measure changes in the probability of conceiving a female associated with the follicular length changing between classes. For these parameters by incorporating a point mass at $\gamma_h = 1.0$ for $h = 1, 2, 3$, we assign positive prior probability to regions where different follicular length classes have the same effect on the probabilities of conceiving a female. Following the strategy of Westfall et al. (30), we let $\pi_h = 0.5^{1/3}$ in order to assign 0.5 prior probability to the null hypothesis of no association between the follicular length class and the day-specific probabilities to conceive a female. The choice $a_{0h} = b_{0h} = 0.1$ allows for a

high degree of uncertainty in the values of γ_h under the alternative hypothesis.

To obtain results for the real data from the joined European and Italian studies, we run the MCMC analysis with a burn-in of 5,000 iterations and a collection interval of 40,000 iterations. Convergence was assessed by examining the stationarity of the distribution of the parameters. Trace-plots and standard diagnostic criteria (Geweke and Gelman and Rubin statistics) of all the parameters provided no evidence against convergence. Effective sample size was obtained for each parameters chain and all were larger than four thousand units. The prior and posterior densities for day-specific parameters (indexed with respect to the ovulation day, $\lambda_{-8}, \dots, \lambda_3$) are plotted in **Figure 2**. Although we suppose vague priors, the posteriors seem, for all parameters, relatively concentrated. In some days (e.g., days $-7, 1, 2$) the probability to conceive a female seems higher and in other days (e.g., days $-6, -3, 3$) it seems lower. However, all the posterior distributions show a relatively large variability (although much less than the prior) and all of them seem to assign a reasonable probability to values around 0.5 (or 0.48, the probability of generating a girl given the sex ratio at birth). It seems that there is not enough evidence to consider as significant the observed differences in probabilities.

Posterior summaries of the parameters are presented in **Table 3**, and the prior and posterior densities for the effect parameters of follicular length classes γ_1, γ_2 , and γ_3 are plotted in **Figure 3**. Each increase of one class in follicular length results in a non-significant increase or decrease in the probabilities of conceiving a female. In fact, the probability that each effect parameters is equal to one is around 50%, indicating that no increase in the effect is significant. Therefore, the follicular length seems not to have any effect on the probability to conceive a female.

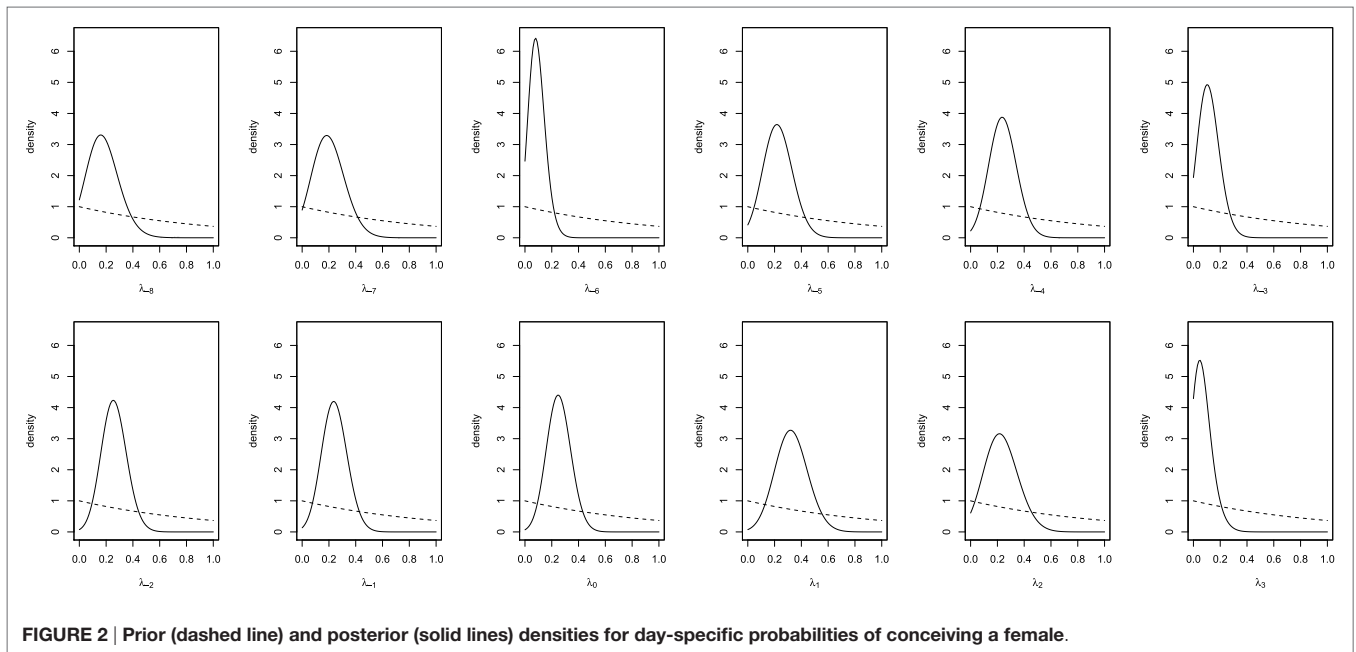


FIGURE 2 | Prior (dashed line) and posterior (solid lines) densities for day-specific probabilities of conceiving a female.

New Zealand data included in the “European fecundability” data set are obtained from a study with a different protocol with respect to the other available data. However, as discussed in Colombo and Masarotto (19), data and results are quite comparable between New Zealand and European studies. Thus, in order to increase the sample size of our analysis, in particular to estimate precision of our estimates, we prefer to include these data in our analysis. However, a sensitivity analysis was performed to confirm the validity of our results by repeating the estimates without these data. The results show very little differences in the point estimates ($\lambda_{-8} = 0.61, \lambda_{-7} = 0.57, \lambda_{-6} = 0.25, \lambda_{-5} = 0.45, \lambda_{-4} = 0.53, \lambda_{-3} = 0.25, \lambda_{-2} = 0.67, \lambda_{-1} = 0.47, \lambda_0 = 0.50, \lambda_1 = 0.81, \lambda_2 = 0.53, \lambda_3 = 0.19, \gamma_1 = 1.23, \gamma_2 = 0.86, \gamma_3 = 1.20$) which are all included in the credibility intervals shown in Table 3.

The estimated day-specific sex of the babies probabilities with 95% credible intervals is plotted in Figure 4. Credibility intervals are quite large and although a large variability between days is

visible, it seems that there are no clear pieces of evidence on days with higher (or lower) probability to conceive a female.

4. DISCUSSION

This article has applied a Bayesian approach for inferences on predictors of day-specific probabilities of conceiving a female in the menstrual cycle. A variable selection-type prior is used for the regression coefficients, which allows uncertainty in the predictors to be included in the model and can be used to construct hypothesis tests comparing null hypotheses of no association with unrestricted alternatives.

TABLE 3 | Posterior summaries of day-specific parameters (λ 's) and of follicular length increasing parameters (γ 's).

Parameter	Mean	Median	SD	95% credible interval	$\hat{P}(\gamma_h = 1)$
λ_{-8}	0.58	0.57	0.19	(0.22, 0.94)	–
λ_{-7}	0.68	0.70	0.19	(0.27, 0.98)	–
λ_{-6}	0.22	0.21	0.11	(0.06, 0.46)	–
λ_{-5}	0.54	0.54	0.13	(0.28, 0.78)	–
λ_{-4}	0.53	0.53	0.11	(0.31, 0.74)	–
λ_{-3}	0.23	0.22	0.10	(0.05, 0.45)	–
λ_{-2}	0.47	0.47	0.09	(0.30, 0.64)	–
λ_{-1}	0.42	0.42	0.09	(0.25, 0.60)	–
λ_0	0.52	0.52	0.09	(0.34, 0.70)	–
λ_1	0.68	0.69	0.10	(0.47, 0.87)	–
λ_2	0.62	0.63	0.15	(0.29, 0.88)	–
λ_3	0.17	0.14	0.12	(0.01, 0.46)	–
γ_1	1.54	1.06	0.84	(0.79, 3.84)	0.43
γ_2	0.82	1.00	0.38	(0.03, 1.07)	0.70
γ_3	1.28	1.00	0.71	(0.33, 3.33)	0.60

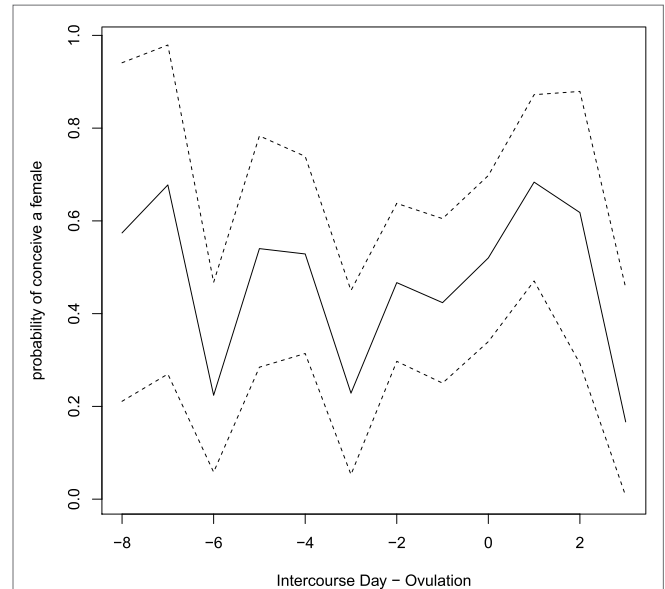


FIGURE 4 | Estimated day-specific probabilities of conceiving a female. Dashed lines show credibility intervals.

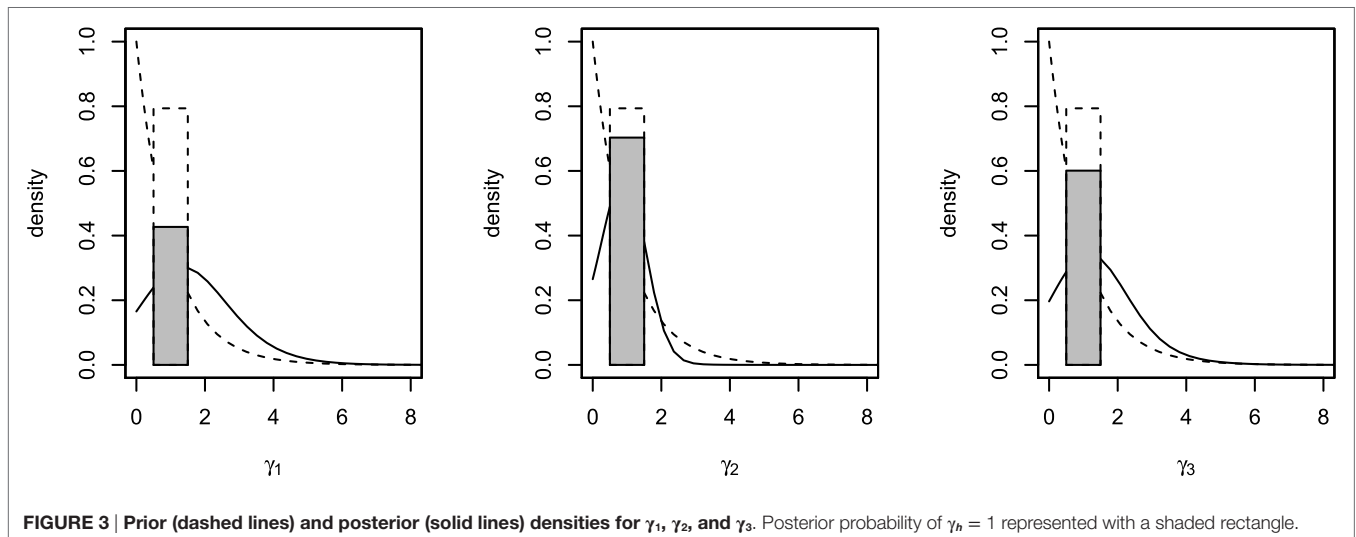


FIGURE 3 | Prior (dashed lines) and posterior (solid lines) densities for $\gamma_1, \gamma_2,$ and γ_3 . Posterior probability of $\gamma_h = 1$ represented with a shaded rectangle.

Our analysis clearly show that we could not show an effect of the follicular length on the sex of the baby, once we consider the day-specific effect, by validating, using independent data, the results by Gray et al. (15).

The significance of the day-specific effect seems to be more difficult to be examined. Although our point estimates seem to show a pattern in the probabilities of conceiving a female, since posterior mode seems to be higher just after the mucus peak day, the variability of the posterior distribution, which reflect the variability in the data, does not allow us to give any evidence for these conclusions. In fact, credibility intervals are very large for each parameter and almost always include the sex ratio at birth (0.485). Therefore, it seems that, up to nowadays, we do not have any actual evidence to suggest different behaviors in the pattern of intercourse for people willing to conceive a female or a male.

REFERENCES

- Smits LJM, De Bie RA, Essed GG, van den Brandt PA. Time to pregnancy and sex offspring: cohort study. *Br Med J* (2005) **331**:1437–8. doi:10.1136/bmj.331.7530.1437
- Weinberg CR, Baird DD, Wilcox AJ. The sex of the baby may be related to the length of the follicular phase in the conception cycle. *Hum Reprod* (1995) **10**(2):304–7.
- Kleegman S. Can sex be planned by the physician? In: Ingelman-Sundberg A, Westin B, editors. *Fifth World Congress of Fertility and Sterility*; 1966 Jun 16–22; Stockholm, Sweden. Amsterdam: Excerpta Medica (1966).
- Guerrero R. Sex ratio: a statistical association with the type and the time of the insemination in the menstrual cycle. *Int J Fertil* (1970) **15**:221–5.
- Guerrero R. Association of the type and time of insemination within the menstrual cycle with the human sex ratio at birth. *N Engl J Med* (1974) **291**:1056–9. doi:10.1056/NEJM197411142912004
- Guerrero R. Type and time of insemination within the menstrual cycle and human sex ratio at birth. *Stud Fam Plann* (1975) **6**:367–71. doi:10.2307/1964917
- Billings EL, Westmore A. *The Billings Method: Controlling Fertility without Drugs or Devices*. Australia: Anne O'Donovan Press (1980).
- Stolkowki J, Lorrein J. Preconceptional selection of fetal sex. *Int J Gynecol Obstet* (1980) **18**:440–3.
- Mathews F, Johnson P, Nei A. You are what your mother eats: evidence for maternal preconception diet influencing foetal sex in humans. *Proc Biol Sci* (2008) **275**:1661–8. doi:10.1098/rspb.2008.0105
- Shettles LB. Factors influencing sex ratios. *Int J Gynecol Obstet* (1970) **8**:643–7.
- Zarutskie PW, Muller CH, Magone M, Soules MR. The clinical relevance of selection techniques. *Fertil Steril* (1989) **52**(6):891–905.
- James WH, Rostron J. Parental age, parity and sex ratio in births in England and Wales, 1968–1977. *J Biosoc Sci* (1985) **17**:47–56. doi:10.1017/S0021932000015467
- Helle S, Helama S, Jokela J. Temperature related birth sex ratio bias in historical Sami: warm years bring more sons. *Biol Lett* (2007) **4**:60–2. doi:10.1098/rsbl.2007.0482
- Astolfi P, Zonta LA. Reduced male births in major Italian cities. *Hum Reprod* (1999) **14**(12):3116–9. doi:10.1093/humrep/14.12.3116
- Gray RH, Simpson JH, Bitto AC, Queenan JT, Li C, Kambic RT, et al. Sex ratio associated with timing of insemination and length of follicular phase in planned and unplanned pregnancies during use of natural family planning. *Hum Reprod* (1998) **13**:1397–400. doi:10.1093/humrep/13.5.1397
- McSweeney L. Successful sex pre-selection using natural family planning. *Afr J Reprod Health* (2011) **15**(1):79–84.
- Perez A, Eger R, Domenichini V, Kambic R, Gray RH. Sex ratio associated with natural family planning. *Fertil Steril* (1985) **43**:152–3.
- Barrett JC, Marshall J. The risk of conception on different days of the menstrual cycle. *Popul Stud* (1969) **23**(2):455–61. doi:10.1080/00324728.1969.10405297

ETHICS STATEMENT

No vulnerable population was involved. Ethics committee: Institutional Review Board of the Fondazione Lanza (Padua, Italy).

AUTHOR CONTRIBUTIONS

BS is the sole author of this work.

ACKNOWLEDGMENTS

The author dedicates this paper in memory of Professor Bernardo Colombo. The author, also, wishes to express his gratitude to Taddeo Mauro for the help he kindly provided for computational exercises and with comments about the ongoing work.

- Colombo B, Masarotto G. Daily fecundability: first results from a new data base. *Demogr Res* (2000) **3**:5. doi:10.4054/DemRes.2000.3.5
- Colombo B, Mion A, Passarin K, Scarpa B. Cervical mucus symptom and daily fecundability: first result of a new database. *Stat Methods Med Res* (2006) **15**:161–80. doi:10.1191/0962280206sm437oa
- Dunson DB, Stanford JB. Bayesian inference on predictors of conception probabilities. *Biometrics* (2005) **61**:126–33. doi:10.1111/j.0006-341X.2005.031231.x
- Zhou H, Weinberg CR. Modeling conception as an aggregated Bernoulli outcome with latent variables via the EM algorithm. *Biometrics* (1996) **52**:945–54. doi:10.2307/2533055
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing (2015). Available from: <http://www.R-project.org/>
- France JT, Graham FM, Gosling L, Hair PI. A prospective study of preselection of sex of offspring by timing intercourse relative to ovulation. *Fertil Steril* (1984) **41**:894–900.
- France JT, Graham FM, Gosling L, Hair P, Knox BS. Characteristics of natural conceptual cycles occurring in a prospective study of sex preselection: fertility awareness symptoms, normal levels, sperm survival, and pregnancy outcome. *Int J Fertil* (1992) **37**:244–55.
- Billings EL, Brown JB, Billings JJ, Burger HG. Symptoms and hormonal changes accompanying ovulation. *Lancet* (1972) **1**(7745):282–4. doi:10.1016/S0140-6736(72)90291-7
- Ecochard R, Boehringer H, Rabilloud M, Marret H. Chronological aspects of ultrasonic, hormonal, and other indirect indices of ovulation. *Br J Obstet Gynaecol* (2001) **108**:822–9. doi:10.1016/S0306-5456(00)00194-7
- Direito A, Bailly S, Mariani A, Ecochard R. Relationships between the luteinizing hormone surge and other characteristics of the menstrual cycle in normally ovulating women. *Fertil Steril* (2013) **99**:279–85. doi:10.1016/j.fertnstert.2012.08.047
- Wilcox AJ, Weinberg CR, Baird DD. Timing of sexual intercourse in relation to ovulation. Effects on the probability of conception, survival of the pregnancy, and sex of the baby. *N Engl J Med* (1995) **333**:1517–1121. doi:10.1056/NEJM199512073332301
- Westfall PH, Johnson WO, Utts JM. A Bayesian perspective on the Bonferroni adjustment. *Biometrika* (1997) **84**:419–27. doi:10.1093/biomet/84.2.419

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Scarpa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Gibbs Sampling Algorithm for Posterior Computation

For posterior computation, following Dunson and Stanford (21), we first introduce for each units auxiliary variables

$$Z_{ik} \sim \text{Poisson}(\exp(u_{ik}^T \beta))$$

such that

$$Y_i = 1 \left(\sum_{k=1}^K X_{ik} Z_{ik} > 0 \right)$$

where $1(x)$ is the indicator function. If we integrate out the auxiliary Poisson variables, which have no effect on parameter interpretation, we have that $Y_i = 0$ if and only if $Z_{ik} = 0$ for all k such that $X_{ik} = 1$, that is there was an intercourse every day in the i th cycle.

Full conditional posterior distributions can then be obtained for each of the parameters and latent variables.

Step 1. Sample from the full conditional distribution of $Z_i = \sum_k X_{ik} Z_{ik}$ by setting $Z_i = 0$ if $Y_i = 0$ and otherwise sampling sequentially from

$$\pi(Z_i | Y_i = 1, \beta, \text{data}) = \text{Poisson} \left(\sum_{k=1}^K X_{ik} p_{ik} \right) \text{ truncated so}$$

that $Z_i > 0$,

$$\pi(Z_{i1}, \dots, Z_{iK} | Z_i, Y_i = 1, \beta, \text{data}) = \text{Multinomial}$$

$$\left(Z_i; \frac{X_{i1} p_{i1}}{\sum_k X_{ik} p_{ik}}, \dots, \frac{X_{iK} p_{iK}}{\sum_k X_{ik} p_{ik}} \right)$$

where $p_i = [p_1, \dots, p_K]$ and $p_{ik} = \exp(\mathbf{u}_{ik}^T \beta) = \lambda_k \prod_{h=1}^{z_i-1} \gamma_h$.

Step 2. Sample $\lambda_1, \dots, \lambda_K$ from their conjugate full conditional distributions:

$$\pi(\lambda_k | \mathbf{Z}_{[X=1]}, \beta_{(-\lambda_k)}, \text{data}) = \text{Gamma} \left(\lambda_k; a_{0k} + \sum_{i: X_{ik}=1} Z_{ik}, b_{0k} + \sum_{i: X_{ik}=1} \prod_{h=1}^{z_i-1} \gamma_h \right),$$

where $\mathbf{Z}_{[X=1]} = Z_{ik}; X_{ik} = 1$ and we integrate out $Z_{ik}; X_{ik} = 0$.

Step 3. Sample $\gamma_1, \dots, \gamma_H$ from their conjugate full conditional distributions:

$$\pi(\gamma_k | \mathbf{Z}_{[X=1]}, \beta_{(-\gamma_k)}, \text{data}) = I_1 - \text{Gamma}(\gamma_h; \tilde{\pi}_h, \tilde{a}_h, \tilde{b}_h),$$

where $\tilde{a}_h = a_h + \sum_{i,k: X_{ik}=1} 1_{(h < w_{ik})} Z_{ik}$, $\tilde{b}_h = b_h + \sum_{i,k: X_{ik}=1} 1_{(h < z_i)} \lambda_k \prod_{j:j \neq h} \gamma_j$, and

$$\tilde{\pi}_h = \frac{\pi_{0h} \exp(-\sum_{i,k: X_{ik}=1} 1_{(h < z_i)} \lambda_k \prod_{j:j \neq h} \gamma_j)}{\pi_{0h} \exp(-\sum_{i,k: X_{ik}=1} 1_{(h < z_i)} \lambda_k \prod_{j:j \neq h} \gamma_j) + (1 - \pi_{0h}) \frac{C(a_h, b_h)}{C(\tilde{a}_h, \tilde{b}_h)}}.$$

where $C(a, b) = b^a / \Gamma(a)$.

Repeat steps 1–3 until apparent convergence and calculate posterior summaries based on a large number of additional iterations.