Bayesian Analysis (2012) **7**, Number 4, pp. 771–800

Flexible Bayesian Human Fecundity Models

Sungduk Kim *, Rajeshwari Sundaram †, Germaine M. Buck Louis ‡ and Cecilia Pyper §

Abstract. Human fecundity is an issue of considerable interest for both epidemiological and clinical audiences, and is dependent upon a couple's biologic capacity for reproduction coupled with behaviors that place a couple at risk for pregnancy. Bayesian hierarchical models have been proposed to better model the conception probabilities by accounting for the acts of intercourse around the day of ovulation, i.e., during the fertile window. These models can be viewed in the framework of a generalized nonlinear model with an exponential link. However, a fixed choice of link function may not always provide the best fit, leading to potentially biased estimates for probability of conception. Motivated by this, we propose a general class of models for fecundity by relaxing the choice of the link function under the generalized nonlinear model framework. We use a sample from the Oxford Conception Study (OCS) to illustrate the utility and fit of this general class of models for estimating human conception. Our findings reinforce the need for attention to be paid to the choice of link function in modeling conception, as it may bias the estimation of conception probabilities. Various properties of the proposed models are examined and a Markov chain Monte Carlo sampling algorithm was developed for implementing the Bayesian computations. The deviance information criterion measure and logarithm of pseudo marginal likelihood are used for guiding the choice of links. The supplemental material section contains technical details of the proof of the theorem stated in the paper, and contains further simulation results and analysis.

Keywords: Conception, Fecundity, Generalized t-distribution, Generalized nonlinear model, Markov chain Monte Carlo, Menstrual Cycle, Posterior distribution

1 Introduction

Human fecundity, defined as the biologic capacity of men and women for reproduction irrespective of pregnancy intentions, is of considerable interest to epidemiologists, clinicians and couples interested in becoming pregnant. Motivation for such studies ranges from ascertaining the causes of infertility to assessing the effects of stress and chemical toxicants such as polychlorinated biphenyls and organochlorine pesticides on human fe-

*Division of Epidemiology, Statistics and Prevention Research *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Rockville, MD, kims2@mail.nih.gov

†Division of Epidemiology, Statistics and Prevention Research *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Rockville, MD, sundaramr2@mail.nih.gov

‡Division of Epidemiology, Statistics and Prevention Research *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Rockville, MD, louisg@mail.nih.gov

§National Perinatal Epidemiology Unit, University of Oxford, Oxford, U.K. cecilia.pyper@npeu.ox.ac.uk

 DOI:10.1214/12-BA726

cundity. Examples of such studies include the Oxford Conception Study (OCS) and the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development's recently completed LIFE Study (Buck Louis et al. 2011). Thus, there remains interest in developing statistical models for analyzing fecundity data (Dunson and Stanford 2005; Dunson and Weinberg 2000; Dunson et al. 2001; Dunson and Zhou 2000; Ecochard 2006; Scheike and Jensen 1997; Sundaram et al. 2011; Weinberg and Gladen 1986; Zhou et al. 1996), particularly in concert with the biology of human reproduction.

Quantitatively, human fecundity is measured by estimating the probability of conception per menstrual cycle for a non-contracepting sexually active couple (Gini 1926). This definition can be expanded to denote the day-specific probability of conception within the estimated fertile window. Barrett and Marshall (1969) first proposed a model for fecundity that was later extended by Schwartz et al. (1980). Consequently, most approaches for estimating day-specific probabilities of conception have focused on the model proposed by Schwartz et al. (1980):

$$P(Y_{ij} = 1|X_{ij}) = \omega \left\{ 1 - \prod_k (1 - \lambda_k^*)^{X_{ijk}} \right\}. \tag{1}$$

Here Y_{ij} indicates whether or not conception occurred in cycle j for woman i and $\boldsymbol{X}_{ij} = (X_{ij1}, \ldots, X_{ijK})'$ is the vector of intercourse indicators over the fertile window of cycle j for couple i. Furthermore, K denotes the length of a fertile window, ω denotes the probability of cycle viability, and λ_k^* denotes the probability of conception in a viable cycle with intercourse only on day k in a fertile window. Various generalizations of the above model have been proposed in the literature. Specifically, Weinberg et al. (1994) incorporated cycle-specific covariates into models while Zhou et al. (1996) and Dunson and Zhou (2000) incorporated within-woman dependency into models; Royston and Ferreira (1999) have proposed an alternative to the Schwartz et al. (1980) model, where only the intercourse acts during the most fertile day contributes to conception. Dunson (2001) proposed a Bayesian hierarchical modeling approach to accommodate day-specific covariates and heterogeneity among women. Subsequently, Dunson and Stanford (2005) extended their model by allowing the day-specific probability λ_k^* to be woman-specific, as well as cycle-specific. Sundaram et al. (2011) have proposed a survival analysis approach to assess day-specific conception probabilities while relaxing the independence assumption of fertilizing ability of acts of intercourse on different days for conception that is inherent in the Schwartz et al. (1980) class of models for day-specific probabilities of conception.

In this paper, we propose a general class of conception models for a binary pregnancy outcome by casting it in the framework of generalized nonlinear models (GNLM). Our proposed class of models includes several conception models available in the literature as special cases. For instance, the Schwartz et al. (1980) class of models corresponds to a choice of an exponential link function in the generalized linear model (GLM) setup. As is usual in practice, a fixed choice of link function may not always provide the best fit for a given data set. Furthermore, it is well known that an incorrect link function may lead to a substantial bias in the mean response estimates (Czado and Santner 1992). In the context of the conception model, this may yield biased estimated probabilities

of conception per menstrual cycle. Furthermore, there is a rich literature on Bayesian approaches that allow an unknown link function in the model. For example Mallick and Gelfand (1994) proposed to use a mixture of betas for the cumulative density function corresponding to the transformation from the real line to the unit interval. Newton et al. (1996) used a Dirichlet process for the unknown link. Chen et al. (1999) proposed a flexible class of parametric links allowing skewness. However, the model proposed by Chen et al. (1999) has the limitation that the intercept term is confounded with the skewness parameter. To overcome the identifiability problem, Kim et al. (2008) proposed generalized skewed t-link models for binary response data. We propose to use a flexible generalized t-link for a binary pregnancy outcome in this paper (Kim et al. 2008).

The paper is organized as follows. Section 2 provides a detailed development of the flexible human fecundity model and examines various properties of the related conception models with the proposed model. In Section 3, we discuss the likelihood function, prior, and posterior distribution, and examine the properties of the resulting posterior distribution along with a discussion of its computation. We demonstrate the performance of the proposed methodologies using a large scale simulation study in Section 4. In Section 5, we apply our methods to the Oxford Conception Study. We conclude the paper with a brief discussion in Section 6.

2 The Models

2.1 Preliminary Findings

Suppose that i denotes a specific couple, j denotes a specific menstrual cycle, and k denotes a day within the fertile window of a menstrual cycle. We assume that there are I couples in the study, and each couple i contributes n_i cycles ($i = 1, \ldots, I$). Let Y_{ij} denote a 0/1 binary outcome variable, where a 1 denotes that the conception occurred during j-th menstrual cycle for i-th couple, and a 0 indicates otherwise. Let X_{ijk} denote the binary intercourse indicator variable, where $X_{ijk} = 1$ denotes that intercourse occurred on day k of menstrual cycle j for couple i, and 0 otherwise. Let $\boldsymbol{X}_{ij} = (X_{ij1}, \ldots, X_{ijK})'$, and $\boldsymbol{\lambda}_{ij} = (\lambda_{ij1}, \ldots, \lambda_{ijK})'$ for $j = 1, \ldots, n_i$ and $i = 1, \ldots, I$. As mentioned in the previous section, the most general extension of Schwartz et. al. model is that proposed by Dunson and Stanford (2005). They modeled the probability of conception in a menstrual cycle j by the i-th couple as follows:

$$P(Y_{ij} = 1|\boldsymbol{X}_{ij}) = 1 - \prod_k (1 - \lambda^*_{ijk})^{X_{ijk}}. \tag{2}$$

Using the following parametrization, $\lambda_{ijk} = -\log(1 - \lambda^*_{ijk})$, the above model can be expressed as follows:

$$\Pr(Y_{ij} = 1|\boldsymbol{X}_{ij}, \boldsymbol{\lambda}_{ij}) = \int_0^{\sum_{k:X_{ijk}=1} \lambda_{ijk}} \exp(-w^*)\, dw^*, \tag{3}$$

where $0 \leq \lambda_{ijk} < \infty$ since λ^*_{ijk} denotes the day-specific probability of conception in cycle j for couple i given that the intercourse occurs only on day k. We note here that in the special case of $\lambda^*_{ijk} = \lambda^*_k$ for all i, j, the conception model in (3) simplifies to the Barrett and Marshall (1969) model. Consequently, we have the following characterization of the conception model in (3). Observe that under conception model (3),

(i) The probability of conception in cycle j for couple i can be represented as the cumulative distribution function (cdf) of the exponential distribution with mean 1.

(ii) The conception model is the exponential link model in the generalized linear model (GLM) setup.

Furthermore, Dunson and Stanford (2005) incorporate a woman-specific frailty and the day-specific covariates through the day-specific conception probabilities, λ^*_{ijk} in (2) using a complementary log-log link. The most popular choices for modeling the day-specific probability of conception, λ^*_{ijk}, are the logit link, probit link, and complementary log-log link. However as noted above (3), the choice of link for the probability of conception in a cycle is an exponential link. Thus, the collection of available models can be viewed as the exponential link conception model with various links for λ^*_{ijk}. Our motivation for an extension is that this fixed choice of an exponential link may not always provide the best fit for a given dataset. In such situations, the conception model is mis-specified, which may yield substantial bias for mean response estimates (Czado and Santner 1992). As an extension of the exponential link conception model in (3), the following flexible conception model is proposed as

$$P(Y_{ij} = 1 | \boldsymbol{X}_{ij}, \boldsymbol{\lambda}_{ij}) = F^* \left(\sum_{k:X_{ijk}=1} \lambda_{ijk} \right), \tag{4}$$

where F^* is a cumulative distribution function with positive support of $R^+ = (0, \infty)$. Note that one can consider several distributions for F^* such as exponential, Weibull, extreme value, gamma, and log normal. In Section 2.2, we propose a class of Bayesian conception models using the latent variable approach, which relaxes the restriction of support of F^* from the positive real line to all of the real line. This formulation allows the development of an efficient Markov chain Monte Carlo (MCMC) algorithm using the conception model in (4).

2.2 The Flexible Human Fecundity Models

In this subsection, we extend the conception model in (3) to a general class of conception models for a binary pregnancy outcome. To this end, we reparametrize $w = \log w^*$ in (3). Then, we propose the flexible conception model for cycle j of couple i as follows:

$$\Pr(Y_{ij} = 1 | \boldsymbol{X}_{ij}, \boldsymbol{\lambda}_{ij}) = F \left(\log \sum_{k:X_{ijk}=1} \lambda_{ijk} \right), \tag{5}$$

where F is the cdf with a support of $R = (-\infty, \infty)$. Note that F is the link function in the GNLM setup. So, some choices for F are logistic, normal or Student t-distribution. To build a regression model, we introduce covariates through λ_{ijk} and model $\lambda_{ijk} = \lambda_k \exp\left(\boldsymbol{Z}'_{ijk}\boldsymbol{\beta}\right)$, where $\boldsymbol{Z}_{ijk} = (Z_{ijk1}, \ldots, Z_{ijkp})'$ denotes the covariate vector on day k for the jth cycle of couple i, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is the corresponding vector of regression coefficients, $k = 1, \ldots, K$, $j = 1, \ldots, n_i$, and $i = 1, \ldots, I$. This relationship between λ_{ijk} and $\boldsymbol{\beta}$ is equivalent to a canonical link for λ_{ijk} in the setting of generalized linear models. With this relationship, the day-specific conception probability in cycle j from couple i with intercourse only on day k can be represented as $\lambda^*_{ijk} = 1 - \exp(-\lambda_k \exp(\boldsymbol{Z}'_{ijk}\boldsymbol{\beta}))$, which is the same as that of Dunson and Stanford (2005) without the random effects.

Using the latent variable approach of Albert and Chib (1993), the binary conception regression model in (5) can be viewed as follows. Let w_{ij} be a latent variable such that

$$Y_{ij} = \begin{cases} 1 & \text{if } w_{ij} > 0 \\ 0 & \text{if } w_{ij} \le 0 \end{cases} \quad \text{and} \quad w_{ij} = \alpha_i + \log \sum_{k:X_{ijk}=1} \lambda_k \exp\left(\boldsymbol{Z}'_{ijk}\boldsymbol{\beta}\right) + \epsilon_{ij}, \tag{6}$$

where $\lambda_k > 0$, $\alpha_i \sim G$, $\epsilon_{ij} \sim F$, and α_i is a couple-specific frailty. Note that $1-\exp(-\lambda_k)$ can be considered a baseline day-specific conception probability. The model in (6) has several nice properties. Observe that the proposed model in (6) has a Bayesian generalized nonlinear model structure. This formulation allows the distribution F to have support of the whole real line. In this case, the covariate $\boldsymbol{Z}_{ijk}$ does not depend on day k for all i and j, the underlying latent variable has a mixed-effects model structure such as $w_{ij} = \alpha_i + \boldsymbol{Z}'_{ij}\boldsymbol{\beta} + \log \sum_{k:X_{ijk}=1} \lambda_k + \epsilon_{ij}$. This representation facilitates an easy implementation of the Gibbs sampling algorithm. Furthermore, (6) defines a rich class of conception models by varying the distribution F. For example, F could belong to a class of scale mixtures of normal distributions (see Chen and Dey (1998)). Also, one can easily express the marginal probability of conception by integrating over the couple-specific frailty α_i in (6) as follows:

$$\Pr(Y_{ij} = 1 | \boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}, \boldsymbol{\lambda}) = \int_{-\infty}^{\infty} F\left(\alpha_i + \log \sum_{k:X_{ijk}=1} \lambda_k \exp\left(\boldsymbol{Z}'_{ijk}\boldsymbol{\beta}\right)\right) g(\alpha_i) d\alpha_i, \tag{7}$$

where $g(\alpha_i)$ is the probability density function (pdf) of α_i. In this paper, we assume that F is a symmetric distribution (symmetric link). Observe that for the same choice of $\Lambda = \sum_{k:X_{ijk}=1} \lambda_{ijk}$, the probability of conception varies for different choices of F. In Figure 1, we display the probability of conception, $\Pr(Y_{ij} = 1 | \boldsymbol{X}_{ij}, \boldsymbol{\lambda}_{ij})$, as a function of $\sum_{k:X_{ijk}=1} \lambda_{ijk}$ for the exponential link and generalized t-link with various degrees of freedom (df). This figure suggests that if the true model is a generalized t-link model for a given data set, then the exponential link model will lead to either underestimation or overestimation of conception probabilities.

Motivated by the flexibility and other nice properties of the generalized t-distribution, we consider the generalized t-distribution as a choice for the distribution F in (7). The

conception model (7) corresponding to the generalized t-link for F will, henceforth, be referred to as the generalized t-link conception model. The generalized t-distribution has the following probability density function. Let $f_{gt,\nu_1,\nu_2}(w)$ denote the probability density function of the generalized t-distribution as introduced by Arellano-Valle and Bolfarine (1995). It is given by

$$f_{gt,\nu_1,\nu_2}(w) = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{\nu_1+1}{2}\right)}{\sqrt{\nu_2}\,\Gamma\left(\frac{\nu_1}{2}\right)} \times \frac{1}{\left(1+\frac{w^2}{\nu_2}\right)^{\frac{\nu_1+1}{2}}}. \tag{8}$$

Here ν_1 is a shape parameter (or degrees of freedom) and ν_2 is a scale parameter. Note that when $\nu_1 = \nu_2 = \nu$, (8) reduces to a Student t-distribution with ν degrees of freedom. The probability density function of a generalized t-distribution is symmetric about zero and bell-shaped, with a small value of ν_1 corresponding to a heavy tailed distribution. The introduction of the second parameter ν_2 offers some additional properties. For instance, by varying the choice of the parameters (ν_1 and ν_2), the density corresponding to the generalized t-distribution may or may not cross the standard normal density. In fact, it may even cross twice. In other words, the distribution has varied shapes. Note that fixing the scale parameter allows the shape parameter to purely control the tails of the link. Further details concerning the generalized t-distribution may be found in Kim et al. (2008). As noted in Figure 1, the generalized t-link conception model allows for various shapes of the link functions.

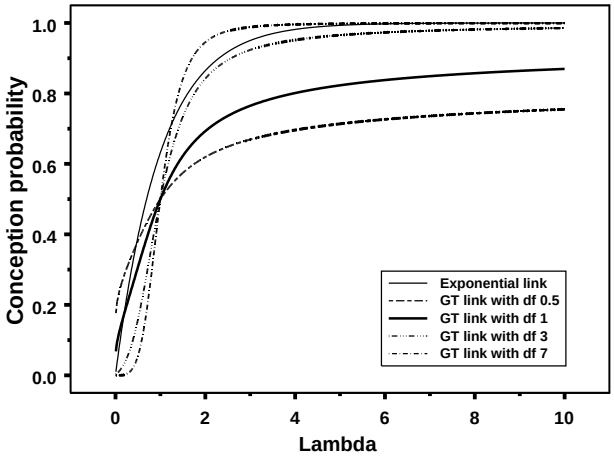


Figure 1: Conception probability plots based on exponential link and generalized t-link (GT link) with degrees of freedom $\nu_1 = 0.5, 1, 3$, and 7, where $\Lambda = \sum_{k:X_{ijk}=1} \lambda_{ijk}$.

3 Posterior Properties and Computation

In this section, we present the details of our estimation procedure including the specification of the priors and the posterior computations.

We first begin by presenting the observed likelihood. Without loss of generality, we assume $\nu_2 = 1$. Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_I)'$ and $D_{obs} = (\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z})$. The observed likelihood based on D_{obs} is given by

$$\begin{aligned} & L(\boldsymbol{\lambda}, \boldsymbol{\beta}, \nu_1, \sigma_\alpha^2 | D_{obs}) \\ & = \prod_{i=1}^{I} \int \prod_{j=1}^{n_i} \left[F_{gt,\nu_1,\nu_2=1} \left(\alpha_i + \log \sum_{k:X_{ijk}=1} \lambda_k \exp\left(\boldsymbol{Z}'_{ijk}\boldsymbol{\beta}\right) \right) \right]^{Y_{ij}} \\ & \qquad \times \left[1 - F_{gt,\nu_1,\nu_2=1} \left(\alpha_i + \log \sum_{k:X_{ijk}=1} \lambda_k \exp\left(\boldsymbol{Z}'_{ijk}\boldsymbol{\beta}\right) \right) \right]^{1-Y_{ij}} \\ & \qquad \times N(\alpha_i; 0, \sigma_\alpha^2) d\alpha_i, \end{aligned} \tag{9}$$

under the assumption of a zero mean normal distribution with variance σ_α^2 for α_i. In general, the likelihood function $L(\boldsymbol{\lambda}, \boldsymbol{\beta}, \nu_1, \sigma_\alpha^2 | D_{obs})$ is analytically intractable. However, it can be computed using a Monte Carlo algorithm. The computational development for $L(\boldsymbol{\lambda}, \boldsymbol{\beta}, \nu_1, \sigma_\alpha^2 | D_{obs})$ is discussed below. But, we first introduce the priors.

We assume that $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$, ν_1, and σ_α^2 are independent *a priori*. Thus, the joint prior for $(\boldsymbol{\lambda}, \boldsymbol{\beta}, \nu_1, \sigma_\alpha^2)$ is of the form $\pi(\boldsymbol{\lambda}, \boldsymbol{\beta}, \nu_1, \sigma_\alpha^2) = \pi(\boldsymbol{\lambda})\pi(\boldsymbol{\beta})\pi(\nu_1)\pi(\sigma_\alpha^2)$. We further assume

$$\begin{aligned} & \boldsymbol{\beta} \sim N_p(0, cI_p), \;\; \pi(\boldsymbol{\lambda}) \propto \prod_{k=1}^{K} \lambda_k^{a_0-1} e^{-b_0\lambda_k}, \\ & \pi(\nu_1) \propto \nu_1^{a_1-1} e^{-b_1\nu_1}, \;\; \text{and} \;\; \pi(\sigma_\alpha^2) \propto (\sigma_\alpha^2)^{-(a_2+1)} e^{-b_2/\sigma_\alpha^2}, \end{aligned} \tag{10}$$

where c, a_0, b_0, a_1, b_1, a_2, and b_2 are the pre-specified hyperparameters. In Sections 4 and 5, we use $c = 1000$ for $\pi(\boldsymbol{\beta})$, $a_0 = 1$ and $b_0 = 0.1$ for $\pi(\boldsymbol{\lambda})$, $a_1 = 1$ and $b_1 = 0.1$ for $\pi(\nu_1)$, and $a_2 = 2$ and $b_2 = 1$ for $\pi(\sigma_\alpha^2)$. These priors, though informative, are widely dispersed. Further, we can consider the following shrinkage prior distribution for a wider fertile window instead of the gamma prior distribution for $\boldsymbol{\lambda}$ in (1): Let $\lambda_k = \exp(\delta_k)$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_K)'$, and $V = \text{diag}\{\xi_1^2, \ldots, \xi_K^2\}$. We also assume the prior distribution for $\boldsymbol{\delta}$ as follows:

$$\boldsymbol{\delta} | \boldsymbol{\xi} \sim N_K(0, V) \;\; \text{and} \;\; \pi(\xi_k^2) \propto \left(\xi_k^2\right)^{a_0-1} \exp(-b_0\xi_k^2), \tag{11}$$

where a_0 and b_0 are the pre-specified hyperparameters. We note that this allows the individual elements $\xi_k^2, k = 1, \ldots, K$, to be independently updated towards 0 which eventually results in shrinkage of the δ_k to a point mass at zero.

Based on the prior distributions specified above, the joint posterior distribution of $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$, ν_1, and σ_α^2 based on the observed data D_{obs} is thus given by

$$\pi(\boldsymbol{\lambda}, \boldsymbol{\beta}, \nu_1, \sigma_\alpha^2 | D_{obs}) \propto L(\boldsymbol{\lambda}, \boldsymbol{\beta}, \nu_1, \sigma_\alpha^2 | D_{obs}) \pi(\boldsymbol{\lambda}) \pi(\boldsymbol{\beta}) \pi(\nu_1) \pi(\sigma_\alpha^2), \tag{12}$$

where $L(\boldsymbol{\lambda}, \boldsymbol{\beta}, \nu_1, \sigma_\alpha^2 | D_{obs})$ is given in (9). Next, we establish the propriety of the posterior distribution in (12) using the prior $\pi(\boldsymbol{\lambda}, \boldsymbol{\beta}, \nu_1, \sigma_\alpha^2) = \pi(\boldsymbol{\lambda})\pi(\nu_1)\pi(\sigma_\alpha^2)$.

Suppose that intercourse occurs at least once in the fertile window of menstrual cycle j for couple i. That is, $X_{ijk} = 1$ for a day k in $1, \cdots, K$ for $j = 1, \ldots, n_I$ and $i = 1, \ldots, I$. Let $c_{ij} = 1$ if $Y_{ij} = 0$ and $c_{ij} = -1$ if $Y_{ij} = 1$. Also let $\boldsymbol{Z}_{ij}^* = c_{ij} 1(X_{ijk} = 1)\boldsymbol{Z}_{ijk}$ for k, where $1(\cdot)$ denotes the indicator function, and $\boldsymbol{Z}_i^* = (\boldsymbol{Z}_{i1}^*, \ldots, \boldsymbol{Z}_{in_i}^*)'$. Define $Z_{l,m}^* = (\boldsymbol{Z}_i^*, l < i < m)$ as the $(m - l) \times p$ matrix with rows $\boldsymbol{Z}_i^*, l < i \leq m$, where $0 \leq l < m \leq I$.

Theorem 3.1. *Assume that* $\pi(\boldsymbol{\beta}) \propto 1$. *Suppose that*

(C1) there exist $h > p$, $0 = m_0 < m_1 < \cdots < m_h \leq I$, *and positive vectors* $d_1, d_2, \ldots, d_h$ *such that* Z_{m_{l-1}, m_l}^* *is of full rank and* $d_l' Z_{m_{l-1}, m_l}^* = 0$ *for* $l = 1, 2, \ldots, h$;

(C2) $\pi(\nu_1)$ *is bounded with* $\pi(\nu_1) = 0$ *for* $\nu_1 < a = p/h$.

(C3) $\sum_k X_{ijk} > 0$ *for* $j = 1, \ldots, n_I$ *and* $i = 1, \ldots, I$;

(C4) $a_0 > 0$ *and* $b_0 > 0$, $a_1 > 0$ *and* $b_1 > 0$, *and* $a_2 > 0$ *and* $b_2 > 0$ *in (1).*

Then the posterior distribution $\pi(\boldsymbol{\lambda}, \boldsymbol{\beta}, \nu_1, \sigma_\alpha^2 | D_{obs})$ *in (12) is proper.*

The proof of the above theorem is provided in the supplementary section. The sufficient conditions stated in the theorem for the propriety of the posterior distribution are quite general, and are typically satisfied by most data. Theorem 3.1 guarantees propriety of the posterior distribution in (12) of $\boldsymbol{\beta}$ using an improper uniform prior. Theorem 3.1 also implies that the regression coefficients $\boldsymbol{\beta}$ are identifiable under the generalized t-link conception model. The result established in Theorem 3.1 enables us to carry out Bayesian inference with an improper prior for the regression coefficients.

We next discuss the proposed efficient MCMC algorithm to carry out posterior computations. The form of the joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\lambda}, \nu_1, \sigma_\alpha^2)$ given in (12) does not lend itself to explicit analytic evaluation. Here, we propose an efficient Markov chain Monte Carlo sampling algorithm to sample from this joint posterior distribution. Since it is difficult to work directly with the generalized t-distribution, we express the generalized t-distribution as a gamma mixture of normal distributions. We introduce a mixing variable τ_{ij} such that $\epsilon_{ij} | \tau_{ij} \sim N\left(0, \frac{1}{\tau_{ij}}\right)$, $\tau_{ij} \sim \text{Gamma}\left(\frac{\nu_1}{2}, \frac{1}{2}\right)$. Let $\boldsymbol{w} = (w_{ij}; j = 1, \ldots, n_i, i = 1, \ldots, I)'$ and $\boldsymbol{\tau} = (\tau_{ij}; j = 1, \ldots, n_i, i = 1, \ldots, I)'$. Then the joint posterior distribution of $(\boldsymbol{w}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\beta}, \nu_1, \sigma_\alpha^2)$ based on the observed data D_{obs} is

given by

$$\begin{aligned}
&\pi(\boldsymbol{w},\boldsymbol{\tau},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,\sigma_\alpha^2|D_{obs})\\
&\propto\prod_{i=1}^{I}\prod_{j=1}^{n_i}\Big[1(Y_{ij}=0)1(w_{ij}\le 0)+1(Y_{ij}=1)1(w_{ij}>0)\Big]\\
&\qquad\times\tau_{ij}^{1/2}\exp\left[-\frac{\tau_{ij}}{2}\left(w_{ij}-\alpha_i-\log\sum_{k:X_{ijk}=1}\lambda_k\exp\left(\boldsymbol{Z}'_{ijk}\boldsymbol{\beta}\right)\right)^2\right]\\
&\qquad\times\left(\sigma_\alpha^2\right)^{-\frac{1}{2}}\exp\left(-\frac{\alpha_i^2}{2\sigma_\alpha^2}\right)\times\frac{\left(\frac{1}{2}\right)^{\frac{\nu_1}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)}\tau_{ij}^{\frac{\nu_1}{2}-1}e^{-\frac{\tau_{ij}}{2}}\\
&\times\prod_{k=1}^{K}\lambda_k^{a_0-1}e^{-b_0\lambda_k}\times\exp\left(-\frac{\boldsymbol{\beta}'\boldsymbol{\beta}}{2\sigma_\beta^2}\right)\times\nu_1^{a_1-1}e^{-b_1\nu_1}\\
&\times\left(\sigma_\alpha^2\right)^{-(a_2+1)}\exp\left(-\frac{b_2}{\sigma_\alpha^2}\right).
\end{aligned}\tag{13}$$

The sample from the joint posterior $\pi(\boldsymbol{w},\boldsymbol{\tau},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,\sigma_\alpha^2|D_{obs})$ given in (13) can be achieved by sampling from the following conditional distributions: (i) $[\boldsymbol{w}|\boldsymbol{\tau},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,\sigma_\alpha^2,D_{obs}]$; (ii) $[\boldsymbol{\alpha},\sigma_\alpha^2|\boldsymbol{w},\boldsymbol{\tau},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,D_{obs}]$; and (iii) $[\boldsymbol{\tau},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1|\boldsymbol{w},\boldsymbol{\alpha},\sigma_\alpha^2,D_{obs}]$. We briefly discuss how to sample from each of the above conditional posterior distributions.

For (i),

$$\begin{aligned}
w_{ij}|\boldsymbol{\tau},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,\sigma_\alpha^2,D_{obs}\sim N\left(\alpha_i+\log\sum_{k:X_{ijk}=1}\lambda_k\exp\left(\boldsymbol{Z}'_{ijk}\boldsymbol{\beta}\right),\frac{1}{\tau_{ij}}\right)\\
\times\Big[1(Y_{ij}=0)1(w_{ij}\le 0)+1(Y_{ij}=1)1(w_{ij}>0)\Big].
\end{aligned}$$

Sampling w_{ij} from the conditional distributions $w_{ij}|\boldsymbol{\tau},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,\sigma_\alpha^2,D_{obs}$ is straightforward.

For (ii), we apply the collapsed Gibbs technique of Liu (1994) via the following identity:

$$[\boldsymbol{\alpha},\sigma_\alpha^2|\boldsymbol{w},\boldsymbol{\tau},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,D_{obs}]=[\boldsymbol{\alpha}|\boldsymbol{w},\boldsymbol{\tau},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,\sigma_\alpha^2,D_{obs}][\sigma_\alpha^2|\boldsymbol{w},\boldsymbol{\tau},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,D_{obs}].\tag{14}$$

That is, we sample σ_α^2 after collapsing out $\boldsymbol{\alpha}$. Then

$$\alpha_i|\boldsymbol{w},\boldsymbol{\tau},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,\sigma_\alpha^2,D_{obs}\sim N\left(\mu_{\alpha_i},\sigma_{\alpha_i}^2\right),$$

where $\sigma_{\alpha_i}^2=\left(\sum_j\tau_{ij}+\frac{1}{\sigma_\alpha^2}\right)^{-1}$ and $\mu_{\alpha_i}=\sigma_{\alpha_i}^2\times\sum_j\tau_{ij}\left(w_{ij}-\log\sum_{k:X_{ijk}=1}\lambda_k e^{\boldsymbol{Z}'_{ijk}\boldsymbol{\beta}}\right)$ for $i=1,2,\ldots,I$. Therefore, for $i=1,\ldots,I$, sampling α_i from the conditional distributions $\alpha_i|\boldsymbol{w},\boldsymbol{\tau},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,\sigma_\alpha^2,D_{obs}$ is straightforward. In (14), the conditional posterior

density for $[\sigma_\alpha^2|\boldsymbol{w},\boldsymbol{\tau},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,D_{obs}]$ has the form

$$
\begin{aligned}
&[\sigma_\alpha^2|\boldsymbol{w},\boldsymbol{\tau},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,D_{obs}] \\
&\propto \prod_{i=1}^{I}\left[1+\sum_j \tau_{ij}\sigma_\alpha^2\right]^{-\frac{1}{2}} \exp\left[\frac{1}{2}\frac{\sigma_\alpha^2\left\{\sum_j \tau_{ij}\left(w_{ij}-\log\sum_{k:X_{ijk}=1}\lambda_k e^{\boldsymbol{Z}'_{ijk}\boldsymbol{\beta}}\right)\right\}^2}{\left(1+\sum_j \tau_{ij}\sigma_\alpha^2\right)}\right] \\
&\quad\times\left(\sigma_\alpha^2\right)^{-(a_2+1)}\exp\left(-\frac{b_2}{\sigma_\alpha^2}\right).
\end{aligned}
$$

We use the Metropolis-Hastings algorithm (Hastings 1970) to sample σ_α^2 from the conditional distribution $[\sigma_\alpha^2|\boldsymbol{w},\boldsymbol{\tau},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,D_{obs}]$.

For (iii), we also apply the collapsed Gibbs technique of Liu (1994) via the following identity:

$$
[\boldsymbol{\tau},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1|\boldsymbol{w},\boldsymbol{\alpha},\sigma_\alpha^2,D_{obs}] = [\boldsymbol{\tau}|\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,\sigma_\alpha^2,D_{obs}][\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1|\boldsymbol{w},\boldsymbol{\alpha},\sigma_\alpha^2,D_{obs}]. \quad (15)
$$

That is, we sample $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$, and ν_1 after collapsing out $\boldsymbol{\tau}$. Given $\boldsymbol{w}$, $\boldsymbol{\alpha}$, $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$, ν_1, σ_α^2, and D_{obs}, the τ_{ij}'s are conditionally independent and

$$
\tau_{ij}|\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\beta},\nu_1,\sigma_\alpha^2,D_{obs} \sim \text{Gamma}\left(a_{\tau_{ij}},b_{\tau_{ij}}\right),
$$

where $a_{\tau_{ij}} = \frac{\nu_1+1}{2}$ and $b_{\tau_{ij}} = \frac{1}{2}\left(1+\left[w_{ij}-\alpha_i-\log\sum_{k:X_{ijk}=1}\lambda_k\exp\left(\boldsymbol{Z}'_{ijk}\boldsymbol{\beta}\right)\right]^2\right)$ for $i=1,2,\ldots,I$ and $j=1,2,\ldots,n_i$. Therefore, we sample τ_{ij} from a gamma distribution. In (15), the conditional posterior density for $[\boldsymbol{\lambda}|\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta},\nu_1,\sigma_\alpha^2,D_{obs}]$ has the form

$$
\begin{aligned}
[\lambda_k|\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta},\nu_1,\sigma_\alpha^2,D_{obs}] &\propto \prod_{i=1}^{I}\prod_{j=1}^{n_i}\left[1+\left\{w_{ij}-\alpha_i-\log\sum_{k:X_{ijk}=1}\lambda_k e^{\boldsymbol{Z}'_{ijk}\boldsymbol{\beta}}\right\}^2\right]^{-\frac{\nu_1+1}{2}} \\
&\quad\times\prod_{k=1}^{K}\lambda_k^{a_0-1}e^{-b_0\lambda_k},
\end{aligned}
$$

for $k=1,2,\ldots,K$. Thus, we use the Metropolis-Hastings algorithm (Hastings 1970) to sample λ_k from conditional distribution $[\lambda_k|\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta},\nu_1,\sigma_\alpha^2,D_{obs}]$. From (15), given $\boldsymbol{\lambda}$, ν_1, $\boldsymbol{w}$, $\boldsymbol{\alpha}$, σ_α^2, and D_{obs}, we have the conditional form for $\boldsymbol{\beta}$ as follows:

$$
\begin{aligned}
[\boldsymbol{\beta}|\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\lambda},\nu_1,\sigma_\alpha^2,D_{obs}] &\propto \prod_{i=1}^{I}\prod_{j=1}^{n_i}\left[1+\left\{w_{ij}-\alpha_i-\log\sum_{k:X_{ijk}=1}\lambda_k e^{\boldsymbol{Z}'_{ijk}\boldsymbol{\beta}}\right\}^2\right]^{-\frac{\nu_1+1}{2}} \\
&\quad\times\exp\left(-\frac{\boldsymbol{\beta}'\boldsymbol{\beta}}{2\sigma_\beta^2}\right).
\end{aligned}
$$

We also use the Metropolis-Hastings algorithm to sample $\boldsymbol{\beta}$ from the conditional distribution $[\boldsymbol{\beta}|\boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \nu_1, \sigma_\alpha^2, D_{obs}]$. Further, from (15), the conditional posterior density for $[\nu_1|\boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\beta}, \sigma_\alpha^2, D_{obs}]$ has the form

$$[\nu_1|\boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\beta}, \sigma_\alpha^2, D_{obs}]$$
$$\propto \prod_{i=1}^{I}\prod_{j=1}^{n_i} \frac{\Gamma\left(\frac{\nu_1+1}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)} \left[1 + \left\{ w_{ij} - \alpha_i - \log \sum_{k:X_{ijk}=1} \lambda_k e^{\boldsymbol{Z}'_{ijk}\boldsymbol{\beta}} \right\}^2\right]^{-\frac{\nu_1}{2}} \times \nu_1^{a_1-1} e^{-b_1\nu_1}.$$

To sample ν_1 from $[\nu_1|\boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\beta}, \sigma_\alpha^2, D_{obs}]$, we use the Metropolis-Hastings algorithm. □

4 A Simulation Study

We present an extensive simulation study to illustrate our proposed human fecundity model based on the generalized t-distribution introduced in (8). The covariates vector $\boldsymbol{Z}_{ijk} = (Z_{ijk1}, Z_{ijk2})'$ was generated from $Z_{ijk1} \sim$ Bernoulli(0.3) and $Z_{ijk2} \sim N(0,1)$, $k = 1, \ldots, K$, $j = 1, \ldots, n_I$, and $i = 1, \ldots, I$, respectively. The maximum follow-up time was for $J = 6$ cycles and the length of the fertile window was assumed to be $K = 6$ days reflecting the typical length usually assumed in the literature. Further, we choose $\boldsymbol{\lambda} = (0.10, 0.20, 0.15, 0.35, 0.30, 0.15)'$, $\boldsymbol{\beta} = (-0.3, 0.2)'$, $\nu_1 = 1.5$, and $\sigma_\alpha^2 = 1.0$, respectively. Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_6)'$ and $\boldsymbol{\theta} = (\theta_1, \theta_2)'$. The intercourse variable X_{ijk} was generated using an independent Bernoulli(p_{ijk}) distribution with

$$p_{ijk} = \frac{\mu_k \exp(\theta_1 Z_{ijk1} + \theta_2 Z_{ijk2})}{1 + \mu_k \exp(\theta_1 Z_{ijk1} + \theta_2 Z_{ijk2})}$$

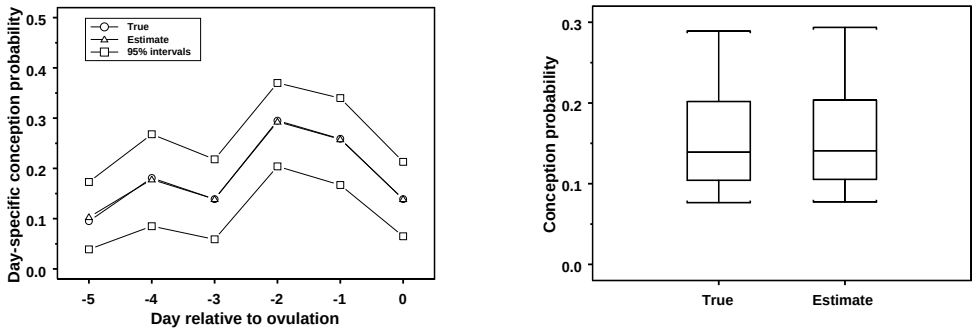
with $\mu = (0.30, 0.35, 0.32, 0.45, 0.42, 0.40)'$ and $\boldsymbol{\theta} = (0.2, -0.1)'$. We simulated $N = 1000$ samples of size $n = 300$ reflecting the typical size in the literature based on each combination of $(\boldsymbol{\lambda}, \boldsymbol{\beta}, \nu_1, \sigma_\alpha^2, \boldsymbol{\mu}, \boldsymbol{\theta})$. The prior distributions used were as specified in Section 3. The simulation results are given in Table 1 under the proposed generalized t-link conception model. Table 1 presents mean, standard deviation (SD), and empirical 95% interval for each of the parameters based on their estimated posterior means, along with the coverage probability (CP) of the 95% highest posterior density intervals (HPD). Observe that all the parameter estimates of $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$, ν_1, and σ_α^2 are close to their true values. The variation in the estimates of $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$, ν_1, and σ_α^2 across the 1000 replicates of simulated samples is small. We also note that the coverage probabilities are reasonably close to the nominal level.

Summary information concerning the performance of the estimation of baseline and overall day-specific conception probabilities is presented in Table I and II of the supplemental material section. Figure 1 (a) shows the baseline day-specific conception probability under the proposed generalized t-link conception model, indicating that the estimates are close to the true values. To further investigate the performance of the estimation procedure, we estimated the subject-specific probabilities of conception for

Table 1: Summary of simulation results

Parameter	True value	Mean	SD	95% interval	CP of 95% HPD
λ_1	0.10	0.110	0.041	(0.040, 0.190)	0.979
λ_2	0.20	0.198	0.059	(0.089, 0.311)	0.942
λ_3	0.15	0.151	0.050	(0.061, 0.246)	0.956
λ_4	0.35	0.348	0.062	(0.228, 0.462)	0.975
λ_5	0.30	0.300	0.061	(0.183, 0.416)	0.976
λ_6	0.15	0.151	0.046	(0.067, 0.239)	0.959
β_1	-0.3	-0.301	0.102	(-0.511, -0.113)	0.972
β_2	0.2	0.201	0.092	(0.029, 0.384)	0.981
ν_1	1.5	1.494	0.201	(1.144, 1.934)	0.968
σ^2_α	1.0	1.101	0.393	(0.508, 2.083)	0.959

each observed cycle for the individual. Figure 1 (b) reports the box-plots comparing the true subject-specific probabilities of conception with the estimated ones. Note that the estimated conception probabilities are close to the true values. Furthermore, throughout our estimation procedure, we observe little bias. The results presented in this section are based on 20,000 Gibbs samples to compute all the estimates, including the mean, standard deviation, and 95% intervals using a burn-in of 10,000 iterations. The convergence of the Gibbs sampler was checked using several diagnostic procedures as recommended by Cowles and Carlin (1996). The computer codes were written in FORTRAN 95 using IMSL subroutines with double precision accuracy.



(a) (b)

Figure 2: (a) Plot of baseline day-specific conception probabilities under the generalized t-link conception model; (b) Box-plots of true and estimated subject and cycle specific conception probabilities under the generalized t-link conception model.

5 Analysis of the Oxford Conception Study

We now present the analysis for a subset of women who participated in the Oxford Conception Study (Pyper et al. 2006), which is a prospective cohort comprising women aged 18-40 years who were attempting to become pregnant after being recruited via various media campaigns in the United Kingdom. Following a baseline interview, women were instructed in the use of home fertility monitors that track urinary reproductive hormones including luteinizing hormone, a proxy for impending ovulation, and also home pregnancy test kits for detection of human chorionic gonadotropin confirmed pregnancy. Women completed daily diaries on menstruation, sexual intercourse and lifestyle factors presumed relevant for human fertility (i.e., cigarette smoking, alcohol consumption). We were able to estimate ovulation for development of our conception probability models using two approaches: 1) using the first day of the menstrual cycle on which the fertility monitor detected luteinizing hormone (LH) indicative of impending ovulation and 2) using the observed and expected dates of menstruation from the daily diaries based upon the Ogino-Knaus (OK) Method. Specifically, the OK method estimates ovulation by counting back 14 days from the first day of menses (Knaus 1929; Ogino 1930), or using the expected first day of menses based upon the previous cycle's data if the woman conceives in that cycle. Both methods are proxies of ovulation, given the absence of a biomarker of ovulation for population research. The gold standard requires vaginal ultrasound verification. We defined the fertile window to be day -5 throught $+1$, with 0 denoting the day of ovulation for both approaches. The analysis was based upon 306 women with LH measurements and 324 women for the Ogino-Knaus method. Our biologic covariates included age (years) and parity defined as a previous live birth (yes/no), while our behavioral covariates included daily usage of cigarettes, alcohol consumption, and caffeine consumption as derived from daily diaries. The covariate age is continuous while smoking, alcohol, caffeine, and parity are binary covariates, where 0 and 1 denote no and yes, respectively. To help the numerical stability in the implementation of the MCMC sampling algorithm, all covariates were standardized. We present here the analysis based on the fertile window defined according to the LH surge. As a sensitivity analysis of our findings, we also reanalyzed the data based on the Ogino-Knauss algorithm for identifying ovulation; the results are presented as supplementary material.

As discussed in Section 2, we considered regression conception models with random effects under different choices of the link functions. We first considered the exponential link conception models with the complementary log-log link (Dunson and Stanford (2005) model, ECLLR) and logit link (ELogitR) for day-specific conception probability λ^*_{ijk} in (3), respectively. Next, we considered the Probit link (ProbitR), Cauchy link (CauchyR), and generalized t-link with ν_1 and $\nu_2 = 1$ (GTR) for F in (5). Note that the ECLLR and ELogitR models are the exponential link conception models with different links for the day-specific conception probabilities λ^*_{ijk}.

Finally, we also carried out a formal comparison of all the above models via the Deviance Information Criterion (DIC) as proposed by Speigelhalter et al. (2002). The

DIC is defined as follows:

$$\text{DIC} = D(\bar{\boldsymbol{\theta}}) + 2p_{_D}, \tag{16}$$

where $\boldsymbol{\theta}$ is the vector of all model parameters, $D(\boldsymbol{\theta})$ is a deviance function and $\bar{\boldsymbol{\theta}} = E[\boldsymbol{\theta}|D_{obs}]$ is the posterior mean of $\boldsymbol{\theta}$. In (16), $p_{_D}$ is the effective number of model parameters, which is calculated as $p_{_D} = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}})$, where $\overline{D(\boldsymbol{\theta})} = E[D(\boldsymbol{\theta})|D_{obs}]$ and $D(\boldsymbol{\theta}) = -2\log L(\boldsymbol{\theta}|D_{obs})$, where $L(\boldsymbol{\theta}|D_{obs})$ is given in (9). We also considered the logarithm of pseudo marginal likelihood (LPML) (Ibrahim et al. 2001) as well. LPML is a well established Bayesian model comparison criterion based on the conditional predictive ordinate (CPO) statistics. Let CPO_{ij} denote the CPO statistic in a menstrual cycle j by i-th couple. LPML is defined as

$$\text{LPML} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}\log(\text{CPO}_{ij}). \tag{17}$$

The larger the LPML, the better the fit of a given model.

Table 2: Model comparison based on LH method

Link	$D(\bar{\theta})$	P_D	DIC	LPML
ECLLR	836.43	13.10	862.64	-429.64
ELogitR	835.02	11.50	858.02	-424.21
Probit	845.19	12.16	869.51	-433.75
Cauchy	829.18	11.41	851.98	-419.17
GTR	825.84	10.77	847.37	-417.44

Table 2 reflects an interesting pattern for the values of DIC and LPML for various conception models with different links based on the LH method. Both the DIC and LPML indicate that the exponential conception models with logit link (ELogitR) fit data better than the exponential conception models with complementary log-log link (ECLLR) for the OCS. These two models are commonly used in the analysis of binary pregnancy outcomes. This is possible as they can both be viewed as exponential conception models with logit and complementary log-log link for the day-specific conception probabilities λ^*_{ijk}. Furthermore, the ELogitR model fits data better than the probit link (ProbitR) for the OCS. The cauchy link model (CauchyR) seems to fit better than ELogitR, ECLLR, and ProbitR. Also, the generalized t-link model (GTR) fits the data best among all the models considered. Overall, Table 2 reveals that the generalized t-link conception (GTR) model fits this data set better than the Dunson and Stanford (2005) (ECLLR) model. Also from Table 2, we see that the exponential link conception model with the logit link and complementary log-log link for λ^*_{ijk} does not fit the data well. The probit link conception link model fits the data worst. Therefore, it is important to consider various choices of links in analyzing the binary pregnancy outcome.

Table 3 reflects the posterior means, the posterior standard deviations (SD), and the 95% highest posterior density (HPD) intervals of the parameters of the regression model with the covariates smoking, alcohol, caffeine, parity, and age under the generalized t-link model based on the LH method for identifying ovulation. Table 3 indicates that the posterior estimates of regression coefficients for smoking, caffeine, and parity are positive, while the posterior estimates of regression coefficients for alcohol and age are negative. The 95% HPD intervals in the table indicate that parity is a significant predictor, while smoking, alcohol, caffeine, and age were not under the LH method. Further, the value of ν_1 is less than 1, which implies that the conception model based on the generalized t-link has a heavy tail. We also report the summary statistics for the overall day-specific conception probability in the supplementary section, Table VI. In Figure 3, we present the plot of posterior mean day-specific conception probability comparing the effect of parity. We can see that the day-specific conception probability is highest on day one (-1) based upon the LH method, but two days (-2) prior to estimated ovulation based on the OK method (supplementary section). This suggests that the LH surge is a better marker for identifying the day of ovulation.

Table 3: Posterior estimates under the generalized t-link model with random effect based on the LH method

Variable	Posterior Mean	Posterior SD	95% HPD Interval
λ_1	0.06526	0.04782	(0.00006, 0.15702)
λ_2	0.10277	0.07231	(0.00031, 0.23941)
λ_3	0.08989	0.05998	(0.00017, 0.20384)
λ_4	0.06068	0.04386	(0.00024, 0.14486)
λ_5	0.20189	0.09540	(0.03340, 0.38498)
λ_6	0.05534	0.04234	(0.00001, 0.13792)
λ_7	0.09741	0.06225	(0.00133, 0.21445)
Smoking (Y/N)	0.21562	0.21177	(-0.20945, 0.62885)
Alcohol (Y/N)	-0.08846	0.24334	(-0.57446, 0.38627)
Caffeine	0.33029	0.25554	(-0.14684, 0.85784)
Parity (Y/N)	0.65516	0.25771	(0.17685, 1.17149)
Age (Y/N)	-0.35461	0.23926	(-0.83310, 0.10492)
ν_1	0.95699	0.16500	(0.65626, 1.29075)
σ_α^2	5.48909	3.17916	(1.20135, 11.62307)

In our analysis, we find that the generalized t-link seems to provide the best fit based upon the DIC and LPML values. However, the Schwartz et al. (1980) model that corresponds to an exponential link based conception model does not seem to provide a good fit based upon the DIC and LPML. As mentioned previously, an improper choice of link function leads to bias in the estimation of the mean response (Czado and Santner 1992). Recall that in the case of the conception model, the mean response corresponds to the probability of conception in a cycle. In Figure 4, we present box-

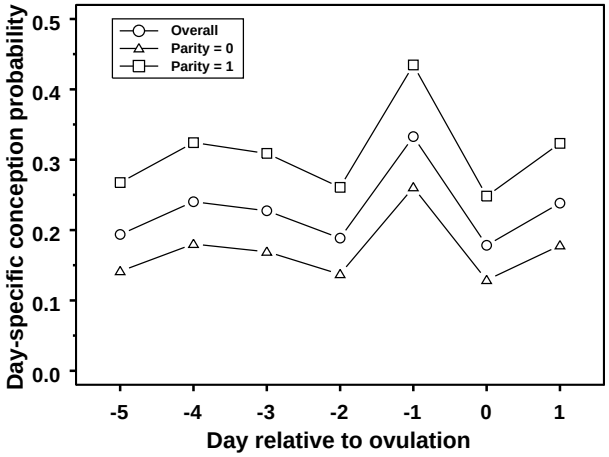


Figure 3: Plots of estimated day-specific conception probability comparing the effect of parity under the generalized t-link model with random effect (GTR model) for the LH method.

plots of estimated conception probability under various conception models with different links for the LH method. In Figure 4, we see that the the conception probability of the Dunson and Stanford (2005) model (ECLLR) tends to be underestimated compared to the exponential link conception models with logit (ECLL) link and the generalized t-link (GTR) model. We further see that the variations in conception probability estimates from the exponential link conception models with complementary log-log (ECLL) link and logit (ECLL) link for λ^*_{ijk} are greater than the generalized t-link (GTR) model. In Figure 5, we present box-plots of estimated conception probabilities by cycles under the generalized t-link conception model with a random effect (GTR model), and the Dunson and Stanford (2005) model (ECLLR model) based on the LH method. From Figure 5, we see that the variation of the estimated conception probability under generalized t-link conception model (GTR model) is smaller than that based on the Dunson and Stanford (2005) model by cycle for the LH method, and that the variation for the Dunson and Stanford (2005) model is much greater than that of generalized t-link conception model.

Note that $\nu_1 = .96$ for the generalized t-link conception model based upon the LH method for the OCS versus $\nu_1 = .8$ for the OK method. From Figure 1, we see that the rate approaching 1 under the generalized t-link conception model is slower than that of the exponential link conception model after a certain point for these values of the parameter ν_1. Consequently, the conception probability under the generalized t-link conception model is very different from that of the exponential link conception model. In conclusion, our analysis indicates that a flexible link function results in a better model fit. Thus, investigators have a prior for analysis in keeping with their research

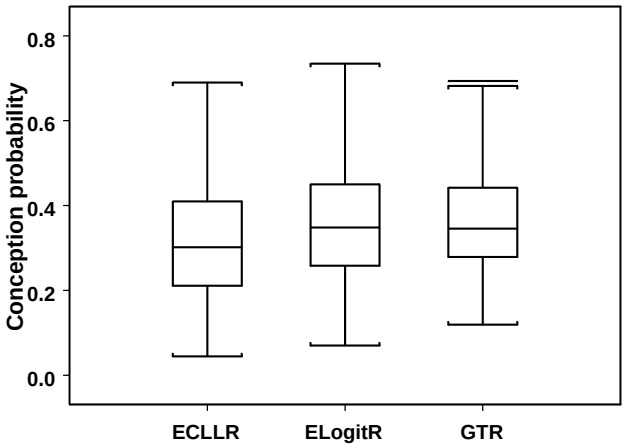


Figure 4: Box-plots of estimated conception probability for various conception models for the LH method.

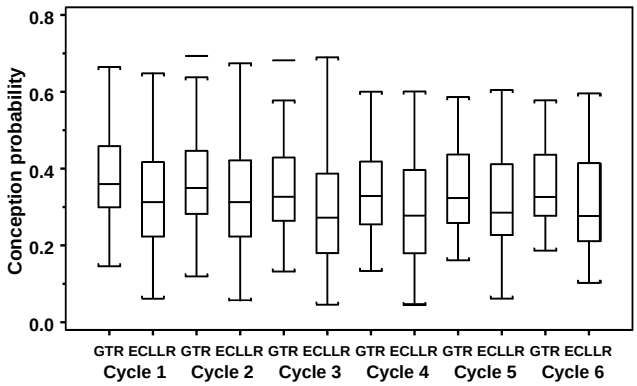


Figure 5: Box-plots of estimated subject-specific conception probability by cycles under the generalized t-link conception model with random effect (GTR) and the Dunson and Stanford (2005) model (ECLLR) for the LH method.

aims.

The computations presented here were based on 20,000 of the 100,000 Gibbs samples. These 20,000 iterations were selected based on a burn-in of 5,000 iterations with every fifth iterations being selected for computation. The computer codes were written in FORTRAN 95 using IMSL subroutines with double-precision accuracy. The convergence of the Gibbs sampler was checked using several diagnostic procedures as recommended by Cowles and Carlin (1996). Approximate convergence is reached after 2000 iterations.

6 Discussion

The early origins of health and disease (DOHaD hypothesis, Hanson and Gluckman (2008)) posits that exposures during sensitive windows of human development permanently reprogram the developing conceptus or fetus for extrauterine life. Such reprogramming may be associated with susceptibility for diminished fecundity or greater risk of chronic disease. To this end, it is imperative to delineate the timing and variability of the fertile window to facilitate accurate estimation of conception. This would be a first step for eventual quantification of exposures during sensitive windows of human development. In so doing, it is imperative to build biologically plausible models for estimating conception.

To identify the day of ovulation, we used urinary LH surge as a biomarker of ovulation. A World Health Orgaization (WHO) Task Force (WHO 1980a,b) has shown the surge in LH to be the best available marker of impending ovulation in absence of a gold standard or serial ultrasonography. Further, it has been shown in the literature that the length of the fertile window is 6 days ending on the day of ovulation (Weinberg et al. 1995; Dunson et al. 1999; Royston and Ferreira 1999; Dunson et al. 2001; Scarpa and Dunson 2007) using LH surge as a marker of ovulation. However, our methods can be applied to longer fertile windows using a shrinkage prior distribution.

We have proposed a class of flexible human fecundity models for analyzing a binary pregnancy outcome, and developed an efficient MCMC algorithm. The proposed conception models provide choices for modeling fecundity data, while ensuring the best fit. The proposed flexible models illustrated that the generalized t-distribution (5) provided a better fit than the exponential link in both analyses using the LH method for identifying the day of ovulation, as well as the Ogino-Knaus approach of identifying the day of ovulation (presented in the Supplementary Section). This framework of modeling conception in the GNLM set up lends itself to Bayesian variable selection for high dimension data where covariates do not depend on day k such as chemical toxicants (e.g., polychlorinated biphenyls). This extension is currently under investigation.

7 Supplementary Materials

7.1 Proof of Theorem 1 (propriety of posterior distribution):

Let $w \sim f_{gt,\nu_1,\nu_2=1}(w)$. For $0 < a < 1$, using a result on the beta function given in Section 6.2 of Abramowitz and Stegun (1972) and after some algebra, we obtain that

$$E\left(|w|^a\right) = \frac{\Gamma(\frac{a+1}{2})\Gamma(\frac{\nu_1-a}{2})}{\sqrt{\pi}\Gamma(\frac{\nu_1}{2})} \leq K(a) \tag{18}$$

for $\nu_1 > a$, where $K(a)$ is a finite positive constant depending only on a. Using (18) and conditions (C2) and (C4), we have

$$\begin{aligned}\int_0^\infty E\left(|w|^a\right)\ \pi(\nu_1|a_0,b_0)\ d\nu_1 &= \frac{\Gamma(\frac{a+1}{2})}{\sqrt{\pi}}\int_0^\infty \frac{\Gamma(\frac{\nu_1-a}{2})}{\Gamma(\frac{\nu_1}{2})}\ \pi(\nu_1|a_1,b_1)\ d\nu_1 \\ &\leq K(a)\int_0^\infty \pi(\nu_1|a_1,b_1)\ d\nu_1 < \infty. \end{aligned}\tag{19}$$

Let $\boldsymbol{w}^* = (c_{11}w_{11},\ldots,c_{I,n_I}w_{I,n_I})'$, $\boldsymbol{\alpha}^* = (c_{11}\alpha_1,\ldots,c_{I,n_I}\alpha_I)'$, and $\boldsymbol{\lambda}^* = (c_{11}1(X_{11k}=1)\log\lambda_k,\ldots,c_{I,n_I}1(X_{I,n_I,k}=1)\log\lambda_k)'$ for k. Using Fubini's theorem, we obtain

$$\begin{aligned}&\int_0^\infty\int_0^\infty\int_0^\infty\int_{R^p} L(\boldsymbol{\beta},\boldsymbol{\lambda},\nu_1,\sigma_\alpha^2|D_{obs})\ \pi(\boldsymbol{\lambda})\ \pi(\nu_1)\pi(\sigma_\alpha^2)\ d\boldsymbol{\beta}\ d\boldsymbol{\lambda}\ d\nu_1 d\sigma_\alpha^2 \\ &= \int_0^\infty\int_{R^I}\int_0^\infty\int_0^\infty\int_{R^p} E\Big[1\Big\{c_{ij}w_{ij} > c_{ij}\Big(\alpha_i + \log\sum_{k:X_{ijk}=1}\lambda_k\exp(\boldsymbol{Z}'_{ijk}\boldsymbol{\beta})\Big), \\ &\qquad 1\le j\le n_i, 1\le i\le I\Big\}\Big] \\ &\qquad \times \pi(\boldsymbol{\lambda})\ \pi(\nu_1)\pi(\sigma_\alpha^2)\ d\boldsymbol{\beta}\ d\boldsymbol{\lambda}\ d\nu_1\ dG(\boldsymbol{\alpha})d\sigma_\alpha^2 \\ &\le K_0\int_0^\infty\int_{R^I}\int_0^\infty\int_0^\infty\int_{R^p} E\Big[1\Big\{c_{ij}w_{ij} > c_{ij}\alpha_i + c_{ij}1(X_{ijk}=1)\log\lambda_k \\ &\qquad + c_{ij}1(X_{ijk}=1)\boldsymbol{Z}'_{ijk}\boldsymbol{\beta}, 1\le j\le n_i, 1\le i\le I\Big\}\Big] \\ &\qquad \times \pi(\boldsymbol{\lambda})\ \pi(\nu_1)\pi(\sigma_\alpha^2)\ d\boldsymbol{\beta}\ d\boldsymbol{\lambda}\ d\nu_1\ dG(\boldsymbol{\alpha})d\sigma_\alpha^2 \\ &= K_0\int_0^\infty\int_{R^I}\int_0^\infty\int_0^\infty E\Big[\int_{R^p} 1\Big\{c_{ij}1(X_{ijk}=1)\boldsymbol{Z}'_{ijk}\boldsymbol{\beta} < c_{ij}w_{ij} - c_{ij}\alpha_i \\ &\qquad - c_{ij}1(X_{ijk}=1)\log\lambda_k, 1\le j\le n_i, 1\le i\le I\Big\}\ d\boldsymbol{\beta}\Big] \\ &\qquad \times \pi(\boldsymbol{\lambda})\ \pi(\nu_1)\pi(\sigma_\alpha^2)\ d\boldsymbol{\lambda}\ d\nu_1\ dG(\boldsymbol{\alpha})d\sigma_\alpha^2 \\ &= K_0\int_0^\infty\int_{R^I}\int_0^\infty\int_0^\infty E\Big[\int_{R^p} 1\{Z^*\boldsymbol{\beta} < \boldsymbol{w}^* - \boldsymbol{\alpha}^* - \boldsymbol{\lambda}^*\}d\boldsymbol{\beta}\Big] \\ &\qquad \times \pi(\boldsymbol{\lambda})\ \pi(\nu_1)\pi(\sigma_\alpha^2)\ d\boldsymbol{\lambda}\ d\nu_1\ dG(\boldsymbol{\alpha})d\sigma_\alpha^2. \end{aligned}\tag{20}$$

Under condition (C1), it directly follows from Lemma 4.1 of Chen and Shao (2001) that there exists a constant K_1 such that $\|\boldsymbol{\beta}\| \le K_1 \min_{1\le l\le h}(\max_{m_{l-1}<i\le m_l}|u_i|)$ whenever

$Z^*\boldsymbol{\beta} \le u$, where $u = (u_1, u_2, \ldots, u_I)'$. Hence, from (20), we have

$$\begin{aligned}
&\int_0^\infty \int_0^\infty \int_0^\infty \int_{R^p} L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \nu_1, \sigma_\alpha^2 | D_{obs})\ \pi(\boldsymbol{\lambda})\ \pi(\nu_1)\pi(\sigma_\alpha^2)\ d\boldsymbol{\beta}\ d\boldsymbol{\lambda}\ d\nu_1 d\sigma_\alpha^2 \\
&\le K_0 \int_0^\infty \int_{R^I} \int_0^\infty \int_0^\infty E\Big[\int_{R^p} 1\{\|\boldsymbol{\beta}\| \le K_1 \min_{1\le l\le h} \max_{m_{l-1}<i\le m_l} |w_i^* - \alpha_i^* - \lambda_i^*|\} d\boldsymbol{\beta} \Big] \\
&\qquad \times \pi(\boldsymbol{\lambda})\ \pi(\nu_1)\pi(\sigma_\alpha^2)\ d\boldsymbol{\lambda}\ d\nu_1\ dG(\boldsymbol{\alpha}) d\sigma_\alpha^2 \\
&\le K_0 K_1 \int_0^\infty \int_{R^I} \int_0^\infty \int_0^\infty \prod_{l=1}^h E\Big(\max_{m_{l-1}<i\le m_l} |w_i^* - \alpha_i^* - \lambda_i^*|^{\frac{p}{h}} \Big) \\
&\qquad \times \pi(\boldsymbol{\lambda})\ \pi(\nu_1)\pi(\sigma_\alpha^2)\ d\boldsymbol{\lambda}\ d\nu_1\ dG(\boldsymbol{\alpha}) d\sigma_\alpha^2 \\
&\le 2^p K_0 K_1 \int_0^\infty \int_{R^I} \int_0^\infty \int_0^\infty \prod_{l=1}^h \sum_{m_{l-1}<i\le m_l} E\left(|w_i^*|^{\frac{p}{h}} + |\alpha_i^*|^{\frac{p}{h}} + |\lambda_i^*|^{\frac{p}{h}} \right) \\
&\qquad \times \pi(\boldsymbol{\lambda})\ \pi(\nu_1)\pi(\sigma_\alpha^2)\ d\boldsymbol{\lambda}\ d\nu_1\ dG(\boldsymbol{\alpha}) d\sigma_\alpha^2 \\
&< \infty,
\end{aligned}$$

by conditions (C1), (C2), (C3), and (C4), which completes the proof. □

7.2 Simulation results

Table I. Summary of baseline day-specific probabilities of conception

Day	True value	Mean	SD	95% interval
-5	0.095	0.103	0.036	(0.039, 0.173)
-4	0.181	0.178	0.048	(0.085, 0.268)
-3	0.139	0.139	0.043	(0.059, 0.218)
-2	0.295	0.293	0.044	(0.204, 0.370)
-1	0.259	0.258	0.045	(0.167, 0.340)
0	0.139	0.139	0.039	(0.065, 0.213)

Table II. Summary of overall day-specific probabilities of conception

Day	True value	Mean	SD	95% interval
-5	0.140	0.151	0.048	(0.060, 0.244)
-4	0.241	0.239	0.053	(0.132, 0.338)
-3	0.194	0.195	0.053	(0.090, 0.287)
-2	0.353	0.351	0.042	(0.261, 0.422)
-1	0.319	0.319	0.045	(0.224, 0.399)
0	0.194	0.195	0.047	(0.105, 0.282)

7.3 Analysis of Oxford data based on Ogino-Knaus method for identifying ovulation

We present the re-analysis of the Oxford data introduced in Section 5. Here the fertile window is identified as day -5 through day $+1$ where the day of ovulation 0 is identified based on the Ogino-Knaus method. We first present the DIC and LPML values for various conception models as discussed in Section 5. Notice that the findings are similar to those of the LH method. Also, the generalized t-link conception (GTR) model fits the data the best. Furthermore, the Schwartz et al. (1980) model that corresponds to an exponential link based conception model does not seem to provide a good fit based upon the DIC and LPML. We next present the parameter estimates and note findings similar to those in Section 5. Parity is significantly associated with probability of conception.

Table III. DIC and LPML values based on the OK method

Link	$D(\bar{\theta})$	P_D	DIC	LPML
ECLLR	938.58	13.15	964.88	-483.37
ELogitR	937.35	11.82	960.98	-477.48
Probit	955.56	12.34	980.24	-492.16
Cauchy	919.64	10.62	940.89	-467.48
GTR	916.76	10.30	937.37	-463.65

Table IV. Posterior estimates under the generalized t-link model with random effect based on the OK method

Variable	Posterior Mean	Posterior SD	95% HPD Interval
λ_1	0.08581	0.06257	(0.00057, 0.20490)
λ_2	0.06236	0.05707	(0.00001, 0.17358)
λ_3	0.09436	0.05938	(0.00481, 0.20903)
λ_4	0.12531	0.07726	(0.00561, 0.27294)
λ_5	0.08816	0.05888	(0.00188, 0.20065)
λ_6	0.06536	0.04626	(0.00021, 0.15424)
λ_7	0.04798	0.03840	(0.00002, 0.12284)
Smoking (Y/N)	0.31020	0.24377	(-0.17002, 0.79821)
Alcohol (Y/N)	-0.09172	0.25773	(-0.60690, 0.40591)
Caffeine (Y/N)	0.03267	0.26221	(-0.48413, 0.56172)
Parity	0.69538	0.29250	(0.14843, 1.28020)
Age	-0.34032	0.27528	(-0.89737, 0.18799)
ν_1	0.78917	0.12443	(0.55301, 1.03225)
σ^2_α	7.56416	4.36351	(1.53295, 16.22523)

In Tables V and VI, we present the day-specific probabilities for both data sets (OK

method and LH method). Also, Figure I represents the effect of parity on day-specific probabilities of conception.

Table V. Day-specific conception probabilities under the generalized t-link with random effect based on the OK method

	Day	Posterior Mean	Posterior SD	95% HPD Interval
Overall	-5	0.24286	0.08160	(0.08020, 0.39645)
	-4	0.20307	0.09285	(0.01986, 0.36780)
	-3	0.25687	0.07265	(0.11599, 0.39977)
	-2	0.28727	0.07537	(0.13874, 0.43273)
	-1	0.24917	0.07592	(0.09641, 0.39148)
	0	0.21784	0.07597	(0.06935, 0.36424)
	+1	0.18748	0.07379	(0.04044, 0.32325)
Parity=0	-5	0.18567	0.07393	(0.04634, 0.33060)
	-4	0.15339	0.08112	(0.00445, 0.29930)
	-3	0.19752	0.06878	(0.06108, 0.32853)
	-2	0.22433	0.07307	(0.08011, 0.36424)
	-1	0.19107	0.07061	(0.05620, 0.33055)
	0	0.16429	0.06798	(0.03554, 0.29606)
	+1	0.13918	0.06497	(0.01889, 0.26182)
Parity=1	-5	0.32526	0.10223	(0.11986, 0.51702)
	-4	0.27494	0.11653	(0.04315, 0.48514)
	-3	0.34233	0.08961	(0.16556, 0.51525)
	-2	0.37793	0.09095	(0.20190, 0.55499)
	-1	0.33296	0.09404	(0.15007, 0.51624)
	0	0.29513	0.09653	(0.09901, 0.47383)
	+1	0.25727	0.09442	(0.07425, 0.43841)

Figures II and III indicate similar findings to those presented in Section 5. We further see that the box-plot for the generalized t-link (GTR) has smaller variation that those of the Dunson and Stanford (2005) model (ECLLR) and similar results for box-plots of estimated conception probability by cycles under the generalized t-link conception model (GTR model) and Dunson and Stanford (2005) model (ECLLR model) based on the OK method. Furthermore, the estimated $\nu_1 = .8$ for the GTR model. In Figure IV, we present a comparison between the generalized t-link with $\nu_1 = .8$ and exponential link and note the differences in their shape. Consequently, using the exponential link in this situation would potentially have led to an underestimation or overestimation of the probability of conceptions in each cycle.

Table VI. Day-specific conception probabilities under the generalized t-link with random effect based on the LH method

	Day	Posterior Mean	Posterior SD	95% HPD Interval
Overall	-5	0.19376	0.07827	(0.03823, 0.33895)
	-4	0.24019	0.08544	(0.06730, 0.39814)
	-3	0.22737	0.08039	(0.06837, 0.37935)
	-2	0.18850	0.07630	(0.03842, 0.33239)
	-1	0.33275	0.06645	(0.19982, 0.45911)
	0	0.17833	0.07548	(0.02833, 0.31889)
	+1	0.23814	0.07470	(0.09140, 0.38232)
Parity=0	-5	0.14140	0.06629	(0.01666, 0.26613)
	-4	0.18015	0.07494	(0.03184, 0.31917)
	-3	0.16905	0.07016	(0.03381, 0.30220)
	-2	0.13715	0.06475	(0.01649, 0.26007)
	-1	0.26067	0.06584	(0.13104, 0.38884)
	0	0.12871	0.06268	(0.01205, 0.24745)
	+1	0.17805	0.06643	(0.04824, 0.30536)
Parity=1	-5	0.26746	0.10226	(0.06375, 0.45834)
	-4	0.32429	0.10870	(0.10871, 0.52963)
	-3	0.30892	0.10302	(0.10313, 0.50330)
	-2	0.26069	0.09970	(0.06031, 0.44507)
	-1	0.43471	0.08104	(0.27657, 0.59151)
	0	0.24816	0.10004	(0.04962, 0.43815)
	+1	0.32320	0.09553	(0.13739, 0.51048)

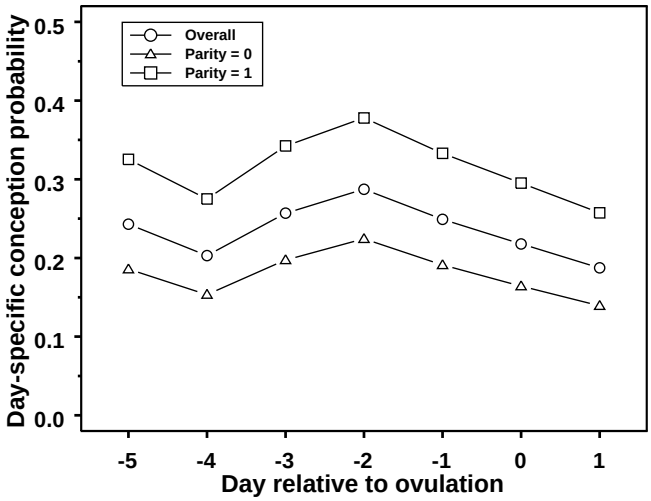


Figure I: Plots of estimated day-specific conception probability comparing the effect of parity under the generalized t-link model (GTR model), for the OK method

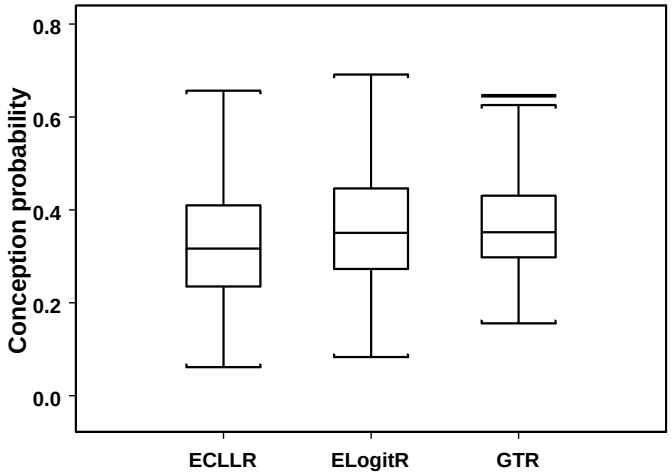


Figure II: Box-plots of estimated conception probability for various conception models for the OK method.

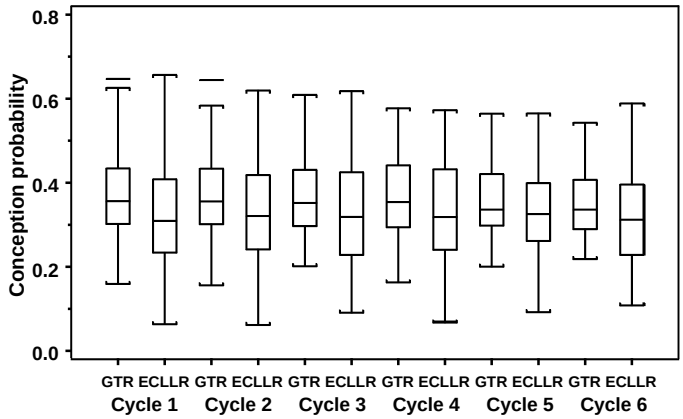


Figure III: Box-plots of estimated conception probability by cycles under the generalized t-link conception model with random effect (GTR) and Dunson and Stanford (2005) model (ECLLR) based on the OK method.

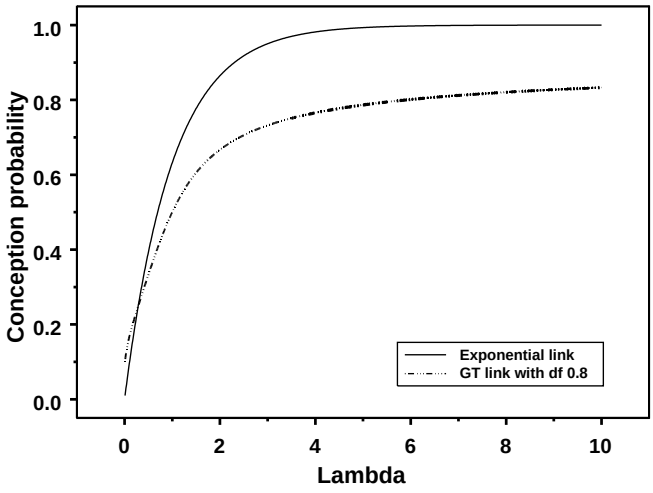


Figure IV: Comparison of the exponential link conception model with the generalized t-link conception model (GT link) with $\nu_1 = 0.8$, where $\Lambda = \sum_{k:X_{ijk}=1} \lambda_{ijk}$.

References

Albert and Chib, S. (1993). "Bayesian analysis of binary and polychotomous response data." *Journal of the American Statistical Association*, 88, 669-79. 775

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications, Inc.

Arellano-Valle, R. B. and Bolfarine, H. (1995). "On some characterizations of the t-distribution." *Statistics Probability Letters*, 25, 79-85. 776

Barrett, J. C. and Marshall, J. (1969). "The risk of conception on different days of the menstrual cycle." *Population Studies*, 23, 455-461. 772, 774

Buck Louis, G.M., Schisterman, E.F., Sweeney, A.M., Wilcosky, T.C., Gore-Langton, R.E., Lynch, C.D., Boyd Barr, D., Schrader, S.M., Kim, S., Chen, Z., and Sundaram, R. (2011). "Designing prospective cohort studies for assessing reproductive and developmental toxicity during sensitive windows of human reproduction and development—the LIFE Study." *Paediatric Perinatal Epidemiology*, 25, 413-424. 772

Chen, M.-H. and Dey, D. K. (1998). "Bayesian modeling of correlated binary responses via scale mixture of multivariate normal link functions." *Sankhya, Series A*, 60, 322-343. 775

Chen, M.-H., Dey, D. K., and Shao, Q.-M. (1999). "A new skewed link model for dichotomous quantal response data." *Journal of the American Statistical Association*, 94, 1172-1186 773

Chen, M.-H. and Shao, Q.-M. (2001). "Propriety of posterior distribution for dichotomous quantal response models with general link functions." *Proceedings of the American Mathematical Society*, 129, 293-302.

Cowles, C. and Carlin, B. P. (1996). "Markov chain Monte Carlo convergence diagnostics: A comparative review." *Journal of the American Statistical Association*, 91, 883-904. 788

Czado, C. and Santner, T. J. (1992). "The effect of link mis-specification on binary regression inference." *Journal of Statistical Planning and Inference*, 33, 213-231. 772, 774, 785

Dunson, D. B. (2001). "Bayesian models for distinguishing effects on the level and duration of fertility in the menstrual cycle." *Biometrics*, 57, 1067-1073. 772

Dunson, D.B., Baird, D.D., Wilcox, C.R. et al. (1999). "Day specific probabilities of clinical pregnancy based on two studies with imperfect measures of ovulation." *Human Reproduction*, 14, 1835-1839. 788

Dunson, D. B. and Stanford, J. B. (2005). "Bayesian inference on predictors of conception probabilities." *Biometrics*, 61, 126-133. 772, 773, 774, 775, 783, 784, 786, 787

Dunson, D. B. and Weinberg, C. R. (2000). "Accounting for unreported and missing intercourse in human fertility studies." *Statistics in Medicine*, 19, 665-679. 772

Dunson, D. B., Weinberg, C. R., Baird, D. D., Kesner, J. S., and Wilcox, A. J. (2001). "Assessing human fertility using several markers of ovulation." *Statistics in Medicine*, 20, 965-978. 772, 788

Dunson, D. B. and Zhou, H. B. (2000). "A Bayesian model for fecundability and sterility." *Journal of the American Statistical Association*, 95, 1054-1062. 772

Ecochard, R.(2006). "Heterogeneity in fecundability studies: issues and modelling." *Statistical Methods in Medical Research*, 15, 141-160. 772

Gini, C. (1926). "Decline in the birth rate and fecundability of women." *The Eugenics Review*, 17, 258-274. 772

Hanson, M.A. and Gluckman, P.D. (2008). "Developmental origins of health and disease: new insights." *Basic & Clinical Pharmacology & Toxicology*, 102, 90-93. 788

Hastings, W.K. (1970). "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika*, 57, 97-109. 780

Ibrahim, J.G., Chen. M.-H., and Sinha, D. (2001). *Bayesian Survival Analysis.* Springer-Verlag, New York. 784

Kim, S., Chen, M.-H., and Dey, D.K. (2008), "Flexible generalized t-link models for binary response data." *Biometrika*, 95, 93-106. 773, 776

Knaus, H. (1929). "Eine neue methode zur bestimmung des ovulationstermines." *Zentralblatt für Gynäkologie*, 53, 2193. 783

Liu, J.S. (1994). "The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem." *Journal of the American Statistical Association*, 89, 958-966. 779, 780

Mallick, B.K. and Gelfand, A.E. (1994). "Generalized linear models with unknown link functions." *Biometrika*, 81, 237-245. 773

Newton, M.A., Czado, C., and Chappell, R. (1996). "Bayesian inference for semiparametric binary regression." *Journal of the American Statistical Association*, 91, 142-153. 773

Ogino, K. (1930). "Ovulationstermin und konzeptionstermin." *Zentralblatt für Gynäkologie*, 54, 464-479. 783

Pyper C., Bromhall L., Dummett S., Altman D.G., Browbill P., and Murphy M. (2006). "The Oxford Conception Study design and recruitment experience." *Paediatric and Perinatal Epidemiology*, 20 (supplement), 51-59. 783

Royston, P. and Ferreira, A. (1999). "A new approach to modeling daily probabilities of conception." *Biometrics*, 55, 1005-1013. 772, 788

Scarpa, B. and Dunson, D.B. (2007). "Bayesian methods for searching for optimal rules for timing intercourse to achieve pregnancy." *Statistics in Medicine*, 26, 1920-1936. 788

Scheike, T. H. and Jensen, T. K. (1997). "A discrete survival model with random effects: an application to time to pregnancy." *Biometrics*, 53, 318-329. 772

Schwartz, D., MacDonald, P. D. M., and Heuchel, V. (1980). "Fecundability, coital frequency and the viability of ova." *Population Studies*, 34, 397-400. 772, 785, 791

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and Van Der Linde, A. (2002). "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society*, Series B, 62, 583-640. 783

Sundaram, R., McLain, A.C., and Buck Louis, G.M. (2011). "A survival analysis approach to modeling human fecundity." *Biostatistics*, doi:10.1093/biostatistics/kxr015. 772

Weinberg, C. R. and Gladen, B. C. (1986). "The beta-geometric distribution applied to comparative fecundability studies." *Biometrics*, 42, 547-560. 772

Weinberg, C. R., Gladen, B. C., and Wilcox, A. J. (1994). "Models relating the timing of intercourse to the probability of conception and the sex of the baby." *Biometrics*, 50, 358-367. 772

Wilcox, A.J., Weinberg, C.R., and Baird, D.D. (1995). "Timing of sexual intercourse in relation to ovulation." *The New England Journal of Medicine*, 333, 1517-1521. 788

WHO Task Forces on Methods for Determination of the Fertile Period (1980a). "Temporal relationships between ovulation and defined changes in the concentration of plasma estradiol-17β, luteinizing hormone, follicle stimulating hormone and progesterone. I. Probit analysis." *American Journal of Obstetrics and Gynecology*, 138, 383-390. 788

WHO Task Forces on Methods for Determination of the Fertile Period (1980b). "Temporal relationships between ovulation and defined changes in the concentration of plasma estradiol-17β, luteinizing hormone, follicle stimulating hormone and progesterone. II. Histologic dating." *American Journal of Obstetrics and Gynecology*, 139, 886-895. 788

Zhou, H. B., Weinberg, C. R., Wilcox, A. J., and Baird, D. D. (1996). "A random-effects model for cycle viability in fertility studies." *Journal of the American Statistical Association*, 91, 1413-1422. 772

Acknowledgements

The research work of the three authors, Drs. Kim, Sundaram, and Louis, is supported by the Intramural Research Program of National Institutes of Health, *Eunice Kennedy*

Shriver National Institute of Child Health and Human Development. This work was done when Dr. Kim was a research fellow with Dr. Sundaram. The research work of Dr Pyper was supported in part by the U.K. National Health Service Executive Primary Care Career Scientist and Service Research and Development Awards.