Model Selection With A Generalized Likelihood Ratio Approach In Fertility Studies

By Haibo Zhou

University of North Carolina, Chapel Hill, U. S. A.

e-mail: zhoubios.unc.edu

Minge Xie,

Rutgers University, Piscataway, U. S. A.

Douglas G. Simpson,

University of Illinois, Champaign, U. S. A.

Clarice R. Weinberg,

National Institute of Environmental Health Sciences, U. S. A.

SUMMARY

Data from fertility studies are naturally cluster-correlated due to the fact that menstrual cycles belonging to the same couple are biologically related. We discuss a generalized likelihood ratio approach for a fertility model for the probability of conception, where the mean structure is given as the product of two independent components, and each of the components can be modeled as a function of a generalized linear model. We use the proposed method to construct confidence intervals for parameters that are restricted on $[0, 1]$, and describe a model selection criteria for comparing different fertility models. The proposed approach provides range-preserving, asymmetric confidence intervals for the fertility parameters. Analysis of a data set from a North Carolina fertility study reveals some interesting findings that can not be observed with the traditional methods. The analysis suggests that prenatal exposure to maternal smoking has a deleterious effect on cycle viability while late menarche has a beneficial effect. The data do not support the existence of a negative effect of frequent ejaculation on sperm potency.

Some Key Words: Fertility; Generalized likelihood ratio; Latent variables; Quantile; Model selection critera; Weighted χ^2.

1. INTRODUCTION

The association between the likelihood of conception and both endogenous and exogenous factors is of interest to many people, including reproductive scientists, and couples who wish to begin or avoid a pregnancy. Estimating and identifying effects of these factors on the fecundability, the probability of conception in a given menstrual cycle for a couple who is not using any contraceptive method, is of particular public health interests. Hypothesis testing and model selection in the traditional fertility models rely on the Wald statistics and the likelihood ratio statistics (chi-square distributed). This is not satisfactory since the correlation among menstrual cycles belonging to the same couples is not accounted for. The confidence intervals using the Wald statistics are not range preserving to reparameterization and the asymptotic distribution of the likelihood ratio statistic is not exactly chi-square, due to the cluster-correlated nature of the data. Valid inference procedures are important in understanding the biological processes involved in human fertility. Our research is motivated by the need to carry out inference related to fertility parameters, and provide criteria for model selection in fertility studies.

We begin with some basic biological background for human reproduction. For women of reproductive age, a single ovum is normally produced and released around the middle of each menstrual cycle in a process called ovulation. For pregnancy to occur, the ovum has to be fertilized by sperm that reach the woman's reproductive tract. These sperm can come from intercourse on different days around the time of ovulation. If the ovum is not fertilized during its viable period, it will degenerate and a new cycle will begin. If the ovum is fertilized, however, the specific act of intercourse during the cycle that caused a conception is usually indeterminate. In addition to the timing and frequency of intercourse, other endogenous and exogenous factors may also affect the probability of conception. Many of these factors are unobservable, but manifest themselves in dependency among outcomes within multiple cycles from the same couple. Other factors, such as age and exposure to a potential toxicant, can be observed directly. Valid inference procedure and model selection criteria for human fertility models should account for the dependency among outcomes within the same couple.

We develop a generalized likelihood ratio (GLR) analysis to carry out the inference for fertility parameters. The proposed GLR method, a generalization of the standard likelihood ratio test under independent settings, is suitable for batch-correlated observations. The statistic has an asymptotic distribution that is a weighted sum of χ_1^2. Through the use of GLR in fertility studies, we provide model selection criteria for fertility models and provide range-preserving, asymmetric confidence intervals for fertility parameters.

The rest of the paper is organized as follows. In Section 2 we present the statistical models for fertility studies. In Section 3 we describe the basic asymptotic theory of the GLR statistics under these models and discuss how to construct a hypothesis test using the GLR statistics. In Section 4, we analyze the data from the North Carolina fertility study (Wilcox et al., 1988) to carry out the inference on fertility parameters and compare a sequence of models. We give final remarks in Section 5. An iterative algorithm for computing the quantiles of the weighted chi-squares is given in the Appendix.

2. MARGINAL MODELING OF FERTILITY

The data from the fertility studies we will consider are cluster-correlated data, where outcomes from the same couple are correlated. To fix notation, for each couple i, where $i = 1, 2, \cdots, N$, there are binary outcomes Y_{ij} for $j = 1, 2, \cdots, n_i$ cycles, where Y_{ij} is 1 if there is a conception signaled by the rise in pregnancy hormone, hCG, for cycle j, and 0 otherwise. We remark that there could be multiple $Y_{ij} = 1$ for couple i if $n_i > 1$ and a conception failed prior to clinical detection. The outcome of each such cycle depends on a susceptibility factor (cycle viability), and the aggregate of a set of unobservable Bernoulli trials (the effects of intercourse on various days). These two independent components are assumed to contribute to the probability of conception in a multiplicative way. For a given cycle j in couple i, we assume there are independent competing risks of fertilization from the days on which there was intercourse, any of which could cause conception in a viable cycle. A menstrual cycle is viable if the ovum is viable and other factors, such as receptivity of the uterus, are such that conception is theoretically possible.

To keep track of when intercourse occurs, let x_{ijk} be an indicator variable such that $x_{ijk} = 1$ if couple i had intercourse on day k of cycle j, and $x_{ijk} = 0$ otherwise. Here k is indexed relative to the day of ovulation (day k=0). It is negative prior to ovulation and positive afterwards. Let $g_1(\beta' u_{ij}) \in [0, 1]$ be the probability that cycle j for couple i is viable, where $u_{ij}{}' = (u_{ij1}, \cdots, u_{ijq})$ is a covariate vector and $\beta' = (\beta_1, \cdots, \beta_q)$ is the corresponding regression parameter vector. Let $g_2(\alpha' v_{ijk}) \in [0, 1]$ be the probability that intercourse on the kth day would have succeeded if the other competing risks had not been present (no intercourse on any other day) and the cycle had been viable, where $v'_{ijk} = (v_{ijk1}, \cdots, v_{ijkp})$ is the associated covariate vector and $\alpha' = (\alpha_1, \cdots, \alpha_p)$ is the corresponding regression parameter vector. Conception occurs in a given cycle if and only if the cycle is viable and at least one of the days with intercourse causes fertilization. The functions g_1^{-1} and g_2^{-1} are assumed probability link functions such as the logit or probit.

The observed data are of the form $\{Y_{ij}, X_{ij}, U_{ij}, V_{ij}\}$, where $\{X_{ij}, U_{ij}, V_{ij}\}$ is a matrix containing observations $\{x_{ijk}, u_{ij}, v_{ijk}\}$ for couple i in cycle j. A marginal type model for clustered data, building on earlier work in fertility modeling, (e.g., Barrett and Marshall 1969, Schwartz, MacDonald and Heuchel 1980, Royston 1982, and Weinberg, Gladen, and Wilcox 1994), has been proposed (Zhou and Weinberg, 1996) that specifies the marginal probability of conception in a given menstrual cycle j as

$$\mu_{ij} = \Pr(Y_{ij} = 1 | X_{ij}, U_{ij}, V_{ij}) = g_1(\alpha' u_{ij}) \left[1 - \prod_k (1 - g_2(\beta' x_{ijk}))^{x_{ijk}} \right]. \tag{1}$$

We will refer to the cycle viability term $g_1(\alpha' u_{ij})$ as the A component and the remaining day-specific probability term in (1) as the P component.

The definition of marginal mean for this model is different from the usual generalized linear model formulation. Indeed, the marginal mean is modeled as a product of two components, A and P, where each component can be viewed as a function of a generalized linear model formulation. An important sub-model of (1) sets $g_2(\beta' v_{ijk}) \equiv p_k$, for $k = -K, \ldots, 0, \ldots, K$. This model is used to study the length of the fertile interval (Wilcox, Weinberg and Baird, 1995). By estimating the probability (p_k) of conception associated with each day and its confidence interval, we can estimate how many consecutive days have $p_k > 0$ and therefore estimate the beginning and end of the fertile window.

Denote $\Theta' = (\alpha', \beta')$. Θ can be estimated by maximizing the pseudo-likelihood function $L(\Theta) = \prod_{i=1}^{N} \prod_{j=1}^{n_i} L_{ij}(\Theta)$ (Zhou and Weinberg, 1996) where

$$L_{ij}(\Theta) = \left[g_1(\alpha' u_{ij}) \left(1 - \prod_k (1 - g_2(\beta' x_{ijk}))^{x_{ijk}} \right) \right]^{Y_{ij}} \left[1 - g_1(\alpha' u_{ij}) \left(1 - \prod_k (1 - g_2(\beta' x_{ijk}))^{x_{ijk}} \right) \right]^{1-Y_{ij}}.$$

Then the maximum pseudo-likelihood estimator $\widehat{\Theta}$ can be obtained by solving the following unbiased estimating equation:

$$U(\widehat{\Theta}) = \sum_{i=1}^{N} \psi_i \left(\widehat{\Theta} \right) = 0, \tag{2}$$

where

$$\psi_i(\Theta) = \sum_{j=1}^{n_i} \frac{\partial}{\partial \Theta} \log\{L_{ij}(\Theta)\}.$$

The robust sandwich estimator of the covariance matrix of $\widehat{\Theta}$ is then $\widehat{\text{cov}}(\widehat{\Theta}) = M_0^{-1} M_1 M_0^{-1^T}$, where $M_0 = \sum_{i=1}^{N} \frac{\partial}{\partial \Theta^T} \psi_i \left(\widehat{\Theta} \right)$ and $M_1 = \sum_{i=1}^{N} \psi_i \left(\widehat{\Theta} \right) \psi_i^T \left(\widehat{\Theta} \right)$. Following Zhou and Weinberg, we use E-M algorithm (Dempster, Laird, and Rubin, 1977) that is based on a set of latent variables to compute $\widehat{\Theta}$. Let the indicator variable Z_{ij} denote whether the particular cycle ij is viable. For conception cycles, Z_{ij} is known to be 1. For nonconception cycles, Z_{ij} is unobservable. Similarly, we let S_{ijk} be a binary indicator that corresponds to whether pregnancy would have occurred had the cycle been viable and had there been no day other than day k with intercourse in cycle j for couple i. S_{ijk} is an unobservable variable except for days on which there was no intercourse ($S_{ijk} = 0$), or for cycles where conception occurred and there was only one day with intercourse ($S_{ijk} = 1$). We estimate Θ from the following complete-data likelihood function:

$$\prod_{i=1}^{N} \prod_{j=1}^{n_i} g_1(\alpha' u_{ij})^{Z_{ij}} \left(1 - g_1(\alpha' u_{ij}) \right)^{1-Z_{ij}} \prod_k g_2(\beta' x_{ijk})^{X_{ijk} S_{ijk}} \left(1 - g_2(\beta' x_{ijk}) \right)^{X_{ijk}(1-S_{ijk})}.$$

The conditional expected value of S_{ijk} and Z_{ij} given the observed data are given in Zhou and Weinberg[6].

As is well known, the Wald test based on the standard error is not transformation invariant, and the confidence intervals for p_k based on the Wald statistics are restricted to be symmetric.

Further there is no valid criterion in the above marginal model for model selection even with nested models. We next discuss a generalized likelihood ratio statistic which can be used for inference about parameters and model selection criterion.

3. GENERALIZED LIKELIHOOD RATIO INFERENCE

Likelihood ratio inference when the underlying distribution function is misspecified has been considered by several authors in the standard i.i.d setting (e.g., Foutz and Srivastava, 1977 and Kent, 1982). Rotnitzky and Jewell (1990) studied the GLR statistics for cluster-correlated data under the semiparametric generalized linear model proposed by Liang and Zeger (1986). They established the asymptotic distribution of the GLR statistics under the "working independent" correlation matrix. The method has been used for binary and ordinal response regression (Xie, Simpson and Carroll, 1997). Our fertility model defined in (1) is different from the generalized linear model considered by Rotnitzky and Jewell. The mean structure in our model is a product of two independent components (A and P), and each of the components can be modeled as a function of a generalized linear model. The two covariate vectors in (1) capture the cluster-level and individual-level information, respectively.

Partition the parameter Θ as $\Theta = (\gamma', \lambda')'$, where $\gamma = (\gamma_1, \cdots, \gamma_r)'$ is the subset of Θ. We are interested in the hypotheses test $H_0 : \gamma = \gamma_0$, and $\lambda = (\lambda_1, \cdots, \lambda_{p+q-r})'$ is the remaining subset of parameters. Let $\widehat{\Theta}_0 = (\gamma_0', \widehat{\lambda}')'$ be the estimator of Θ subject to $\gamma = \gamma_0$, then the generalized likelihood ratio statistic is defined as

$$T_L(\gamma) = 2\left[\log L(\widehat{\theta}) - \log L(\widehat{\theta}_0)\right].$$

If $Y_{ij}, j = 1, \cdots, n_i$ are independently distributed, $T_L(\gamma)$ is asymptotically distributed as chi-square with degree of freedom r under very mild regularity conditions. In our model, the independence assumption is violated and $T_L(\gamma)$ is no longer asymptotically Chi-square distributed. Define matrices $J(\Theta)$ and $H(\Theta)$ as

$$J(\Theta) = E\{U(\Theta)\, U(\Theta)'\},$$

and

$$H(\Theta) = -E\left\{\frac{\partial}{\partial\Theta'}U(\Theta)\right\}.$$

For the asymptotic distribution theory of GLR, we have the following general result, which is adapted from similar results in the literature (Foutz and Sirvastava, 1977, Kent, 1982, and Rotnitzky and Jewell, 1990). A sketch of proof is outlined in the Appendix.

Theorem 1 *Under mild regularity conditions and model specification (1) with $\gamma = \gamma_0$, $T_L(\gamma)$ converges in distribution to $\sum_{j=1}^{r} d_j\chi_j^2$ as the number of clusters increases, where $\chi_j^2, j = 1, \cdots, r$ are independent χ_1^2 random variables, and $d_1, \cdots, d_r$ are the eigenvalues of $H_{\gamma\gamma.\lambda}(\Theta)(H^{-1}(\Theta)J(\Theta)H^{-1}(\Theta))_{\gamma\gamma}$, where $H(\Theta)_{\gamma\gamma.\lambda} = H(\Theta)_{\gamma\gamma} - H(\Theta)_{\gamma\lambda}H^{-1}(\Theta)_{\gamma\gamma}H(\Theta)_{\lambda\gamma}$.*

The asymptotic distribution of the GLR statistics $L_T(\gamma)$ in the theorem enables us to construct a confidence region for parameters of interest based on the generalized profile likelihood function. Specifically, we define the generalized $100\%\alpha$ confidence region for γ as the set,

$$\{\gamma : L_T(\gamma) \leq Q_\alpha(\Theta)\}, \tag{3}$$

where $Q_\alpha(\Theta)$ is the $100\%\alpha$ quantile of the weighted chi-square distribution given in Theorem 1. We would replace $Q_\alpha(\Theta)$ by $Q_\alpha(\widehat{\Theta})$ in practice.

To utilize the above theorem in fertility studies, we need to estimate the matrices $J(\Theta)$ and $H(\Theta)$, and to calculate the quantiles of the weighted chi-square distribution. To estimate $H(\Theta)$, one can simply use the empirical $M_0(\widehat{\Theta})$. For $J(\Theta)$, we suggest $(1 - r/N)M_1(\widehat{\Theta})$, although in principle any consistent estimate would work. Several methods can be used to calculate the quantile of the weighted sum of chi-squares. For $r = 1$, the appropriate distribution is simply a scaled χ_1^2 distribution. Boos (1992) discussed the difficulties in practical use of generalized likelihood ratio test when $r > 1$. Some authors (e.g., Rao and Scott, 1981, Robert, Rao and Kumar, 1987 and Rotnitzky and Jewell, 1990) approximated the weighted sum of chi-square distribution with a standard chi-square distribution. We used an iterative algorithm for calculating quantiles of the weighted Chi-square distribution. This algorithm is provided in the Appendix.

4. NORTH CAROLINA FERTILITY STUDY

We analyze the data from the Early Pregnancy Study (Wilcox et al., 1988) conducted in North Carolina using the proposed method. We use the GLR statistics to construct an asymmetric confidence interval for the probability of conception due to a specific day's intercourse, and to draw inference about some important covariates including women's age, her age at menarche, whether she was exposed to her mother's smoking while she was *in utero*, and the effect of frequent intercourse on the probability of conception. Finally, we will use the GLR statistics as model selection criteria.

The Early Pregnancy Study (Wilcox et al. 1988) enrolled couples with no known fertility problems who were ready to begin a pregnancy. The couples enrolled in the study when they stopped contraception. Women collected a first urine specimen each morning, and kept a daily diary recording whether they had had unprotected intercourse in the preceding 24 hours and whether they had had any menstrual bleeding during that time. Participation continued through 8-weeks into a pregnancy or for 6 months if no clinical pregnancy was identified. Most women were in their mid-twenties to early thirties. Urine specimens were assayed for levels of metabolites of ovarian hormones and these results were used to identify the day of ovulation (Baird et al., 1991). A special feature of the North Carolina study is that "success" as defined by the implantation, can be detected through bioassay on the daily urine specimen (Baird et al., 1991, Wilcox, Weinberg and Baird, 1995). This is signaled by hCG, the human pregnancy hormone, which rises exponentially around the time of implantation. Following Wilcox, Weinberg and Baird (1995), the outcome variable in our analysis for a particular menstrual cycle is defined by whether there is a rise in the human gonadotropin (hCG) around the putative time of implantation. Hence, because we are treating both occult conceptions (lost prior to clinical diagnosis) and clinical pregnancies as success, there could be multiple successes (implantations) for a couple with multiple cycles in the study. The present analysis is based on 555 menstrual cycles from 198 women, with 177 successes (implantations) and 378 failures.

4.1. Estimating of the day-specific probability of conception

As we mentioned in Section 2, a sub-model of (1) with $g_2(\beta' v_{ijk}) \equiv p_k$ allows the nonparametric modeling of the day-specific conception probability. The Wald statistic is especially limited in estimating the day-specific probability of conception as the probabilities are confined to [0,1] and their confidence intervals need to be asymmetric. Applying generalized likelihood ratio inference to fertility models, we calculate the 95% confidence intervals for the conception probabilities associated with intercourse on each of the 6 days up to and including the day of ovulation in each menstrual cycle. We contrast the confidence intervals with those obtained from the Wald statistics.

To calculate the 95% confidence intervals, we fit the following simple non-parametric model. Denote $p_k, k = -5, -4, \cdots, 0$ as the probability of fertilization due to intercourse on the kth day in a viable cycle, where day 0 is the ovulation day. For the North Carolina data, only those 6 days had a nonzero probability (Wilcox, Weinberg and Baird, 1995). Take $g_1(\alpha' u_{ij}) \equiv A$, and $g_2(\beta' v_{ijk}) \equiv p_k$, then model (1) becomes

$$\mu_{ij} = \Pr(Y_{ij} = 1 | X_{ij}, U_{ij}, V_{ij}) = A \left[1 - \prod_{k=-5}^{0} (1 - p_k)^{x_{ijk}} \right].$$

This is equivalent to letting design matrix V to be a 555×6 matrix, where each column is a indicator variable for intercourse on the corresponding day. The estimation procedure is the one outlined in Section 2. The cycle viability probability, A, is estimated at .37 with estimated standard error of 0.04. Table 1 lists the fitted point estimates and standard errors for the p_k's, their eigenvalues and asymptotic 95%th quantile based on Theorem 1 and the 95% quantile of the naive χ_1^2. The corresponding 95% fitted generalized profile likelihood confidence intervals based on (3) and those based on the Wald statistics for p_k are plotted in Figures 1-2. Clearly except for p_{-4}, all the 95% confidence intervals for other p_k's based on the Wald standard errors extend outside the parameter space either above or below (Figure 1).

We compare a sequence of models with the GLR criteria to see if the two component model fits the data better than would a single constant. Table 2 provides 4 models we considered. The p-values listed in Table 2 are based on the asymptotic sum of weighted chi-squares. The first one is a simple grand mean model. The second one is a two components model where the success probability of each day's intercourse act is a constant p and

$$\mu_{ij} = \Pr(Y_{ij} = 1 | X_{ij}, U_{ij}, V_{ij}) = A\left[1 - (1-p)^{\sum_{k=1}^{6} x_{ijk}}\right].$$

The two-components model fit the data statistically better than the grand mean model ($p = 0.022$). Notice from Figure 1 that the probability of success rises fairly linearly in time. Model 3 in Table 2 captures this by modeling $p_k = (1 + \exp(-\alpha - \beta * k))^{-1}$, i.e. logit$(p_k)$ is a simple linear function of the time index, k. The hypothesis test for $H_0 : \beta = 0$ is significant with p-value at 0.018. Jumping to model 4, the same non-parametric model studied in Table 1, we see that with 5 more degrees of freedom than model 2, this model does not significantly improve the fit over model 2. Therefore model 3 is clearly the parsimonious choice. Note here we used the property of transformation invariance of the GLR statistics for different models. The fitted point estimate $\hat{\beta}$ in model 3 is 0.63 with standard error of 0.346. The 95% confidence interval based on the asymptotic normality of Wald statistic is then (-0.07, 1.28) which give an equivocal result for testing $H_0 : \beta = 0$. Using GLR we obtain the 95% confidence interval (0.13, 1.42) which excludes 0 at level 0.05.

4.2. Inference on covariates and selection of the best fitted model

Based on the chosen parsimonious model for intercourse (model 3 in Table 2), we now incorporate covariates into the fertility model and use the GLR statistics to select a best fitted fertility model. The covariates considered here are $PSMK$, which is a binary variable indicating whether the women was exposed to her mother's smoking while she was *in utero*, AGE, which is the women's current age, $MAGE$, which is the age at menarche (age at first menstrual period), and $DDAY$, where $DDAY_k = 1$ if there was consecutive intercourse on both day k and $k-1$ and $DDAY_k = 0$ otherwise. Each of these covariates could plausibly affect fertility. Components of cigarette smoke are known to be harmful to the oocytes in animal studies (e.g., Kristensen et al., 1995 and MacKenzie, and Angevine, 1981). Fecundability decreases as women age (Bongaarts, 1983). It is also suspected that too frequent intercourse may reduce the potency of the sperm (Auger et al., 1995). We model both A and P with logit link function, that is, covariate vectors

u_{ij} and v_{ijk} in the respective A and P components in (1) is modeled as the following

$$\begin{aligned} g_1(\alpha_0 + \alpha_1' \times u_{ij}) &= \{1 + \exp(-\alpha_0 - \alpha_1' \times u_{ij})\}^{-1}, \\ g_2(\beta_0 + \beta_1 \times k + \beta_2' \times v_{ijk}) &= \{1 + \exp(-\beta_0 - \beta_1 \times k - \beta_2' \times v_{ijk})\}^{-1} \end{aligned} \quad (4)$$

Table 3 summarizes the model selection process using the GLR statistics. The hypothesis tests are based on nested models listed under three sections, a, b and c, where covariates are only modeled in P component, only in A component and in both A and P components, respectively. From section a, the PSMK and MAGE are significant. Model $a5$ is the best fitted model if we only model the covariates in P components. Section b in Table 3 show that PSMK is still significant in A component and MAGE is only marginally significant (p=0.064). AGE and MAGE are marginally significant given PSMK is in the model (p=0.081 and p=0.064 respectively). Results from section a and b suggest that MAGE and AGE are slightly better fitted with the P component while PSMK is better fitted with the A component. We further consider some nested models in section c of Table 3 to allow covariates to influence both components. Since $DDAY$ would be expected to act through effects on sperm quality, we only model it in P. The p-value for testing $DDAY$ is not significant (p=0.76, model $c.5$ versus $c.3$). The most parsimonious model in this section is model $c.3$, where PSMK is significant in A and MAGE is significant in P. Model $c.6$ suggests that the addition of AGE is marginally significant (p=0.081). Since participant's ages are relatively close (most in mid twenties to thirty), and age is known to affect women's fecundability (older women are less likely to conceive), it would be reasonable to keep AGE in the final model. Hence the final model by the GLR statistics is model $c.6$ with PSMK and AGE in the cycle viability component, and days relative to ovulation and age at menarche in the day-specific conception probability component.

Table 4 lists the estimated regression coefficients and the standard error for model $c.6$ of Table 3. We contrasted the Wald test and the GLR test by comparing the p-values for testing effect of each covariate in the model. Both tests agree on the effect of PSMK. The Wald test yields only marginal significance for days relative to ovulation (p=0.057) while nonsignificant effects on MAGE (p=0.101) and AGE (p=0.112). Clearly a different model would be selected based on the Wald test. The MAGE effect is only picked up by the GLR test.

4.3. Summary of the North Carolina Fertility Study analysis

In summary, the data suggest that prenatal exposure to maternal smoking has a deleterious effect on cycle viability. This is consistent with animal experiments showing a loss of primordial oocytes in newborns prenatally exposed through the mother to components of cigarette smoke (e.g., Kristensen et al., 1995 and MacKenzie, and Angevine, 1981). In this way, subfertility could be seen as a subtle birth defect associated with maternal smoking. The data also provide some evidence for a negative trend with age, and a beneficial effect of late menarche. While the best-fitting model puts the menarche effect in the model component related to intercourse, rather than in the component related to cycle viability, the data do not allow us confidently to choose the one pathway over the other. While it seems more likely a priori that a female factor such as age at menarche would modify cycle viability, female factors can also influence ovum fertilization. The data do not support the existence of a negative effect of frequent ejaculation on sperm potency.

5. DISCUSSION

We have discussed and applied a generalized likelihood ratio statistic for inference and model selection in fertility models. The proposed method would be applicable in various medical research settings where the available data are naturally cluster- or batch-correlated. The GLR method is a dependent-data generalization of the standard likelihood ratio test, which is only valid in independent settings. The GLR method reduces to the standard likelihood ratio method if independence holds. Like the standard likelihood ratio statistics, the GLR statistic has some advantages, such as invariance under parameter transformation, over the Wald-type test. The proposed method can be used for model selection, as demonstrated in the Early Pregnancy Study. Unlike the usual likelihood ratio statistic, we have shown the GLR statistic follows a weighted sum of chi-squares distribution asymptotically. We used a marginal model for the batch-correlated binary outcomes where the probability of success is modeled as a product of two components with each component a function of probabilities that follow a generalized linear model. This model allows the distinction of two levels of covariate, the cycle level and the day level.

Application of the GLR method to data from Early Pregnancy Study showed some very interesting results that can not be observed using existing methodology. Using GLR as model selection criteria also yields a more biologically plausible model than would from the Wald test. The data suggest that prenatal exposure to maternal smoking has a deleterious effect on cycle viability. The data also provide some evidence for a negative trend with age, and a beneficial effect of late menarche. While the best-fitting model puts the menarche effect in the model component related to intercourse, rather than in the component related to cycle viability, the data do not allow us confidently to choose the one pathway over the other. The data do not support the existence of a negative effect of frequent ejaculation on sperm potency.

We present algorithms to evaluate the quantiles of the appropriate distributions in the Appendix. Considering the current advances in computing power, a feasible alternative approach might be to simulate the empirical weighted chi-square distribution quantile. However, such an approach will not produce an unique quantile of interest each time.

ACKNOWLEDGEMENTS

The authors thank Drs. D.D. Baird and A.J. Wilcox for sharing the Early Pregnancy Data.

APPENDIX A: Sketch of Proof for Theorem 1

The proof of Theorem 1 first establishes the asymptotic normality of $\sqrt{n}(\widehat{\Theta} - \Theta)$. Notice that the estimating equation (2) corresponds to the independence equations in Liang and Zeger (1986) with working correlation matrix taken to be the identity matrix, except that the specification of the marginal mean now is a product of two independent components given by (1). Therefore normality can be developed by extending the argument of Liang ad Zeger (1986). The impact of (1) is on the form of D matrix where $D_i = \partial\mu_i/\partial\Theta$. With normality established, the theory by Rotnitzky and Jewell (1990) for cluster-correlated data can be invoked to establish the asymptotic distribution of the generalized likelihood ratio statistics. The result can also be derived from a similar argument of Kent (1982) in the independent case.

APPENDIX B: Calculating Quantiles of the Weighted Chi-square

We adopt Ruben's (1962) series to our weighted sum of χ_1^2 distribution, and provide a numerical evaluation of T based on a Newton-Raphson iteration procedure. Write $F^*(t) = Pr(\sum_{j=1}^{r} d_j W_j \leq t | d_1, \ldots, d_r)$ and $f^*(t) = \partial F^*(t)/\partial t$ as the cdf and df for the weighted sum of chi-squares, where W_i are independent χ_1^2 random variables. The following equation holds uniformly on every finite interval of t (Ruben, 1962, Theorem 2):

$$F^*(t) = \sum_{j=0}^{+\infty} c_j F_{r+2j}(\frac{t}{\delta}), \quad f^*(t) = \sum_{j=0}^{+\infty} \frac{c_j}{\delta} f_{r+2j}(\frac{t}{\delta}) \tag{5}$$

where $F_{r+2j}(t)$ and $f_{r+2j}(t)$ are the cdf and df for the central chi-square random variable with degrees of freedom $r+2j$, δ is an arbitrary positive number and c_j are constants which are recursively defined by $c_0 = \prod_{l=1}^{r} (\delta/d_i)^{1/2}$, $c_j = (2j)^{-1} \sum_{s=0}^{j} \{\sum_{l=1}^{r} (1-\delta/d_l)^{j-s}\} c_s$, $j = 1, \ldots, r, \ldots$. Particularly, when $\delta < \lambda_{\min} = \min\{\lambda_i, i = 1, \ldots, r\}$, we have $c_i > 0$ and $\sum_{i=1}^{k} c_i = 1$. We use $\delta = 0.90625\lambda_{\min}$ as suggested by Farebrother (1984). Since both $F(t)$ and $f(t)$ decrease with degrees of freedom, the tail sums of series (5) are bounded by

$$\sum_{j=N}^{+\infty} c_j F_{r+2j}(\frac{t}{\delta}) \leq (1 - \sum_{j=0}^{N-1} c_j) F_{r+2N}(\frac{t}{\delta}),$$

and

$$\sum_{j=N}^{+\infty} \frac{c_j}{\delta} f_{r+2j}(\frac{t}{\delta}) \leq \frac{1}{\delta}(1 - \sum_{j=0}^{N-1} c_j) f_{r+2N}(\frac{t}{\delta}),$$

respectively. The left hand sides can be evaluated at a given accuracy since the bounds of above inequalities go to zero as N increases. Denote T_α as the α-th quantile of $\sum_{i=1,r} d_i W_i$, which satisfies $F(T_\alpha) = 1 - \alpha$. We calculate T_α by the following Newton-Raphson iteration:

$$T_\alpha^{(i+1)} = T_\alpha^{(i)} - [f(T_\alpha^{(i)})]^{-1}[F(T_\alpha^{(i)}) - 1 + \alpha]. \tag{6}$$

The calculation was done with a special FORTRAN subroutine and S-plus driver function. The starting value used was $t_0 = (d_{\min} + d_{max})/2\chi_r^2(\alpha)$ where $\chi_r^2(\alpha)$ is the α-th quantile of standard central χ_r^2 distribution, and $d_{\min}$, d_{max} are minimum and maximum eigenvalues in Theorem 1 respectively.

REFERENCES

Auger, J., Kunstmann, J. M., Czyglik, F. and Jouannet, P. (1995) Decline in semen quality among fertile men in Paris during the past 20 years. *New England Journal of Medicine*, 332, 281-5.

Baird, D. D., Weinberg, C. R., Wilcox, A. J., McConnaughey, D. R., and Musey, P. I. (1991). Using the ratio of urinary estrogen and progesterone metabolites to estimate day of ovulation. *Statistics in Medicine*, 10, 255-266.

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

Bongaarts, J. (1983). The proximate determinants of natural marital fertility. *Determinants of Fertility in Developing Countries*, Vol. 1. Academic Press, 103-38.

Boos, D. D. (1992). On generalized score tests. *The American Statistician*, 46, 327-333.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series* B 39, 1-38.

Farebrother, R. W. (1984). The distribution of a positive linear combination of χ^2 random variables. *Applied Statistics*, 33, 332-339.

Foutz, R. V. and Srivastava, R. C. (1977). The performance of the likelihood ratio test when the model is incorrect. *The Annals of Statistics*, 5, 1183-1194.

Kent, J. T. (1982). Robust properties of likelihood ratio tests. *Biometrika*, 69, 19-27.

Kristensen, P., Eilertsen, E., Elín, E., Haugen, A., Skaug, V. and ϕvrebo, S. (1995). Fertility in mice after prenatal exposure to benzo[a]pyrene and inorganic lead. *Environmental Health Perspectives* 103, 588-590.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.

MacKenzie, K. M. and Angevine, D. M. (1981). Infertility in mice exposed *in utero* to benzo(a)pyrene. *Biol. Reprod* 41, 580-586.

Rao, J. N. K. and Scott, A. J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. *J. Am. Statist. Assoc.*, 76, 221-30.

Roberts, G., Rao, J. N. K. and Kumar, S. (1987). Logistic regression analysis of sample survey data. *Biometrika*, 74, 1-12.

Rotnitzky, A. and Jewell, N. P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* 77, 485-97.

Royston, J. P. (1982) Basal body temperature, ovulation and the risk of conception, with special reference to the lifetime of sperm and egg. *Biometrics*, 38, 397-406.

Ruben, H. (1962). Probability content of regions under spherical normal distributions. IV: The distribution of homogeneous and non-homogeneous quadratic functions of normal variables *The Annals of Mathematical Statistics*, 33, 542-570.

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

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.

Wilcox, A. J., Weinberg, C. R., O'Conner, J. F., et al. (1988). Incidence of early pregnancy loss. *New England Journal of Medicine*, 319 189-194.

Wilcox, A. J., Weinberg, C. R., Baird, D. D. (1995). Timing of intercourse in relation to ovulation: Effects on the probability of conception, survival of pregnancy, and sex of the baby. *New England Journal of Medicine*, 333(23): 1517-1521.

Xie, M., Simpson, D.G. and Carroll, R.J. (1997). Scaled link functions for heterogeneous ordinal response data. In *Modelling Longitudinal and Spatially Correlated Data: Methods, Applications and Future Directions*, Springer-Verlag, Lecture Notes in Statistics, p. 23-36.

Zhou and Weinberg, C. R. (1996). Modeling conception as an aggregated Bernoulli outcome with latent variables via the EM algorithm. *Biometrics* 52, 945-954.

Table 1. Semiparametric Model for Intercourse-related Probabilities in a 6 Day Window. p_0 is the probability for the Day of Ovulation. $T_{.95}$ is the 95th percentile of the rescaled χ_1^2s. $\chi^2_{.95,1}$ is the 95th percentile of the central chi-square with degree of freedom 1.

parameter	estimates/s.e.	eigenvalue	$T_{.95}$	$\chi^2_{.95,1}$
p_{-5}	0.25/0.22	0.80	3.08	3.84
p_{-4}	0.47/0.20	1.10	4.21	3.84
p_{-3}	0.31/0.24	1.04	4.00	3.84
p_{-2}	0.75/0.22	0.95	3.67	3.84
p_{-1}	0.82/0.21	1.05	4.03	3.84
p_0	0.84/0.25	1.02	3.90	3.84

Table 2. Models for regression analysis of North Carolina EPS data. E is the prenatal smoking exposure status. k here indexes the day relative to ovulation (day 0), where $k = -5, \cdots, -1, 0$. Ovulation day in each cycle is estimated by the pattern of change in the ratio of two hormone levels, estrogen and progesterone.

Model	Components of the Model		Dev.	d.f.	Model	p-value
	Cycle Viability	Day-specific			Compared	
1.	A		694.9	1	–	
2.	A	p	689.7	2	1	0.022
3.	A	$\text{logit}^{-1}(\alpha + \beta \times k)$	684.1	3	2	0.018
4.	A	$p_{-5}, \cdots, p_0$	683.1	6	3	0.284

Table 3. Models for regression analysis of North Carolina EPS data. PSMK is the prenatal smoking exposure status, MAGE is the woman's age of first menstrual period, AGE is her current age, k here indexes the day relative to ovulation (day $k = 0$), where $k = -5, \cdots, -1, 0$, and DDAY is the indicator of the second consecutive day with intercourse. 1's in columns 3 and 4 denote intercepts. The models for components A and P are based on formulae in (1) and (4).

Model	Deviance	U_{ij} in A	v_{ijk} in P	d.f	p-value	weights	Models compared
a.1	684.10	1	1 k	3			
a.2	673.56	1	1 k psmk	4	0.001	1.04	a1
a.3	677.19	1	1 k mage	4	0.015	1.16	a1
a.4	683.82	1	1 k age	4	0.644	1.31	a1
a.5	665.41	1	1 k mage psmk	5	0.009	1.21	a2
					0.001	1.10	a3
a.6	665.14	1	1 k mage psmk age	6	0.688	1.69	a5
b.1	684.10	1	1 k	3			
b.2	669.53	1 psmk	1 k	4	0.001	1.26	b1
b.3	679.36	1 mage	1 k	4	0.037	1.08	b1
b.4	682.26	1 age	1 k	4	0.287	1.63	b1
b.5	665.38	1 mage psmk	1 k	5	0.064	1.02	b2
					0.001	1.24	b3
b.6	661.66	1 mage psmk age	1 k	6	0.081	1.22	b5
c.1	673.56	1	1 k mage	4			
c.2	669.53	1 psmk	1 k	4			
c.3	664.03	1 psmk	1 k mage	5	0.003	1.25	c1
					0.024	1.07	c2
c.4	661.37	1 psmk	1 k mage1 mage2	6	0.025	1.24 0.96	c2
c.5	663.93	1 psmk	1 k mage Dday	6	0.764	1.14	c3
c.6	660.32	1 psmk age	1 k mage	6	0.081	1.22	c3
					0.014	1.06	c7
c.7	666.76	1 psmk age	1 k	5			

Table 4. Results of regression analysis of North Carolina EPS data for Model C3 in Table 3. PSMK is the prenatal smoking exposure status, MAGE is the women's age of first menstrual period, AGE is her current age, DAY here indexes the day relative to ovulation (day 0), where $DAY = -5, \cdots, -1, 0$, and DDAY is the indicator of two day consecutive intercourse. The models for A and P are based on formulae in (1) and (4). s.e. is the standard error from the GEE analysis.

Model				p-values	
Components	parameter	estimate	s.e.	Wald test	GLR test
A					
	pmsk	-0.777	0.229	< 0.001	< 0.001
	age	-0.054	0.034	0.112	0.081
P					
	Day	0.553	0.290	0.057	0.023
	mage	0.520	0.317	0.101	0.014

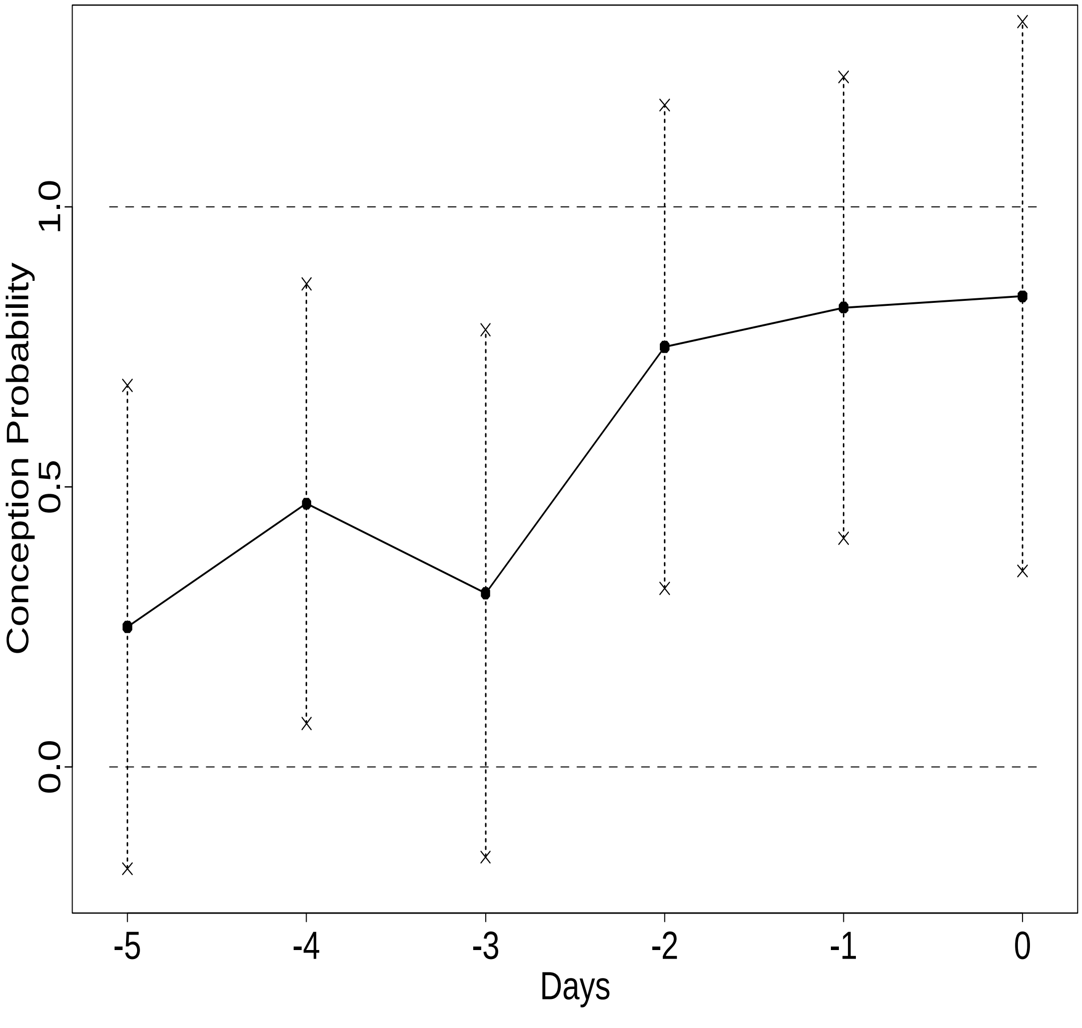

Figure 1: *Conditional conception probabilities (conditional on cycle viability) and their 95% profile likelihood confidence intervals based on Wald statistics for Model 4 in Table 2. Day 0 corresponds to the estimated day of ovulation.*

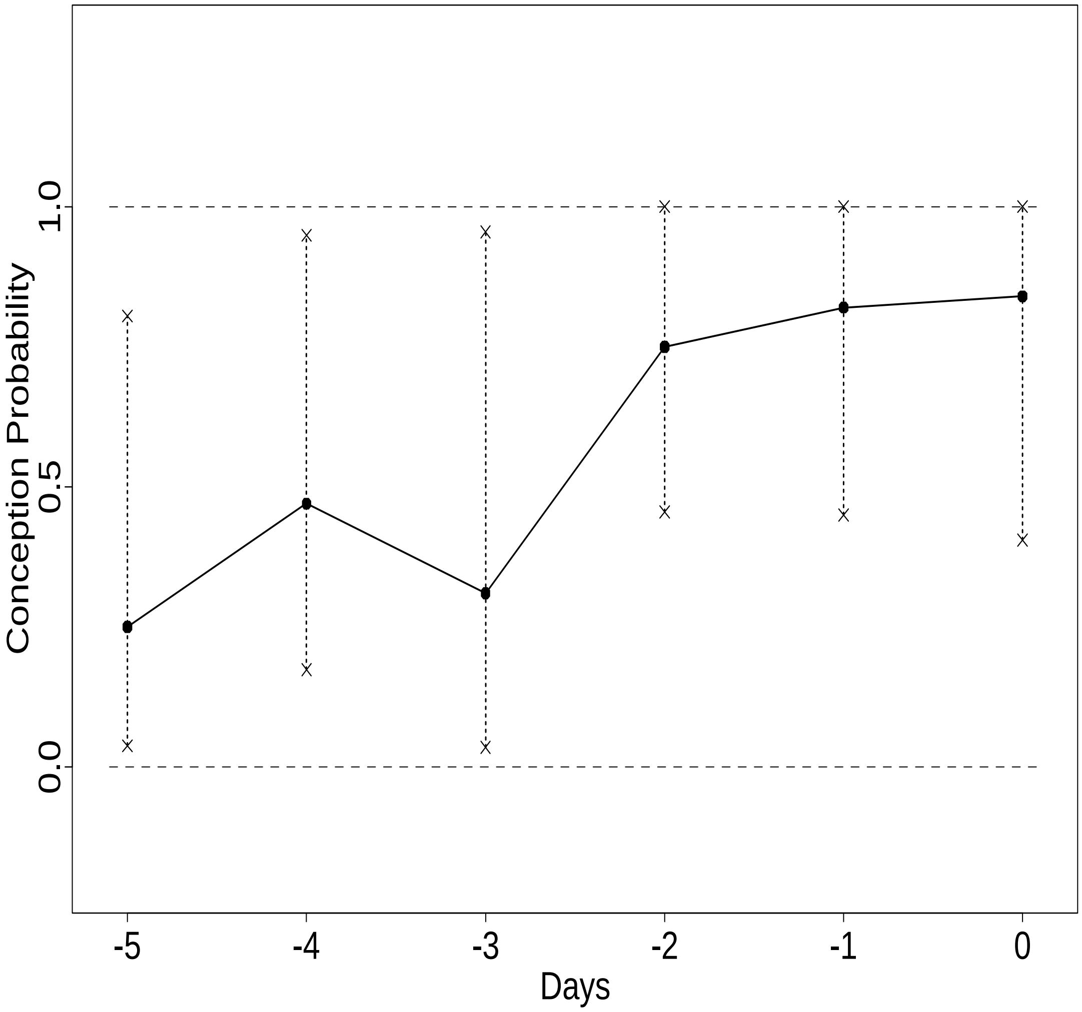

Figure 2: *Conditional conception probabilities (conditional on cycle viability) and their 95% profile likelihood confidence intervals based on GLR statistics for Model 4 in Table 2. Day 0 corresponds to the estimated day of ovulation.*