Bayesian Inference for Prevalence in Longitudinal Two-Phase Studies

Alaattin Erkanli,[1,*] **Refik Soyer,**[2]

and

Elizabeth J. Costello[1]

[1] Developmental Epidemiology Program, Department of Psychiatry and Behavioral Sciences,

Duke University Medical Center, Box 3454, Durham, North Carolina 27701, U.S.A.

[2] Management Science Department, School of Business and Public Management, George Washington University, Washington, DC 20052, U.S.A.

SUMMARY

We consider Bayesian inference and model selection for prevalence estimation using a longitudinal two-phase design where subjects initially receive a low-cost screening test followed by an expensive diagnostic test administered at several occasions. The change in the subject's diagnostic probability over time is described using four mixed-effects probit models where the subject-specific effects are captured by latent variables. The computations are performed using Markov Chain Monte Carlo methods. These models are then compared using the Deviance Information Criterion (DIC). The methodology is illustrated with an analysis of alcohol and drug use in adolescents using data from the Great Smoky Mountains Study (GSMS).

**Corresponding author's email address*: al@psych.mc.duke.edu

Key words: Bayesian inference; DIC; Gibbs sampling; Model selection; Prevalence estimation; Probit regression.

1. Introduction

The aim of multi-phase designs for estimating the prevalence of a rare disorder is to optimize precision, controlling for cost. For example, one might use a double-sampling design to screen the population using a moderately effective but relatively inexpensive instrument at the first phase (Deming, 1977), and then administer a more accurate, more expensive diagnostic test to a sub-sample of the screened subjects. In a longitudinal two-phase experiment, as defined here, the second (diagnostic) phase is repeated at different points in time. This work is motivated by the need for sophisticated models in analyzing complex data from the *Great Smoky Mountains Study* (GSMS). The GSMS (Costello et al., 1996) is a large-scale longitudinal study funded by the National Institute of Mental Health to understand the development and treatment of child psychiatric ilness at the individual, family, and service system levels. A double-sampling design was used because of the study's multiple goals. As pointed out by Pickles, Dunn and Vasquez-Barquero (1995), two-phase screening experiments closely resemble case-control studies, with the subjects with high screen scores as the "cases" and subjects with low screen scores as "controls". Usually, subjects at the diagnostic phases constitute a biased sample of the subjects selected at the screening phase. For example, in the GSMS all the subjects in the high screening group and only 10% of those in the low screening group were included in the diagnostic phases. Thus, the analyses must be adjusted for selection bias. Additional complexity arises when, as is common to all longitudinal studies, some subjects have missing observations over time.

The non-Bayesian literature on screening experiments and prevalence estimation has concentrated mostly on cross-sectional data (see Pickles et al., 1995, and the references therein) where weighted-likelihood methods are used to compansate for sampling bias. The resulting parameter estimates, however, are valid only asymptotically. On the other hand, by utilizing the complete likelihood based on both the screening and diagnostic data, Bayesian analyses have the natural ability of adjusting the posterior distribution of the prevalence for possible sampling bias (Gianola and Fernando, 1986).

Previous methodological developments in Bayesian inference for screening tests for cross-sectional experiments focus mostly on large sample analytical approximations to posterior densities; see, for example, Johnson and Gastwirth (1991). However, these approximations are complicated and can not in general provide satisfactory answers in longitudinal studies since the number of parameters is typically large. There is also a growing literature on simulation-based computational algorithms for Bayesian analyses of continuous and discrete longitudinal data; see, for example, Albert and Chib (1996), and Cargnoni, Mueller and West (1997) and the references therein. To the best of our knowledge, however, none of these focuses on two-phase screening experiments for prevalence estimation.

Recently, Lawrence, Theresa and Louis (1995), and Erkanli, Soyer and Stangl (1997) described Markov Chain-Monte Carlo methods for prevelance estimation using cross-sectional data. Neither, however, discussed prevalence estimation using longitudinal data. In this paper, Bayesian inferences for prevalence are developed using four different probit models for the diagnostic probabilities. The required computations are performed using Gibbs sampling (Gelfand and Smith, 1990). These models are then compared via the Deviance Information Criterion (DIC) recently introduced by Spiegalhalter et al. (1998).

In Section 2, the basic sampling structure of two-phase designs for screening experiments is described. In Section 3, probit models for the diagnostic probabilities are introduced and the prior distributions of the related parameters in each model are described. In Section 4, computational methods for posterior inference and model selection are discussed. In Section 5, the prevalence of substance use in adolescents is assessed by using data from the GSMS. Conclusions are in Section 6.

2. Two-Phase Designs

2.1 *Screening Phase*

The first (screening) phase of a two-phase, multi-wave screening experiment consists of selecting N subjects randomly from a large target population and administering a screening test to these subjects. These subjects are then classified into screen high, $S+$, and screen low groups, $S-$, according to their test scores. Consider the binary event $s_i = 1$ if the i-th subject belongs to $S+$ and $s_i = 0$ otherwise, for $i = 1, \ldots, N$. Let $n_+ = \sum_{i=1}^{N} s_i$ and $n_- = N - n_+$ be the sizes of $S+$ and $S-$ groups, respectively. Assuming independence of $\{s_i\}$ conditional on $\pi = P(s_i = 1) = P(S+)$, the sampling distribution of n_+ is binomial with parameters N and π and the likelihood for π is given by

$$l_s = l(\pi) = \pi^{n_+}(1-\pi)^{N-n_+}. \tag{1}$$

Here, the distribution of n_+ describes a simple screening process. Often, the screening phase is performed conditional upon other characteristics of the subjects. For example, in the Great Smoky Mountains Study, the screening interview was given after a post-stratification of the subjects into a cross-classification of sex (F, M) and three age cohorts (9, 11, and 13). Early work (Costello, et al., 1996) has shown, however, that the screening proportions were similar for these groups and, consequently, the likelihood in (1) was considered for future analyses. In general, the heterogeneous screening populations can be modeled using $\pi_i = P(s_i = 1|\mathbf{x}_i)$, for some covarite vector $\mathbf{x}_i$, and a link function, e.g., the probit link $\Phi^{-1}(\pi_i) = \mathbf{x_i}'\boldsymbol{\beta}$, to assess the covariate effects.

2.2 *Diagnostic Phase*

The second (diagnostic) phase consists of randomly selecting a subsample of the screen high and screen low groups and administering an in-depth diagnostic assessment several occasions. We define $n_1 = f_1 n_+$ and $n_0 = f_0 n_-$ where f_0 is the fraction of n_- subjects and f_1 is the fraction of n_+ subjects, $0 \leq f_k \leq 1$ for $k = 0, 1$, who are given the in-depth interview. We assume for practical reasons that n_0 and n_1 are integers. Let d_{itk}, for $i = 1, \ldots, n_k$, $t = 1, \ldots, T$ and $k = 0, 1$, denote the binary *score* of the diagnostic test for the i-th individual at occasion t, in group k. Thus, $d_{itk} = 1$ if the subject is diagnosed positively at time t, and $d_{itk} = 0$, otherwise. In GSMS, by the sampling design, only $n = n_0 + n_1 < N$ of the subjects are included in the second-phase. Since the sampling fractions f_k do not depend on the diagnostic outcomes, the sampling distribution for the remaining outcomes corresponding to the $m = N - n = n_+(1 - f_1) + n_-(1 - f_0)$ subjects, who are not included in the second-phase, are independent of the second-phase sampling mechanism. Furthermore, d_{it0} and $d_{j\tau 1}$ are statistically independent for all i, j, t and τ.

Here, the sampling fractions are assumed as fixed at the first-phase. It is possible, however, to consider other designs in which they vary over the waves. For example, the sampling fractions can be updated dynamically through the equations $n_0(t) = f_0(t) n_-(t-1)$ and $n_1(t) = f_1(t) n_+(t-1)$, for $t = 2, 3, \ldots, T$, where $n_0(t)$ and $n_1(t)$ are the number of subjects to be included in the t-th wave to receive the diagnostic test, and $n_-(t-1)$ and $n_+(t-1)$ are the number of subjects with the diagnosis at the wave $t-1$ in $S-$ and $S+$ groups, respectively. For $t = 1$, we have $n_-(0) = n_-$, $n_+(0) = n_+$, $n_0(1) = n_0$, and $n_1(1) = n_1$, as before.

3. Models for Diagnostic Probabilities

For prevalence estimation in cross-sectional screening experiments, (i.e., $T = 1$), Erkanli et al. (1997) have recently developed computationally efficient methods for prevalence estimation based on Dirichlet priors. Although these priors can be used here with some modifications, they are not very useful for modeling correlations between diagnostic variables in longitudinal studies. To describe a suitable model for the diagnostic process, let $E(d_{itk}) = h_{itk} = P(d_{itk} = 1|s_i = k)$, where the expectation is taken with respect to the distribution of d_{itk}, and consider the probit link $\Phi^{-1}(h_{itk}) = \mu_{tk} + \beta_{itk}$, where μ_{tk} is the unknown common mean for group k at time t, and β_{itk} is a latent variable describing the subject-specific variations around μ_{tk} at time t. In this paper, we consider the following models:

1. $\mu_{tk} = \mu_k + \beta_{tk}$, $\beta_{itk} = \alpha_{ik}$,
2. $\mu_{tk} = \mu_k + \beta_{tk}$, $\beta_{itk} = \alpha_{ik} + b_{itk}$,
3. $\mu_{tk} = \mu_k + \beta_{1k}t$, $\beta_{itk} = \alpha_{ik} + \gamma_{1ik}t$,
4. $\mu_{tk} = \mu_k + \beta_{1k}t + \beta_{2k}t^2$, $\beta_{itk} = \alpha_{ik} + \gamma_{1ik}t + \gamma_{2ik}t^2$.

Model 1 is similar to a one-way analysis of variance model for the probits $\Phi^{-1}(h_{itk})$ without a subject-and-time interaction term. A single subject-specific variable, α_{ik}, is used to account for within-subject correlations. Model 1 fits a population mean $\mu_k + \beta_{tk}$ to the probits $\Phi^{-1}(h_{itk})$ with α_{ik} describing individual variations around it. The parameter β_{tk} describes the variations over time in the population around the common mean μ_k. In the language of generalized linear models, μ_k and β_{tk} are referred as fixed-effect parameters, and α_{ik} as a random-effect parameter. As a convenient set of priors for model 1 we consider non-informative but proper normal priors for μ_k and

β_{tk}, and $\alpha_{ik}|\sigma^2_{\alpha k} \overset{\text{i.i.d}}{\sim} N(0, \sigma^2_{\alpha k})$, with $(\sigma^2_{\alpha k})^{-1} \sim \text{Gamma}(a_1, b_1)$, where a_1 and b_1 are known.

Model 2 is a simple extension of model 1 where $\mu_k, \{\beta_{tk}\}$ and $\{\alpha_{ik}\}$ are as in model 1, and b_{itk} is a subject-and-time interaction term. As in model 1, $\mu_k + \beta_{tk}$ is the population mean, but now the individual variations, $\alpha_{ik} + b_{itk}$, are functions of time. Here, $b_{itk}|\sigma^2_{bk} \overset{\text{ind.}}{\sim} N(0, \sigma^2_{bk})$, with $(\sigma^2_{bk})^{-1} \sim \text{Gamma}(a_2, b_2)$, where a_2 and b_2 are known.

Model 3 describes a linear growth curve $\mu_k + \beta_{1k}t$ for the probits $\Phi^{-1}(h_{itk})$, with $\alpha_{ik} + \gamma_{1ik}t$ denoting individual variations around it. Here, μ_k and α_{ik} are as in model 1, β_{1k} has a non-informative but proper normal prior, and $\gamma_{1ik}|\sigma^2_{\gamma_1 k} \overset{\text{i.i.d}}{\sim} N(0, \sigma^2_{\gamma_1 k})$, with $(\sigma^2_{\gamma_1 k})^{-1} \sim \text{Gamma}(a_3, b_3)$, where a_3 and b_3 are known.

Model 4 is a quadratic extension of model 3 where $\mu_k + \beta_{1k}t + \beta_{2k}t^2$ is the population growth curve and $\alpha_{ik} + \gamma_{1ik}t + \gamma_{2ik}t^2$ describes the individual variations around it. Here, β_{2k} has a non-informative normal prior, and $\gamma_{2ik}|\sigma^2_{\gamma_2 k} \overset{\text{i.i.d}}{\sim} N(0, \sigma^2_{\gamma_2 k})$ with $(\sigma^2_{\gamma_2 k})^{-1} \sim \text{Gamma}(a_4, b_4)$, where a_4 and b_4 are known.

We describe the details of specifying the known hyperparameters of these priors for the substance use example in Section 5. Let $\boldsymbol{\theta}_k$ denote the components of μ_{tk} and β_{itk}, and $\boldsymbol{\psi}_k$ denote the unknown hyperparameters of the distribution of $\boldsymbol{\theta}_k$. The variance components in $\boldsymbol{\psi}_k$, and conditional on $\boldsymbol{\psi}_k$ the components in $\boldsymbol{\theta}_k$, are assumed to be mutually independent. Moreover, conditional on $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$, π and the vectors $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are mutually independent, with the prior $\pi \sim U(0,1) = \text{Beta}(1,1)$. Finally, we assume that $d_{itk}|\boldsymbol{\theta}_k, \boldsymbol{\psi}_k, \pi \overset{\text{ind.}}{\sim} \text{Bernoulli}(h_{itk})$ for all $i = 1, \ldots, n_k$ and $t = 1, \ldots, T$; $k = 0, 1$.

The prevalence of diagnosis is described as a mixture of the prevalences in each of the $S+$ and $S-$ groups as follows. Let a potential subject with the diagnosis in the population be indexed by j. Denote the subject's potential screening score as s_j and diagnostic score at time t as d_{jt}, respectively. Define the prevalence as the probability of the subject's diagnosis at time t as $p_{jt} = P(d_{jt} = 1) = E(d_{jt})$. By conditioning on s_j and using the law of total probability, we have $p_{jt} = h_{jt0}(1-\pi) + h_{jt1}\pi$. Then, by averaging p_{jt} with respect to the prior distributions of β_{jt0} and β_{jt1}, we obtain the overall population prevalence p_t at time t as $p_t = \lambda_{t0}(1-\pi) + \lambda_{t1}\pi$, where $\lambda_{tk} = E_{\beta_{jtk}}(h_{jtk}) = \Phi\left[\mu_{tk}/\{1 + \mathrm{Var}(\beta_{jtk})\}^{1/2}\right]$. For instance, in model 1, $\mathrm{Var}(\beta_{jtk}) = \sigma^2_{\alpha k}$, and in model 4, $\mathrm{Var}(\beta_{jtk}) = \sigma^2_{\alpha k} + t^2\sigma^2_{\gamma_1 k} + t^4\sigma^2_{\gamma_2 k}$. If a heterogeneous screening population is considered, we can replace π in the definition of p_{jt} by $\pi_j = P(s_j = 1|\mathbf{x}_j)$. The overall prevalence is then a function of both the covariate vector $\mathbf{x}_j$ and the parameter vector $\boldsymbol{\beta}$.

The *sensitivity* and *specificity* of a screening test, i.e., $\eta_{t1} = P(s = 1|d_t = 1)$ and $\eta_{t0} = P(s = 0|d_t = 0)$, are obtained, using the diagnostic information and Bayes' Theorem, as $\eta_{t1} = P(d_t = 1|s = 1)P(s = 1)/p_t = \lambda_{t1}\pi/p_t$ and $\eta_{t0} = P(d_t = 0|s = 0)P(s = 0)/(1 - p_t) = (1 - \lambda_{t0})(1 - \pi)/(1 - p_t)$.

The computations of the posterior distributions of π, and $\boldsymbol{\theta}_k$ and $\boldsymbol{\psi}_k$, for $k = 0, 1$, and related quantities are discussed in the next section.

4. Posterior Inference

The likelihood for $\boldsymbol{\theta}_k$ from the diagnostic phase is $l(\boldsymbol{\theta}_k) = \prod_{i=1}^{n_k}\prod_{t=1}^{T} h_{itk}^{d_{itk}}(1-h_{itk})^{1-d_{itk}}$, where h_{itk} are defined as in Section 3. Denoting the likelihood for the diagnostic phase conditional on the screening status as $l_{d|s}$, the overall likelihood is obtained as

$$l(\pi, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1) = l_s l_{d|s} = \pi^{n_+}(1-\pi)^{N-n_+}\left\{\prod_{k=0}^{1} l(\boldsymbol{\theta}_k)\right\}. \tag{2}$$

If a heterogeneous screening population is considered, then the screening contribution in (2) can be easily modified as $l_s = \{\prod_{i=1}^{n_+}\pi_i\}\{\prod_{i=1}^{n_-}(1-\pi_i)\}$. In Equation 2, we interpret the screening contribution l_s as a weight function to offset the sampling bias in prevalence rates, e.g., when the subjects with high screen scores are oversampled; $f_1 = 1$ and $f_0 < 1$. Because the likelihood of π and $\boldsymbol{\theta}_k$, $k = 0, 1$, factors, and the priors are independent, the marginal posterior distribution of π is a Beta$(n_+ + 1, N - n_+ + 1)$. The joint posterior distribution of $\boldsymbol{\theta}_k$ and $\boldsymbol{\psi}_k$, conditional on the data $\mathbf{d} = \{d_{itk}\}$, for $k = 0, 1$, is proportional to

$$\prod_{k=0}^{1} l(\boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k | \boldsymbol{\psi}_k) p(\boldsymbol{\psi}_k), \tag{3}$$

where $p(\cdot)$ denotes the prior densities. Although we can not evaluate (3) directly, we can simulate samples from it using Markov Chain Monte Carlo methods. By utilizing the product form of the likelihood and the independence of the priors described above, we can construct a Gibbs sampler that simulates successively from the conditionals

$$p(\boldsymbol{\theta}_k | \mathbf{d}, \boldsymbol{\psi}_k) \propto l(\boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k | \boldsymbol{\psi}_k), \tag{4}$$

and

$$p(\boldsymbol{\psi}_k | \mathbf{d}, \boldsymbol{\theta}_k) \propto p(\boldsymbol{\theta}_k | \boldsymbol{\psi}_k) p(\boldsymbol{\psi}_k). \tag{5}$$

The conditional distribution in (4) does not have a closed analytical form; however, it can be shown to be log-concave. Thus, the adaptive rejection sampling method (Gilks and Wild, 1992) or, alternatively, the Metropolis-Hastings algorithm (Carlin and Louis, 1996, page 173) can be used for simulation from (4). The conditional distribution in (5) is available in closed analytical form as a product of (independent) inverse gamma distributions.

So far we assumed that each subject has exactly T complete repeated measurements. Under the assumption of ignorable missingness, missing observations can be simulated using the conditional predictive distribution Bernoulli(h_{itk}) in the k-th group. Augmenting the conditional distributions in (4) and (5) with this distribution provides sampling based approximations for the population parameters and missing observations. If the missing mechanism is known to be informative, then the methods described in this paper can not be used directly. We refer the interested reader to Cowles et al. (1996), and more recently, to Ten Have, et al. (1998), for mixed-effect tobit and logistic models for handling informative drop-out processes.

4.1 *Posterior Inference for Prevalence, Sensitivity and Specificity*

Having simulated Monte Carlo sequences $\{\boldsymbol{\theta}_k^{(r)}\}_{r=1}^R$ and $\{\boldsymbol{\psi}_k^{(r)}\}_{r=1}^R$ from the joint posterior distribution in (3), and an additional Monte Carlo sequence $\{\pi^{(r)}\}_{r=1}^R$ generated independently from Beta$(n_+{+}1, N{-}n_+{+}1)$, the posterior distributions of λ_{tk} and p_t can now be approximated by direct substitution using the corresponding formulas. For example, $p_t^{(r)} = \lambda_{t0}^{(r)}(1{-}\pi^{(r)}){+}\lambda_{t1}^{(r)}\pi^{(r)}$. The posterior distributions of sensitivity and specificity are also similarly approximated by using the sequences $\{\eta_{1t}^{(r)}\} = \{\lambda_{t1}^{(r)}\pi^{(r)}/p_t^{(r)}\}$ and $\{\eta_{t0}^{(r)}\} = \{(1-\lambda_{t0}^{(r)})(1-\pi^{(r)})/(1-p_t^{(r)})\}$.

4.2 *Model Selection*

Since we do not believe any of the models 1-4 to be the "true" model, our approach is to compare these models in terms of a global criterion. Here, we use the *Deviance Information Criterion* (DIC), recently described in Speilgelhalter et al. (1998). For a generic parameter vector $\boldsymbol{\theta}$, which may or may not include random-effects and latent variables, DIC is defined as $\overline{D} + p_D$, where $D = -2\log l(\boldsymbol{\theta})$, $\overline{D} = E_{\boldsymbol{\theta}|\mathbf{d}}[D]$, and $p_D = \overline{D} - D(\widehat{\boldsymbol{\theta}})$, with $\widehat{\boldsymbol{\theta}}$ denoting the posterior mean. Here, the quality of a fit is measured by $\overline{D}$ and complexity is measured by the *effective* number of parameters, p_D. In fact, using asymptotic arguments, Speilgelhalter et al. (1998) show that for fixed-effects models DIC $\approx$ AIC with $p_D \approx p = \dim(\boldsymbol{\theta})$. For the mixed-effects models 1-4 above, $\overline{D}$

is the posterior mean of $D(\pi, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1) = -2(\log l_s + \log l_{d|s})$, which is estimated by the sampled values for D. The number p_D measures the contribution of the latent variables to the total number of parameters. Also, the $-2\log l_s$ term can be dropped from DIC, since the screening contribution l_s to overall likelihood and the prior distribution of π is the same for each model.

5. Drug and Alcohol Use in Adolescents

In this section, we investigate the progression of substance use over time in young adolescents enrolled in the *Great Smoky Mountains Study* (GSMS), using the Bayesian models introduced in Section 3. The GSMS is an epidemiologic study of young adolescents living in rural Southern Appalachia (Costello et al., 1996). Briefly, a representative sample of $N = 4500$ youths who were 9, 11, and 13 years old at baseline were selected from eleven counties in western North Carolina using a household equal probability design. A screening questionnaire was administered to a parent, by telephone or in person. All children ($n_+ = 1125$) who scored above a predetermined cut point on the behavioral and substance use questions, plus a 1-in-10 random sample of the rest, were recruited for the longitudinal study. Among the 4500 subjects who received the screening test, a final total of $n_0 + n_1 = 1073$ respondents were enrolled in the longitudinal study. Between 77% and 95% of the subjects completed interviews at each wave for the first four waves of annual data collection. For each child enrolled in the longitudinal study, a parent and the child were interviewed annually. Previous cross sectional analyses based on demographics characteristics such as age, sex, and income and the screening data (Costello at al., 1996) show no significant differences among the families who agreed to participate but could not be interviewed at wave 1 and the families who were selected for the diagnostic sample but refused to participate. Of the 1073 families who completed at least one interview, 7% completed only one, 7% completed 2, 16% completed 3, and 70% completed all 4. There is no evidence that the noncompletion rate is informative. We therefore assumed ignorable missingness throughout.

The measures used for psychiatric diagnoses and substance use were based on the *Child and Adolescent Psychiatric Assessment* (CAPA) (Angold et al., 1995), an interview that elicits information about symptoms that contribute to a wide range of diagnoses. For the purpose of this paper, we used a binary diagnostic variable that is obtained using a computer algorithm to implement the diagnostic criteria. Table 1 gives the summary statistics of substance use at each sampling wave (1–4) for each screening group, the number of subjects interviewed, and the MLEs of p_t at each wave.

Table 1 here

5.1 *Computations*

The posterior distribution of the proportion π of screening is available in closed form as a Beta$(1010, 2888)$ distribution with the mean 0.25 and standard deviation of 0.007, respectively. The posterior computations for the other unkown quantities were performed using the BUGS (Spiegelhalter et al., 1996) software. We used 1000 iterations for burn-in period and 5000 for convergence. Starting values needed for the Gibbs sampling are obtained via simulation from the prior distributions. In all the models, we assumed the prior $N(0, 10)$ for the regression parameters in μ_{tk}, and the hyperparameters $a_1 = a_2 = a_3 = a_4 = 1$, $b_1 = b_2 = b_3 = b_4 = 0.1$ for the precision parameters in $\boldsymbol{\psi}_k$, $k = 0, 1$. These choices induce relatively flat priors for the regression coefficients and precisions of the latent variables, but not too flat to create any numerical problems for the Gibbs sampler. Also, to achieve good convergence properties, we reparameterized each model using the corner constraints (Clayton, 1996). To assess convergence, we monitored (for each model) the realizations of the parameters λ_{tk}, p_t, and η_{tk} using five different starting points; all seemed to settle after 1000-1200 initial iterations. The results here are based on a single run of the Gibbs sampler. We then computed the posterior summaries for the model parameters and DIC for each model.

5.2 *Comparison of the models*

The posterior means for the overall prevalence of substance use, and the values of DIC for each model are displayed in Table 2. While all the models give posterior estimates that are close to the MLE's of prevalence of substance use (see Table 1 above), the posterior estimates produced by models 3 and 4 are somewhat *smoother* than those obtained using models 1 and 2. This can be attributed to the smoothing effect of the linear and quadratic growth curve models for the probits. In terms of the DIC, model 2 has clearly performed better, although model 4 (quadratic model) is not too far away. Model 1 has the least complexity, but it does not fit as good as the other models. The second least complex model is the third model, i.e., the linear growth curve model, which fits slightly better than the model 1, but worse than the models 2 and 4.

Table 2 here

The posterior distributions of the prevalence of substance use conditional on the screening status, and specificity and sensitivity of the screening test under model 2 are displayed in Figure 1. The box plots in Figure 2 show the posterior distribution of the overall prevalence of substance use at each wave (the posterior standard deviations for p_t under model 2 are 0.008, 0.009, 0.015, and 0.015, respectively, at the waves 1, 2, 3, and 4). As expected, the prevalence estimates rose in the later waves of data, reflecting the marked rise in substance use as the children in the study grew into adolescence. As evidenced in Figures 1 and 2, the posterior means of p_t are pulled towards the posterior distribution of the conditional prevalence λ_{t0} at all waves, suggesting that the Bayesian model accounts for the sampling bias. We also observe that, although the prevalence estimates increase over time, the sensitivity and specificity estimates of the screening test exhibit no such behavior, suggesting stability of screening test over time in predicting the diagnosis.

Figures 1 and 2 here

6. Conclusions

We demonstrate in this paper that it is possible to use even complex multi-sampling longitudinal designs as the basis for reasonably reliable prevalence estimates. This is important for many fields of research where case-finding is expensive; for example, in medical research into rare disorders. Moreover, the Bayesian models used here provide a basis for future analyses of multi-phase, longitudinal designs for prevalence estimation.

ACKNOWLEDGMENTS

We would like thank to Dr. Dalene Stangl of Institute of Statistics and Decision Sciences, Duke University, and Dr. Adrian Angold of Department of Psychiatry and Behavioral Sciences, Duke University Medical Center for helpful discussions. We are especially grateful to editor Professor R. J. Carroll for his suggestion to compare alternative models, and the associated editor and two referees for their valuable inputs, which together led to substantial improvement of the paper.

REFERENCES

Albert, J. and Chib, S. (1996). Bayesian modeling of binary repeated measures data with application to crossover trials. In *Bayesian Biostatistics*, Berry, D.A. and Stangl, D. K. (eds), 577-599. New York: Marcel and Decker.

Angold, A., Anthony, C., Prendergast, M., Rutter, M., and Siminoff, E. (1992). *Child and Adolescents Psychiatric Assessment (CAPA)* . Developmental Epidemiology Program, Department of Psychiatry, Duke University.

Cargnoni, C., Mueller, P., and West, M. (1997). Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models. *Journal of the American Statistical Association* **92**, 640-647.

Carlin, B. P. and Louis, T. A. (1996). *Bayes and Empirical Bayes Methods For Data Analysis*. London: Chapman and Hall.

Clayton D. G., (1996). Generalized linear mixed models. In: *Markov Chain Monte Carlo in Practice*, Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds), 275-301. London: Chapman and Hall.

Costello, E. J., Angold, A., Burns, B., Stangl, D., Tweed, D., Erkanli, A., and Worthman, C. (1996). The Great Smoky Mountains Study of Youth: Prevalence and correlates of DSM-III-R disorders. *Archives of General Psychiatry*, 53, 1129-1136.

Cowles, M. K, Carlin, B. P. and Connett, J.E. (1996). Bayesian tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness. *Journal of the American Statistical Association* **91**, 68-98.

Deming, W. E. (1977). An essay on screening, or two-phase sampling, applied to surveys of a community. *International Statistical Review* **45**, 29-37.

Erkanli, A., Soyer, R., and Stangl, D. (1997). Bayesian inference in two-phase prevalence studies. *Statistics in Medicine* **16**, 1121-1133.

Gelfand, A. E., and Smith, A. M. F. (1990). Sampling-based approaches to calculating marginal densities.*Journal of the American Statistical Association* **85**, 398-409.

Gilks, W. R., and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**, 337-348.

Gianola, D. and Fernando, R. L. (1986). Bayesian methods in animal breeding theory. *Journal of Animal Science* **63**, 217-244.

Johnson, W. O., and Gastwirth, J. L. (1991). Bayesian inference for medical screening tests: approximations useful for the analysis of Acquired Immune Deficiency Syndrome. *Journal of the Royal Statistical Society, Series B* **53**, 427-439.

Lawrence, J., Theresa, G. W., and Louis, C. (1995). Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology* **141**, 263-272.

Pickles, A., Dunn, G., and Vasquez-Barquero, J. L. (1995). Screening for stratification in two-phase epidemiological surveys. *Statistical Methods in Medical Research* **4**, 73-89.

Spiegelhalter, D., Thomas, D., Best, N., and Gilks, W. R. (1996). *BUGS Manual 0.5.* Cambridge: MRC Biostatistics Unit.

Spiegelhalter, D., Thomas, D., Best, N., and Carlin, B. P. (1998). Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models. *Journal of Royal Statistical Society, Series B.* In press.

Ten Have, R. T., Kunselman, A. R., Pulkstenis, E. P., and Landis J. R. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics* **54**, 367-383.

Table 1: Number of cases with substance use in $S-$ and $S+$. The numbers in the fourth and fifth columns are the number of subjects interviewed at each ave. The MLE's of prevalence (standard errors) are displayed in the last column.

Waves	Cases in $S-$	Cases in $S+$	$n_0 = 270$	$n_1 = 803$	$\widehat{p}_t$
1	8	38	257	758	0.036 (0.009)
2	7	63	244	726	0.044 (0.009)
3	22	76	241	688	0.097 (0.015)
4	15	86	216	604	0.088 (0.015)
Total	52	263	958	2776	0.065 (0.006)

Table 2: Posterior means for prevalence and DIC for each model.

Waves	Model 1	Model 2	Model 3	Model 4
1	0.038	0.039	0.037	0.036
2	0.049	0.050	0.059	0.066
3	0.104	0.106	0.088	0.097
4	0.105	0.106	0.117	0.116
$\overline{D}$	1045.2	745.7	894.3	752.0
p_D	248.15	325.13	269.88	331.96
DIC	1293.35	1070.83	1164.18	1083.96

Figure 1: Posterior distributions for conditional prevalences, sensitivity and specificity under Model 2.

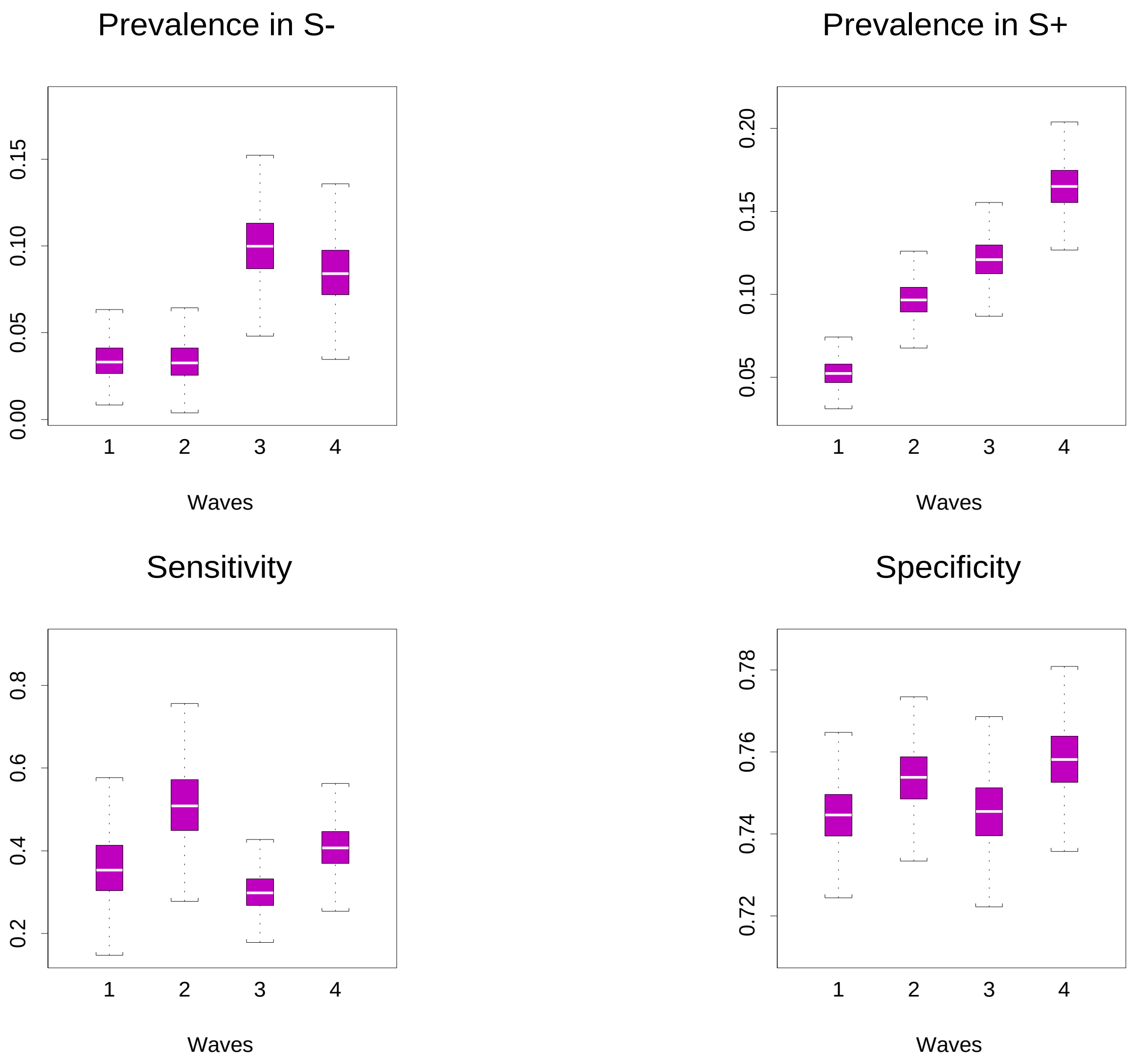

Figure 2: Posterior distributions of prevalence of substance use under Model 2.