Age at First Marriage in Malawi: a Bayesian multilevel analysis using a discrete-time model

Samuel O. Manda and Renate Meyer

Department of Statistics, University of Auckland, Auckland, New Zealand

Summary. This paper presents a hierarchical discrete-time survival model for the analysis of data on timing of marriage. The data were collected as part of the Knowledge, Attitudes, and Practices in Health (MKAPH) survey for Malawi held in 1996. The model assesses the determinants of transition to marriage and accounts for family and community random effects with cross-level correlation structure. A nonparametric technique is used to dynamically model the baseline discrete hazard. The model produces improved estimates of fixed effects and the associated standard errors, in addition to providing intracluster and intercluster correlations at the levels of the hierarchy. A full Bayesian approach is employed using Markov Chain Monte Carlo for posterior computation.

Keywords: Gibbs sampling; Marital reproductive behaviour; Survival analysis; Dynamic model; Random effects

1. Introduction

In reproductive behaviour, an increase in age at first marriage is a key feature of fertility transition in any country. In Malawi, median age at first marriage among women is 18 years and the total fertility rate (TFR), defined as the average number of births a woman would have if she survived to age 50, is 6.7 (Government of Malawi (1992)). In 1977 and 1987, the TFRs were 7.6 and 7.4 respectively, and these figures are among the highest in sub-Sahara Africa. Thus, Malawi could be considered to be in pre-fertility transition stage. The Malawi Government has been devising policies regarding delaying entry into marriage and childbearing through improved family planning and health care, increased school intakes particularly of female students, and more employment opportunities. The data available for this study come from the 1996 Malawi Knowledge, Attitudes, and Practices in Health (MKAPH) survey. The survey used a hierarchically clustered sampling design - in particular at the family and community clustering levels. This paper examines the socioeconomic and demographic factors influencing timing of first marriage in Malawi and, in particular, it assesses the extent of association among ages at marriage within and across families and communities.

Early marriage, in the absence of contraception, leads to an early childbearing age which often endangers the health of the woman and her offspring and thus causes high child and maternal mortality (Acsadi and Johnson-Acsadi (1990)). Also, early marriage reduces education and employment opportunities available to women, hindering them from contributing effectively to the social and economic development of a country. Education is the prime factor determining entry into marriage as an increase in educational attainment positively correlates with an increased age at first marriage (Caldwell (1976); Caldwell and Caldwell (1997)). Although improvements have been achieved, levels of women's education in the rural areas of the sub-Sahara region remain low (Kirk and Pillet (1998)). Thus, the status of education among rural women in the region indicates that the environment is not conducive to any significant reduction in age at first marriage. Birth cohort differentials in age at first marriage provide an indication of trends over time. However, trend analysis is hampered by recall errors among older women of events that happened many years ago (Kirk and Pillet (1998); Singh (1998)). Thus, only trends in recent years are worth examining. Many factors affecting timing of marriage are strongly interrelated such that their effects can not be considered in isolation. Indeed, as noted by Kavinya (2001), most secondary certificate holders migrate to urban centres in search for waged employment and the most educated women are in the young birth cohorts.

As elaborated in the following paragraphs, an appropriate model for this application needs to take account of the discrete measurement of time to marriage. The model should also measure the influence of the covariate factors as well as allowing for heterogeneity among families and communities. Possible cross-level association between families and communities need to be investigated. Finally, rather than restraining the underlying baseline hazard to a certain functional form, we want to allow for flexible modelling of the baseline hazard function.

In demographic studies, the proportional hazards regression is used widely to model data on event histories such as age at marriage (Santow and Bracher (1994); Brien, Lillard and Waite (1999)). The proportional hazards model is based on the assumption that time duration is measured on a continuous scale. However, in demography, time is often measured as a discrete variable, typically in months or years resulting in a large number of tied time points. Estimation of the model with such heavily tied failure times would generally give inconsistent estimators (Prentice and Gloeckler (1978)). Breaking ties by slight adjustments does not solve the problem as this introduces accuracy bias into the data. In these situations, discrete time duration models are more appropriate. Discrete time survival models treat the indicators of death at each time point as independent Bernoulli trials, and use a series of binary regression models at each time point (Allison (1982); Arjas and Haar (1987);Fahrmeir and Knorr-Held (1997)). The discrete-time duration models have been extended to account for unobserved heterogeneity or frailty that is not explained by the observed and measured variables (Scheike and Jensen (1997); Xue and Brookmeyer (1997); Lewis and Raftery (1999)). But their analyses included only a random effect at the individual level to capture unobserved heterogeneity.Biggeri, Bini and Grilli (2001) used penalised quasi-likelihood to estimate a discrete time survival model with two independent nested random effects for the analysis of time to obtain a first job among graduates in Italy.

The inclusion of cluster-specific random effects to control for the unobserved heterogeneity in clustered discrete time data is of substantive and statistical interest. Modelling the differences between families and communities with respect to age at first marriage provides an understanding of the consequences of unmeasured social, behavioural, and environmental factors that are related to marriage. An analysis of clusters exhibiting extreme random effects may provide clues to high risk factors which were previously unmeasured. From a methodological view point, clustering induces correlation in the data at the family and community levels. Thus, it is inappropriate to use procedures that assume independence among observations. Analyses that neglect clustering tend to underestimate the variance of the fixed effect parameter estimates, and can lead to an inflation of the significance of the effects. Moreover, in non-linear models, neglecting clustering may result in biased and inconsistent estimates of parameters. We also want to investigate cross-level correlation. We want to see whether the effects are amplified when, for example, a family which adheres to traditional customs rigidly resides in a very traditional community. If this is the case, then neglecting cross-level correlation may be wrong. Similarly, a 'poor' family may have much more advantage than, say, a mid-level social family as a result of living in a community with late marriage entries. This could happen, if for example, there was a program in an urban city to target high-risk families with an intervention to improve reproductive health. An appropriate model that combines both within and cross level correlation of the random effects is the bivariate normal model. It has previously been used to model correlation between random effects at the same level (Gustafson (1997); Cook and Ng (1997)).

In many applications, the baseline hazard is modelled as piecewise constants which has limited flexibility. An ideal situation is to model both baseline hazards and some covariate effects as functions of time. This may involve using parametric functional forms such as polynomials for baseline hazards (Biggeri, Bini and Grilli (2001)) and time-varying effects. But this may not capture hidden patterns such as peaks and bumps. In this paper, the baseline hazard is modelled dynamically using a smoothing mechanism which can show inherent patterns (Fahrmeir and Knorr-Held (1997)). In using cross-level dependence between family and community frailties, a nonparametric baseline hazard and a discrete-time duration model, this paper is a significant generalisation of the analyses in Sastry (1997) and Bolstad and Manda (2001)). We perform full Bayesian inference using Markov chain Monte Carlo (MCMC) simulation techniques. Fahrmeir and Knorr-Held used MCMC methods to analyse a discrete-time duration model to employment data from Germany without explicitly controlling for random effects. Lewis and Raftery (1999) also used MCMC methodology for estimating a discrete-time event history model with heterogeneity in the study of fertility decline in Iran.

2. The Data

The data available for analysis in this study were collected in 1996 as part of the Malawi Knowledge, Attitudes, and Practices (MKAPH) survey conducted by the National Statistical Office (NSO). The survey provided information on malaria prevention, family planning, marriage, partner relations, and sexually transmitted diseases, including AIDS. The MKAPH survey used six sampling strata: urban and rural domains for each of the three administrative regions of the country, Northern, Central and Southern regions. The survey employed a two-stage sampling design. In the first stage, a systematic stratified sample of 106 enumeration areas (EAs) was selected. In the second stage, a systematic sample of families was selected within each selected EA. Both the rural and northern regions were oversampled to ensure that reliable estimates in these areas were calculated, in addition to ensuring the sampling errors within them were kept to a minimum.

In the sampled families, all women aged 15-49 years were eligible for the survey. A total of 2,683 women were interviewed from 2008 families distributed over 106 EAs. The EAs are census tracts which correspond to villages, and these provide community units. The number of women per community ranged from 6 to 57 with mean and medium numbers 22 and 24 respectively. For the MKAPH survey, a family was defined as one or more persons related or unrelated living together and sharing food resources. We take this sampling unit as a family unit. Each woman respondent was asked about all her current and past relations in and outside marriage. Of these women, 2172 (eighty-one percent) had been or are currently married with mean and median age at first marriage estimated at 17.8 and 17.0 years respectively. Fig. 1 shows the estimated empirical discrete hazard function, which clearly has a quadratic trend with high marriage rates in ages between 16 and 21

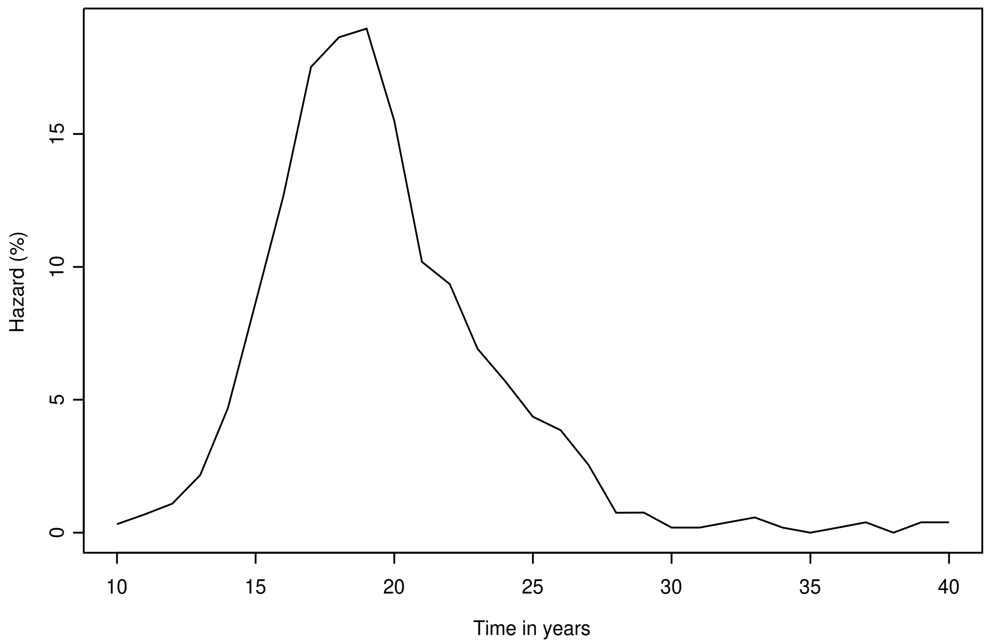

Fig. 1. Empirical hazard function in percentage

years.

We considered four covariates for inclusion in the analysis. As the boundaries of the three regions largely define ethnic lines, region of residence was used in the analysis to study ethnic differentials. Type of place of residence (whether urban or rural) was also included even though it is strongly correlated with education. The effect of education was indicated by the following categories: No education, primary and secondary or higher. In order to capture the effect of changing times on age at first marriage, we included a birth cohort of women in the analysis. The birth cohort categories were 1947-1951, 1952-1956, 1957-1961, 1962-1966, 1967-1971, 1972-1976 and 1977-1981. Table 1 shows the distribution of the covariate variables, with corresponding coefficients. In comparisons with census figures, the urban areas and the northern region were clearly oversampled. However, the results presented in this study were not weighted. A discussion of analysing weighted data from demographic and health surveys can be found in Korn and Graubard (1999)). In order to investigate early marriage rates by birth cohorts, proportions of birth cohorts married before various exact ages were calculated and are shown in Fig. 2. If we only consider births after 1957, then there appears to be a decline in early marriage rates. The proportion of women married by age 15 has declined from 25 percent among women born between 1957 and 1961 to 10 percent among women born in 1977-81. It is also apparent that older women tend to displace their age at first marriage closer to the survey date as at least 11% of women born before 1955 reported age at first marriage to be 22 years or older, which is significantly

Table 1. Summary of covariate variables used in the model

Variable	Frequency	Percent	Coefficient
Year of birth			
1977-81	626	23.3	0.0
1972-76	548	20.4	β_1
1967-71	435	16.2	β_2
1962-66	388	14.5	β_3
1957-61	260	9.7	β_4
1952-56	262	9.8	β_5
1947-51	164	6.1	β_6
Residence			
Urban	1333	49.7	0.0
Rural	1350	50.3	β_7
Region			
Northern	622	23.2	0.0
Central	1038	38.7	β_8
Southern	1023	38.1	β_9
Education level woman			
No education	733	27.3	β_{10}
Primary	1571	58.6	β_{11}
Secondary+	379	14.1	0.0
Total	2683	100.0	

older than those in sub-Sahara Africa. This clearly inflates the median age at first marriage (Government of Malawi (1992)).

We are interested in measuring the degree of association among women in the same family and community that is still present after controlling for the observed covariates. The association is a result of unobserved factors. For example, at the family level, the unobservables may include poverty and social status of the family which may influence entry into marriage. At the community level, the unobservables may include the amount of bride wealth, approved age at marriage and sanctions against premarital childbirth that would have the effect of increasing or decreasing the time to first birth or marriage.

3. The Model

In discrete time survival analysis, the time scale is partitioned into m intervals $\{I_t = [a_{t-1}, a_t)\}$ with $0 = a_0 < a_1 < \cdots a_m < \infty$). We assume that all censoring happens at the end of the intervals, and the intervals need not necessarily to be of equal length. We observe the discrete time $T = t$ with t in $\{1, \ldots, m\}$ denoting an observed event in interval I_t. Now consider the kth individual in family j in community i with $x_{ijk(t)}$, a vector of covariates at time t ($i = 1, \ldots, N$; $j = 1, \ldots, J_i$; $k = 1, \ldots, K_{ij}$). Conditional on the unobserved covariates or log-*frailties* b_{ij} and b_i for family ij and community i respectively,

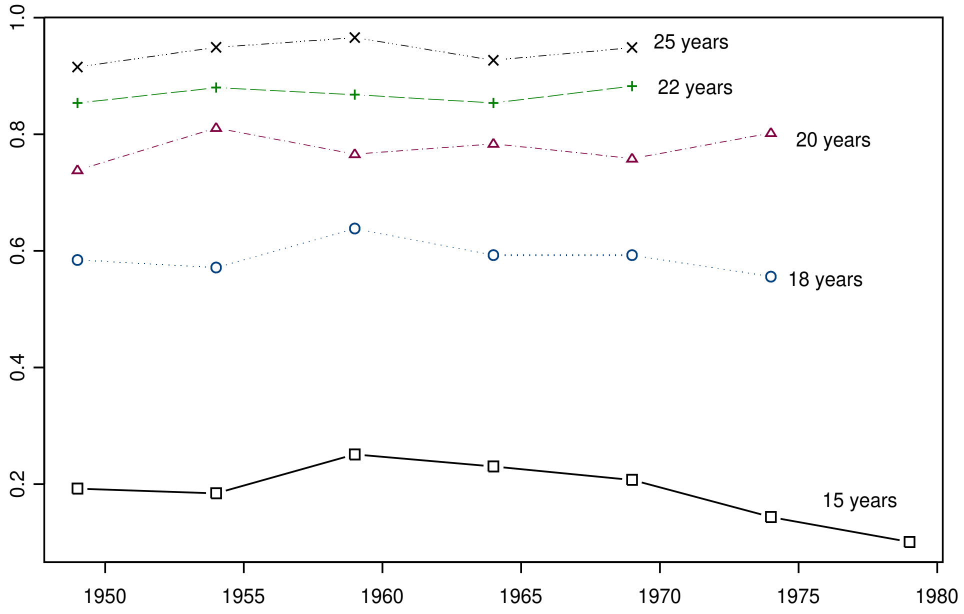

Fig. 2. Proportions (using Life-table estimates) of birth cohorts who had married by different exact ages

the discrete hazard function for this individual is given by:

$$h(t|x_{ijk(t)}, b_{ij}, b_i) = Pr(T = t|T \geq t; x_{ijk(t)}, b_{ij}, b_i) \tag{1}$$

which is the conditional probability of failure in interval $[a_{t-1}, a_t)$ given survival to the beginning of the interval. The corresponding discrete survival function is given by

$$S(t|x_{ijk(t)}, b_{ij}, b_i) = Pr(T > t|x_{ijk(t)}, b_{ij}, b_i) = \prod_{t=1}^{t}(1 - h(t|x_{ijk(t)}, b_{ij}, b_i)) \tag{2}$$

which is the probability of surviving beyond a_t. In a more general setting the observed response t is regarded as the expression of an underlying continuous process Y, commonly known as tolerance with distribution function F (Im and Gionola (1988)). The observed event is time t when the tolerance exceeds a certain threshold value, and a model for the discrete hazard function is usually given by

$$h(t|x_{ijk}, b_{ij}, b_i) = \mathrm{F}(\alpha_t + \beta' x_{ijk(t)} + b_{ij} + b_i) \tag{3}$$

where α_t represents a time-varying baseline effect and β is a vector of unknown fixed covariate effects. The extreme-value distribution function: $\mathrm{F}(\alpha_t + \beta' x_{ijk(t)} + b_{ij} + b_i) = 1 - \exp(-\exp(\alpha_t + \beta' x_{ijk(t)} + b_{ij} + b_i))$ and the logistic distribution: $\mathrm{F}(\alpha_t + \beta' x_{ijk(t)} + b_{ij} + b_i) = \exp(\alpha_t + \beta' x_{ijk(t)} + b_{ij} + b_i)/(1 + \exp(\alpha_t + \beta' x_{ijk(t)} + b_{ij} + b_i))$ are often used. The former is derived by assuming the Cox's proportional hazards model for the underlying failure times on a continuous scale, hence the fixed effects are the log of relative risks. The models are similar for short intervals (Thompson (1977)). In this study, we use the logistic model where each component of the parameter vector β is interpreted as the logarithm of the odds ratio of getting married to not getting married for the corresponding fixed effect covariate at each time point, and the random effects b_i and b_{ij} are the family and community specific logarithms of odds of marrying. An α_t represents the logarithm of the baseline odds of marrying at time t.

Now, suppose that woman ijk has observed discrete failure time t_{ijk}. The woman's contribution to the likelihood is given by:

$$L_{ijk} = \prod_{t=1}^{t_{ijk}} \frac{\left(e^{\alpha_t + \beta' x_{ijk(t)} + b_{ij} + b_i}\right)^{d_{ijkt}}}{1 + e^{\alpha_t + \beta' x_{ijk(t)} + b_{ij} + b_i}} \tag{4}$$

where $d_{ijkt} = 1$ if she was first married within $[a_{t-1}, a_t)$ years of age and $d_{ijkt} = 0$ otherwise. The contribution is just the likelihood obtained by treating the censoring indicators d_{ijkt} as independent Bernoulli trials across t (Prentice and Gloeckler (1978); Scheike and Jensen (1997); Xue and Brookmeyer (1997)). The three-level discrete time survival model presented here needs advanced estimation methods to estimate the baseline and fixed effects $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots \beta_p$ as well as all b_{ij}, b_i and the variance components ρ, σ_f^2 and σ_c^2. For an analysis of employment data with independent random effects, Biggeri, Bini and Grilli (2001) used the penalised quasi-likelihood, an approximate method. But as pointed out in Rodriguez and Goldman (2001), these approximate likelihood type methods result in substantial underestimation of both fixed and random effects when the random effects are large. In this paper, all the model parameters are estimated using a Bayesian approach via Markov chain Monte Carlo. We believe this to be a better methods to handle the complexity of the model.

4. Bayesian Inference

4.1. Setting Prior Distributions

In Bayesian analysis, the proposed model of the observed data is combined with the prior distribution of all the unknown model parameters. We first consider modelling of baseline effects α_t. The baseline effects could be modelled as a simple constant hazard function or Levy process. However, a correlated process prior is more appropriate for a process that evolves over time (Gray (1994); Qiou, Ravishanker and Dey (1999)). In particular, it is assumed that α_t follows a first-order autocorrelated prior process:

$$\alpha_t|\alpha_{t-1} = \alpha_{t-1} + e_t; \quad t = 1, \ldots, m \tag{5}$$

where $\alpha_0 = 0$ and $e_t \sim \text{Normal}(0, \nu_t^2)$ and independent of previous α_t's. For variable interval lengths, it seems reasonable to set $\nu_t^2 = \delta_t \nu_0^2$ where $\delta_t = a_t - a_{t-1}$, the length of interval I_t. The joint prior density of the baseline hazards is given by

$$\pi(\alpha) \propto (\nu_0^2)^{-m/2} \exp\left(-\frac{1}{2\nu_0^2}\sum_{t=1}^{m}\frac{(\alpha_t - \alpha_{t-1})^2}{\delta_t}\right). \tag{6}$$

This is a dynamic model as considered in Gamerman and West (1987), Gray (1994) and Fahrmeir and Knorr-Held (1997), who also allowed the baseline hazard and fixed-effect vectors to vary over the discretised time scale by using autoregressive prior processes. As pointed out by Fahrmeir and Knorr-Held, dynamic models do not impose any functional form for the dependence of parameters on time, and as such they can be regarded as nonparametric.

Next we consider the modelling of the cross-level log-frailties b_{ij} and b_i. These are typically assumed to be i.i.d variables from normal distributions (Biggeri, Bini and Grilli (2001)). In order to capture the cross-level dependence between log-frailties, the pair (b_{ij}, b_i) is given a bivariate normal distribution:

$$\begin{pmatrix} b_{ij} \\ b_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_f^2 & \rho\sigma_f\sigma_c \\ \rho\sigma_f\sigma_c & \sigma_c^2 \end{pmatrix}\right) \tag{7}$$

By setting the mean to zero, the specific log-frailties are interpreted as deviations from the overall average on the log-odds ratio scale. The fixed effects are modelled as multivariate normal with mean 0 and covariance matrix $d_0 I$ where d_0 was set to 1000, giving a prior which has its mass well spread over the region around zero, but also allows a wide range of other values.

For the variance components, we follow the usual practice of assigning them inverted gamma prior distributions. This is equivalent to specifying a gamma prior distribution for the corresponding precision parameters. Thus, the prior for the precision parameter of the baseline jump, $(\tau_0 = 1/\nu_0^2$, has a gamma distribution:

$$f(\tau_0) \propto \tau_0^{\kappa-1} \exp(-\omega\tau_0)$$

where κ and ω are positive parameters which should be specified. Ideally we will be looking at many intervals with small widths such that large jumps would be very unlikely. So we are looking for a prior which favours small values of ν_0^2. A Gamma$(0.001, 0.001)$ prior distribution for τ_0 favours small values of ν_0^2 but covers other values reasonably well.

The bivariate normal density for (b_{ij}, b_i) is factored into the product of the marginal density:

$$f(b_i|\sigma_c^2) \propto (\sigma^2)^{-1/2} \exp\{-\frac{1}{2\sigma_c^2} b_i^2\}$$

and the conditional density:

$$f(b_{ij}|b_i, \sigma_f^2, \sigma_c^2, \rho) \propto \left[\sigma_f^2(1-\rho)\right]^{-1/2} \exp\{-\frac{1}{\sigma_f^2(1-\rho)}(b_{ij} - \frac{\rho\sigma_f}{\sigma_c} b_i)^2\}.$$

In specifying prior distributions for the variance components $\sigma_c^2 (= 1/\tau_c)$ and $\sigma_f^2 (= 1/\tau_f)$, care should be taken when using very flat or non-informative priors as these can cause convergence problems and can lead to improper posterior distributions. To guard against these problems, proper priors will be used for the variance components.

We base our prior elicitation on data from nineteen Sub-Sahara countries (excluding Malawi) on timing of first birth as presented in Singh (1998). The use of evidence from data on first birth is appropriate since there is a very strong linear correlation (r=0.91) between time to first birth and time to first marriage as shown by Acsadi and Johnson-Acsadi (1990) for Sub-Sahara and other developing countries, and both are determined by the same factors. 95% of the nineteen countries, had odds ratio of early childbearing in the range 0.607 to 3.121 with mean and median values 1.900 and 1.990, respectively, among rural women compared to urban women; the range was 1.084 to 8.110 with mean and median values 3.77 and 2.900, respectively, among women with less than seven years of schooling compared to those with seven or more years of schooling. Clearly place of residence and education are important predictors. After allowing for these factors, we would expect the odds ratio between two families and communities to fall within these range, and the random effects to have small variations. The model assumes that 90% of the families have log(odds) of their women getting into early marriage in a range with width $2 \times 1.64/\sqrt{\tau_f}$. We also thought it plausible that there was a three-fold variation between the odds of two families at the extreme ends of the 90% interval, with a five-fold variation being very unlikely. This implies that 9 is a prior guess for τ_f. This leads to a prior $\tau_f \sim \text{Gamma}(9, 1)$ whose implied distribution for σ_f^2 is reasonable, even though it precludes very low values and very high values. The same prior distribution was also used for τ_c. For the cross-level correlation, there is not enough prior information, so a uniform(-1,1) prior is placed on ρ.

4.2. Computing the Joint Posterior Distribution

For simplicity, prior independence is assumed of all the model parameters. The joint posterior distribution of $\theta = \{\alpha, \beta, b_{ij}, b_i, \nu, \tau_f, \tau_c, \rho\}$ is

$$\begin{aligned} f(\theta|Y) \quad \propto \quad & \left[\prod_{i=1}^{N}\prod_{j=1}^{J_i}\prod_{k=1}^{K_{ij}}\prod_{t=1}^{t_{ijk}} \frac{\left(e^{\alpha_t+\beta' x_{ijk(t)}+b_{ij}+b_i}\right)^{d_{ijkt}}}{1 + e^{\alpha_t+\beta' x_{ijk(t)}+b_{ij}+b_i}}\right] \\ \times \quad & \left[\prod_{i=1}^{N}\prod_{j=1}^{J_i} \tau_f^{1/2}(1-\rho)^{-1/2} \exp(-\frac{\tau_f}{2(1-\rho)}(b_{ij} - \frac{\rho\sqrt{\tau_c}}{\sqrt{\tau_f}} b_i)^2\right] \\ \times \quad & \left[\prod_{i=1}^{N} \tau_c^{1/2} \exp(-\frac{\tau_c}{2} b_i^2)\right] \end{aligned}$$

$$\times \left[\prod_{t=1}^{m} \tau_0^{1/2} \exp(-\frac{\tau_0}{2}\frac{(\alpha_t - \alpha_{t-1})^2}{\delta_t})\right]\left[\prod_{r=1}^{p} \exp(-\frac{1}{2}\beta_r^2/d_0)\right]$$
$$\times \quad \tau_f^{\xi-1} \exp(-\phi\tau_f)\,\tau_c^{\xi-1} \exp(-\phi\tau_c)\,\tau_0^{\kappa-1} \exp(-\omega\tau_0) \tag{8}$$

where Y is the total vector of the observed responses. An analytical solution for the joint posterior distribution is not easily obtainable as it involves complex integrals. We use the Gibbs sampler (Geman and Geman (1984); Tierney (1994)), a Markov chain Monte Carlo method for sampling from a joint distribution. It is widely used in Bayesian statistics to approximate joint posterior distributions involving difficult integrals (Gelman, et al. (1995)), and is implemented in the WinBUGS computer software (Spiegelhalter, Thomas and Best (2000)). The algorithm requires that sampling is to be done from all full conditional posterior distributions of the parameters, usually univariate conditional posterior distributions. Each of the univariate full conditional posterior distributions is proportional to (8), and involves terms that are relevant to the parameter under consideration.

The univariate full conditional posterior distribution for a baseline effect is

$$f(\alpha_t|\beta, \{b_i\}, \{b_{ij}\}, \tau_0, Y) \propto \left[\prod_{(i,j,k)\in R_t} \frac{\left(e^{\alpha_t+\beta' x_{ijk(t)}+b_{ij}+b_i}\right)^{d_{ijkt}}}{1 + e^{\alpha_t+\beta' x_{ijk(t)}+b_{ij}+b_i}}\right]$$
$$\times \quad \exp[-\frac{\tau_0}{\delta_t}(\alpha_t - (\alpha_{t+1} + \alpha_{t-1})/2)^2],\ t = 1, \ldots, m$$

where R_t is the set of individuals at risk at time t; for $t = 1$ set $\alpha_{t-1} = \alpha_{t+1}$and for $t = m$ set $\alpha_{t+1} = \alpha_{t-1}$. This conditional distribution is log-concave in α_t, so it is amenable to adaptive-rejection sampling of Gilks and Wild (1992). The univariate posterior conditional distribution for a single fixed-effect parameter β_r is

$$f(\beta_r|\alpha, \beta_{-r}, \{b_i\}, \{b_{ij}\}, \tau_0, Y) \propto \left[\prod_{i=1}^{N}\prod_{j=1}^{J_i}\prod_{k=1}^{K_{ij}}\prod_{t=1}^{t_{ijk}} \frac{\left(e^{\alpha_t+\beta' x_{ijk(t)}+b_{ij}+b_i}\right)^{d_{ijkt}}}{1 + e^{\alpha_t+\beta' x_{ijk(t)}+b_{ij}+b_i}}\right] \exp[-\beta_r^2/d_0]$$

where β_{-r} is the vector of fixed-effect parameters excluding the rth element. This conditional is log-concave in β_r, so it is sampled by adaptive rejection sampling.

The univariate full conditional posterior distribution for a family random effect parameter is

$$f(b_{ij}|\alpha, \beta, b_i, \tau_f, \tau_c, \rho, Y) \propto \left[\prod_{k=1}^{K_{ij}}\prod_{t=1}^{t_{ijk}} \frac{\left(e^{\alpha_t+\beta' x_{ijk(t)}+b_{ij}+b_i}\right)^{d_{ijkt}}}{1 + e^{\alpha_t+\beta' x_{ijk(t)}+b_{ij}+b_i}}\right]$$
$$\times \left[\exp(-\frac{\tau_f}{2(1-\rho)}(b_{ij} - \frac{\rho\sqrt{\tau_c}}{\sqrt{\tau_f}}b_i)^2\right]$$

and that of a community random effect is given by

$$f(b_i|\alpha, \beta, b_{i1}, \ldots, b_{ij_i}, \tau_c, Y) \propto \left[\prod_{j=1}^{J_i}\prod_{k=1}^{K_{ij}}\prod_{t=1}^{t_{ijk}} \frac{\left(e^{\alpha_t+\beta' x_{ijk(t)}+b_{ij}+b_i}\right)^{d_{ijkt}}}{1 + e^{\alpha_t+\beta' x_{ijk(t)}+b_{ij}+b_i}}\right] \exp(-\frac{\tau_c}{2}b_i^2)$$

both of which are log-concave in the respective parameters.

The precision of the jump in the baseline effect has the univariate full conditional posterior distribution given by

$$f(\tau_0|\alpha) \propto \tau_0^{(\kappa+m/2)-1} \exp\left[-(\omega + \frac{\sum(\alpha_t - \alpha_{t-1})^2}{2dt})\tau_0\right]$$

which is a gamma distribution. This parameter is sampled directly. The univariate full conditional posterior distribution for the precision of the family random effect is

$$f(\tau_f|\{b_i\},\{b_{ij}\},\tau_c,\rho,Y) \propto \left[\prod_{i=1}^{N}\prod_{j=1}^{J_i} \tau_f^{1/2} \exp(-\frac{\tau_f}{2(1-\rho)}(b_{ij} - \frac{\rho\sqrt{\tau_c}}{\sqrt{\tau_f}}b_i)^2\right] \tau_f^{\xi-1}\exp(-\phi\tau_f)$$

and that of the community random effect precision is

$$\begin{aligned} f(\tau_c|\{b_i\},\{b_{ij}\},\tau_f,\rho,Y) &\propto \left[\prod_{i=1}^{N}\prod_{j=1}^{J_i} \tau_f^{1/2} \exp(-\frac{\tau_f}{2(1-\rho)}(b_{ij} - \frac{\rho\sqrt{\tau_c}}{\sqrt{\tau_f}}b_i)^2\right] \\ &\times \left[\prod_{i=1}^{N} \tau_c^{1/2}\exp(-\frac{\tau_c}{2}b_i^2)\right] \tau_c^{\xi-1}\exp(-\phi\tau_c). \end{aligned}$$

both of which are sampled by slice sampling (Neal (1997)). Finally, the univariate posterior conditional distribution for the cross-level correlation is

$$f(\rho|\{b_i\},\{b_{ij}\},\tau_f,\tau_c,Y) \propto \left[\prod_{i=1}^{N}\prod_{j=1}^{J_i} (1-\rho)^{-1/2} \exp(-\frac{\tau_f}{2(1-\rho)}(b_{ij} - \frac{\rho\sqrt{\tau_c}}{\sqrt{\tau_f}}b_i)^2\right]$$

which also is sampled by slice sampling.

5. Results

The time axis was partitioned into two-year intervals from 10 to 36, so that the discrete time ranges from 1 to 13. Regression and interval parameters were initially set to zero, frailty variances to unity, and cross-level correlation to zero. Given these initial variance component values, starting values for the frailty terms were generated from their priors. A single chain of 50,000 iteration was run, and using convergence diagnostics based on plots of sampled values and Raftery and Lewis (1992) methods, 10,000 runs were sufficient for the burn-in period. The remaining 40,000 iterations passed the Heidelberger and Welch stationary and halfwidth tests. Also, the integrated autocorrelation times (IACT) (Sokal (1996)) were calculated for parameters of interest on the 40,000 iterations. The IACT measures the number of correlated sample in the MCMC chain having the same efficiency as a single independent sample. The maximum value of IACT was 50, which is a reasonable value considering the complexity and dimensionality of the parameter space of the model. All post-convergence iterations (10001-50000) were retained for posterior summaries.

The estimated mean baseline hazard function is shown in Fig. 3, with 95% credible intervals. It shows the baseline hazard has an increasing trend up to year 25, then declining until around year 30. Subsequently, the hazard increases then declines. Increases in the

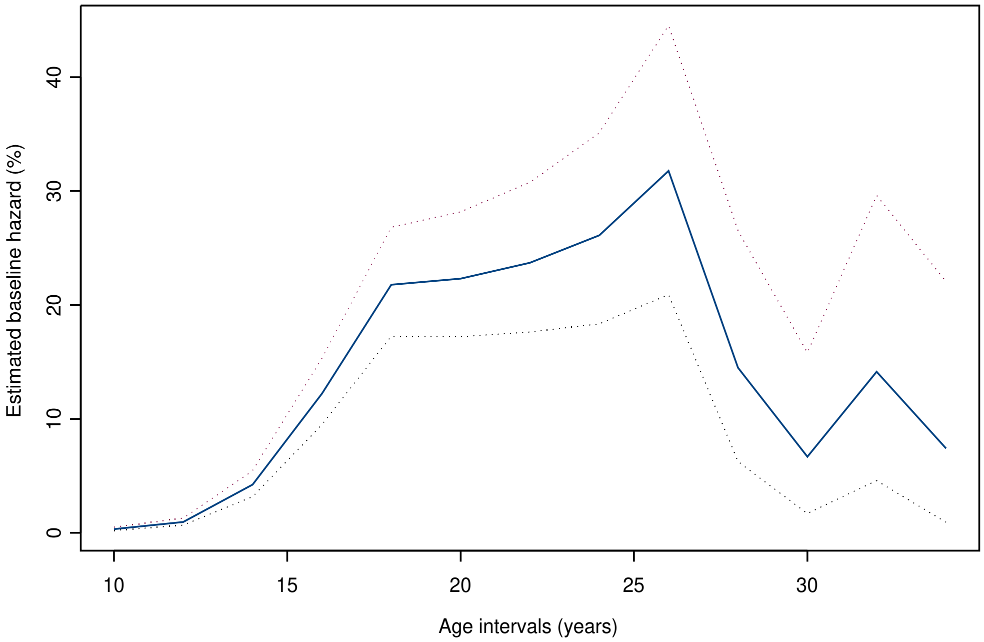

Fig. 3. Estimated mean baseline hazard function with 95% credible interval

Table 2. Posterior estimate of the fixed effect parameters

Parameter	Mean	STD	Median	Mean Odds Ratio
Year of birth				
1977-81	0.0	-	-	1.000
1972-76	0.559	0.105	0.561	1.788
1967-71	0.683	0.107	0.683	2.050
1962-66	0.629	0.113	0.629	1.951
1957-61	0.727	0.125	0.729	2.169
1952-56	0.519	0.127	0.519	1.765
1947-51	0.251	0.151	0.252	1.344
Residence				
Urban	0.0	-	-	1.000
Rural	0.112	0.078	0.112	1.132
Region				
Norther	0.0	-	-	1.000
Central	-0.532	0.099	-0.532	0.574
Southern	-0.423	0.098	-0.426	0.652
Woman's education				
No Education	1.248	0.119	1.246	3.832
Primary education	1.157	0.099	1.156	3.422
Secondary+	0.0	-	-	1.000

baseline hazard from 30 years of age may be due to sparse data in these intervals, and more likely to reporting bias mentioned earlier. Thus, we restrict trend analysis up to 30 years. In contrast to a peak hazard at about 18 years as seen in Fig. 1, we now have a peak hazard at 25 years. This reflects the fact that now we are estimating the baseline hazard for women in the low risk group, ie. who were born in 1973-81, in urban areas, and who obtained secondary education. The hazard function in Fig. 1 is global.

Table 2 shows the posterior means, standard deviations (STDs) and medians for the fixed effects parameters together with the posterior means of the odds ratios ($\exp(\beta_r)$). The most significant variable was the woman's education. The odds of marrying early among women with less than 8 years of education (Primary) are about 3 times higher than the odds among women with more than 8 years of education. Birth cohort and region were also found to be important determinants of rates of first marriage. Place of residence, though not significant, shows that early marriage rates among rural women are increased by 10%. Posterior correlations between the fixed effects parameters were also examined, and showed that only estimates defined within the same variable were strongly related; no educated and primary education ($corr = 0.750$), central and southern regions ($corr = 0.633$) and between birth cohorts estimates ($0.403 <= corr <= 0.635$).

Fig. 5 shows posterior densities of standard deviations of the random effects as well as of the jumps in the baseline effect. The posterior mean of the family and community effect standard deviations are 0.390 ± 0.070 and 0.327 ± 0.049 respectively while the standard deviation of baseline jump has a posterior mean of 2.038 ± 0.459. The cross-level correlation is estimated with posterior mean -0.340 ± 0.177. Lewis and Raftery (1999) derived

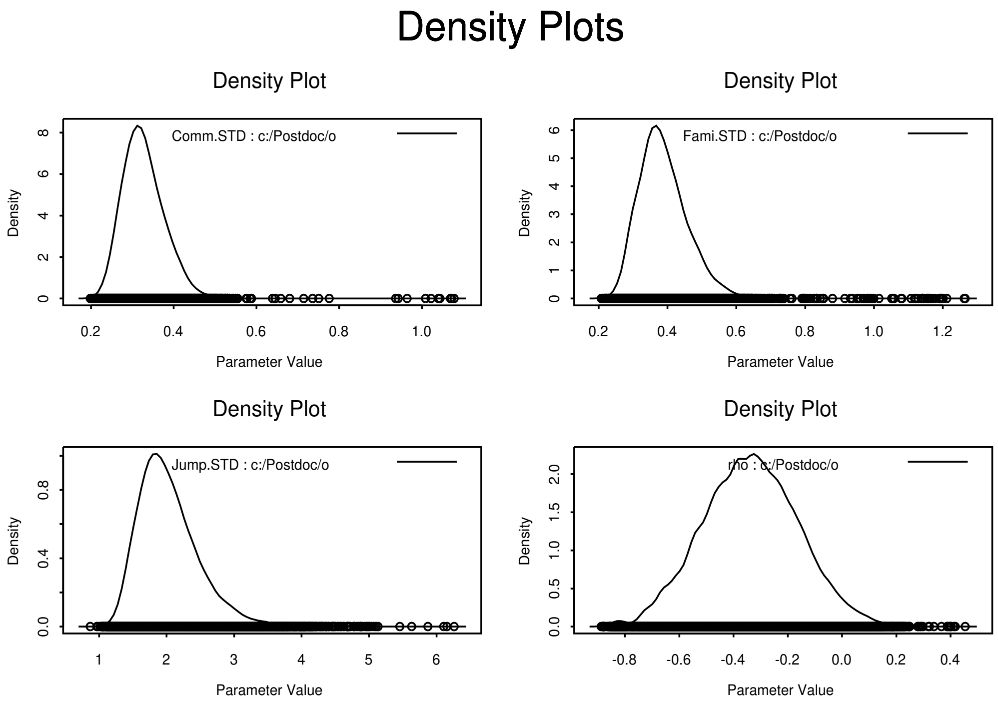

Fig. 4. Posterior densities for the standard deviations of random effects and baseline jump and for the cross-level correlation

an inverted Gamma$(0.470, 0.023)$ prior distribution for the variance of the women-specific random effect in the context of Iran Fertility analysis. We also used this prior for the variance components. The posterior means of the fixed-effects parameters were essentially the same as those obtained for our prior. There was a slight difference in the random effect parameters, which now have posterior means 0.500 ± 0.148, 0.234 ± 0.118 and -0.188 ± 0.305, respectively for standard deviations for the family and community effects and the cross correlation. This shows that estimates of random effect parameters are all significant, with more variations due to the family effect than to the community effect. Thus, variations in entry to marriage depends more on differences at the family level than at the community level. The cross-level correlation indicates that a family with high early marriage rate may as well reside in community with a generally lower rate of early marriage.

In a model, where the random effect parameters are significant, inference of the risk factors is not as simple as in the logistic model without random effects where the exponents of fixed-effect parameters are the odds ratio of marrying early. Now the exponents must be adjusted by a magnitude depending on the family and community effects. However, we can consider families and communities with specified random effect values. Fig. 5, shows estimated hazard functions for women in various education levels with other covariates and random effects set to 0. The hazard functions have similar shapes, but women with secondary education have smaller hazards in all age intervals except in very low ages.

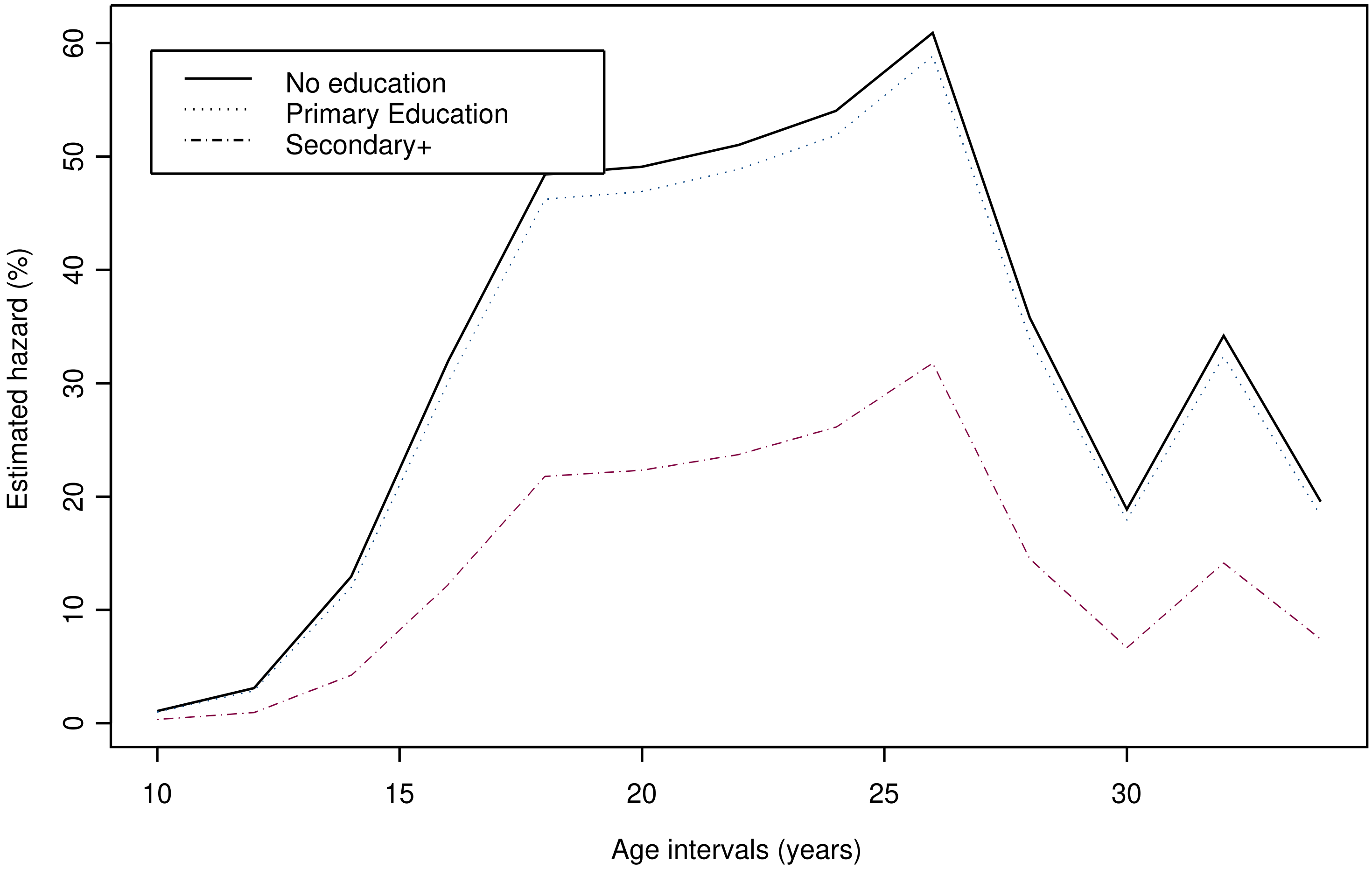

Fig. 5. Estimated mean hazard function by woman's education with the other covariates and random effects set to 0

The posterior means and standard deviation presented in Table 2 were compared to the estimates found using the discrete-time survival model without family and community random effects (results not shown). The estimates of the fixed-effect parameters and their standard deviations in the model without random effects were slightly underestimated, though substantially the conclusions on the fixed effects parameters are unchanged. Thus, for this study the conclusions are unaffected by any unobserved heterogeneity which is reassuring given that most analyses of this type ignore unobserved differences. We also tried to use flat priors such Gamma(0.01,0.01) and Gamma(0.001,0.001) for the precision of the random effects. We experienced convergence difficulties for the random effect parameters, and therefore the results are not given here. This is the more reason for introducing prior opinion as it was expected that the likelihood would be fairly flat.

6. Conclusion

This paper introduces a Bayesian hierarchical model for the analysis of discrete-time survival data. The data concerns age at first marriage among women in Malawi. This is an important application as age at first marriage is a major determinant of fertility of a society. Besides modelling the influence of explanatory variables, the model allows for unobserved heterogeneity at two levels of hierarchy as well as cross-level association. Rather than making rigid parametric assumptions, the model also includes a flexible nonparametric prior for the discrete baseline hazard. The model is fitted using the Gibbs sampler, a simulation-based Bayesian computational algorithm. The implementation is simple and effective, and provides a sample from the joint posterior distributions of all the model parameters. In contrast to maximum-likelihood based techniques that provide only point estimates, we can employ the sample from the joint posterior to explore the behaviour of parameters and any function of the parameters of interest.

Women's education was found to be the most significant factor that determines entry into marriage. Better educated women enter marriage at substantially higher ages. There were also substantial birth cohort differences, with young birth cohorts delaying entry into marriage. We identified an important rural-urban difference, with rural women marrying at much younger ages. These findings are consistent with results obtained in elaborate analyses of Singh (1998) and Kirk and Pillet (1998) that young educated women in urban areas are at a substantially reduced risk of early marriage. The observed differentials in education operate through other elements of socioeconomic development that are associated with education. Educated women have the motivation and independence to regulate their entry into marriage. The government of Malawi has acknowledged the importance of improving girl's education to curb the negative social and economic consequences of early marriage. The government has embarked on programs that strengthen education and employment and family planning programs in an effort to contribute to lowering early marriage rates. But these services need strong and sustained commitment by the government, a task made even harder by the adverse economic conditions and resources.

In Malawi fertility levels are high and there is an encouragement of early reproductive and marriage especially in rural areas (Kavinya (2001)). Changing the social climate would have an impact on attitudes towards early marriage and motherhood which are the traditional ways in which a woman affirms her value and identity within a family and community. Lack of opportunities can still cause early marriage even where high education is desired by many. Policy makers, therefore have the task and responsibility to initiate population education regarding the advantages of delayed marriage. It is important to build a public consensus about an acceptable marriageable age. Acsadi and Johnson-Acsadi (1990) argue for enacting marriage laws to minimise the harmful effects of early marriage and that proper sanctions against early marriage are needed and should be enforced. But these efforts would be unfruitful if levels of women's education remain very low in Malawi, and opportunities of employment are lacking.

The results also indicate that there are not only substantial differences in early marriage rates among families, but also among communities. The unobserved family and community heterogeneity are important as family and community level covariates have not been included in the model. The unobserved family heterogeneity includes social and economic status of the family while the unobserved community heterogeneity takes account of marriage norms and other factors relating to marriage at the community level not included in the model. The strength of both variations implies that targeting both families and communities in a program to improve reproductive health would bear fruits. It has also been shown that there is negative association between the family effects and community effects. In terms of policies, this implies that targeting communities with an intervention to improve reproductive health may not necessary benefit individual families. A campaign that is geared towards individual families would achieve the desired goals.

Acknowledgements

The work of Dr Manda was supported by a grant from the Wellcome Trust, while that of Dr Meyer was supported by the Royal Society of New Zealand Marsden fund. The authors thank the Macro International Inc. and the Malawi's National Statistical Office for the permission to use the data.

References

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. In *Sociological Methodology* (eds S. Leinhardt),pp. 61–98. Jossey-Bass: San Francisco.

Acsadi, G. T. F and Johnson-Acsadi, G. (1990). Effects of timing of marriage on reproductive health. *A World Bank Symposium*, Washington.

Arjas, E. and Haar, P. (1987). A logistic regression model for hazard: asymptotic results. *Scandinavian Journal of Statistics*, **14**, 1–18.

Biggeri, L., Bini, M. and Grilli, L. (2001). The transition from university to work: a multilevel approach to the analysis of time to obtain the first job. *Journal of the Royal Statistical Society*, **A**, **164**, 293–305.

Bolstad, W. M. and Manda, S. O. M. (2001). Investigating child mortality in Malawi using family and community random effects: a Bayesian analysis. *Journal of the American Statistical Association*, **96**,12–19.

Brien, M. J., Lillard, L. A. and Waite, L. J. (1999). Interrelated family-building behaviours: cohabitation, marriage, and nonmarital conception *Demography*, **36**, 535–551.

Caldwell, J. (1976). Towards a Restatement of Demographic Transition Theory. *Population and Development Review*, **2**, 321–366.

Caldwell, J. and Calwell, P. (1997). What do we know about fertility transition? In *The Continuing Demographic Transition* (eds G. W. Jones, Douglas, R. M., J. C. Caldwell and R. M. D'Souza), pp. 13–35. Clarendon Press, Oxford.

Cook, R. J. and Ng, E. T. (1997). A logistic bivariate normal model for overdispersed two-state Markov processes. *Biometrics*, **53**, 358-364.

Fahrmeir, L. and Knorr-Held, L. (1997). Dynamic discrete-time duration models: Estimation via Markov chain Monte Carlo *Sociological Methodology* **27**, 417–452.

Gamerman, D. and West, M. (1987). An application of dynamic survival model in unemployment studies. *The Statistician* **36**, 269–274.

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

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B.(1995). *Bayesian Data Analysis.* London: Chapman and Hall.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721-741.

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

Gray, R. J. (1994). A Bayesian analysis of institutional effects in a multicentre clinical trial. *Biometrics* **50**, 244–253.

Gustafson, P. (1997). Large hierarchical Bayesian analysis of multivariate survival data. *Biometrics* **53**, 230–242.

Government of Malawi. (1992). Demographic and Health Survey. Malawi National Statistics Office, Zomba.

Im, S. and Gionola, D. (1988). Mixed models for binomial data with an application to lamb mortality. *Applied Statistics*, **37**, 196–204.

Kavinya, A. (2001). Spousal Communication and Reproductive Decision-Making in Malawi. Ph D. Thesis, University of Waikato, Hamilton..

Kirk, D. and Pillet, B. (1998). Fertility levels, trends and differentials in sub-Sahara African in the 1980s and the 1990s. *Studies in Family Planning* **29**, 1–22.

Korn, E. L. and Graubard, B. I. (1999). *Analysis of Health Surveys.* John Wiley Sons, New York.

Lewis, S. M. and Raftery, A. E.(1999). Bayesian analysis of event history models with unobserved heterogeneity via Markov chain Monte Carlo: Application to the explanation of fertility decline. *Sociological Methods and Research* **28**, 35-60.

Neal, R. (1997). Markov chain Monte Carlo methods based on slicing the density function. Technical Report, Department of Statistics, University of Toronto.

Prentice, R. L. and Gloeckler, L. A. (1978). Regression analysis of grouped survival data with applications to breast cancer data. *Biometrics* **34**, 57–67..

Qiou, Z., Ravishanker, N. and Dey, D. K. (1999). Multivariate survival analysis with positive stable frailties. *Biometrics* **55**, 637–644.

Raftery, A. L. and Lewis, S. (1992). Comment: One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science* **7**, 493–497.

Rodriguez, G. and Goldman, N. (2001). Improved estimation procedures for multilevel models with binary responses: a case study. *Journal of the Royal Statistical Society*, **A**, **164**, 339–355.

Santow, G. and Bracher, M. (1994). Change and continuity in the formation of first marital unions in Australia. *Population Studies*, **48**, 475–496

Sastry, N. (1997). A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. *Journal of the American Statistical Association* **92**, 426–435.

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.

Singh, S.(1998). Adolescent childbearing in developing countries: a global review. *Studies in Family Planning*, **29**, 117-136.

Sokal, A. D. (1996). Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms. Lectures at the Cargese Summer School on " Functional Integration: Basics and Applications,.

Spiegelhalter, D., Thomas, A. and Best, N. (2000). *WinBugs Version 1.3; User Manual.* MRC Biostatistics Unit, Cambridge, UK.

Thompson, W. A. (1977). On the treatment of grouped observations in life studies. *Biometrics* **39**, 141-147.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* **22**, 1701–1762.

Xue, X. and Brookmeyer, R. (1997). Regression analysis of discrete time survival data under heterogeneity. *Statistics in Medicine* **16**, 1983–1993.