# Frailty models with applications to the study of infant deaths on birth timing in Ghana and Kenya

Stephen Obeng Gyimah[1]

1 Associate Professor, Department of Sociology, Queens University, Kingston, Ontario, Canada and Sabbatical Fellow, African Population & Health Research Center, Nairobi, Kenya

Abstract

In hazard models, it is assumed that all heterogeneity is captured by a set of theoretically relevant covariates. In many applications however, there are ample reasons for unobserved heterogeneity due to omitted or unmeasured factors. If there is unmeasured frailty, the hazard will not only be a function of the covariates but also of the unmeasured frailty. This paper discusses the implications of unobserved heterogeneity on parameter estimates with application to the analysis of infant death on subsequent birth timing in Ghana and Kenya using DHS data. Using Lognormal Accelerated Failure Time models with and without frailty, we found that standard models that do not control for unobserved heterogeneity produced biased estimates by overstating the degree of positive dependence and underestimating the degree of negative dependence. The implications of the findings are discussed.

**Keywords**: Unobserved heterogeneity, frailty, Birth timing, Ghana, Kenya.

## Introduction

Event history models have become a standard statistical tool in analyzing longitudinal and duration data in social science research. In demographic applications, parametric, non-parametric, and semi-parametric models are often used to model transition data. In such applications, it is assumed that all heterogeneity is captured by theoretically relevant covariates (Trussell and Richards, 1985; Trussell and Rodriguez, 1990; Van den Berg, 2001). In many situations, however, there are ample reasons to suspect omitted or unmeasured factors. That is, while some individuals will be more at risk of experiencing the event, it is unlikely the underlying reasons for this variability will be fully captured by the observed covariates. If there is unmeasured frailty, the hazard will not only be a function of the covariates but also of the frailty. To assess the true effects of the observed covariates under this circumstance, some have stressed the need to explicitly account for unobserved heterogeneity (Aalen 1994; Baker and Melino 2000; Baschieri and Hinde 2007; Blossfeld and Hamerle 1992; Heckman et al., 1985; Heckman and Singer, 1984; Jenkins, 2005; Lancaster, 1990; Manda and Meyer 2005; Vaupel et al., 1979).

Indeed, results from several empirical and simulation studies have shown that accounting for unobserved heterogeneity significantly improves overall model fitness (Heckman et al., 1985; Sastry, 1997; Trussell and Rodriguez, 1990; Blossfeld and Hamerle, 1992). There is evidence that the failure to correct for unobserved heterogeneity can lead to hazards that either decline steeply or rise more slowly than the true hazard, and the parameter estimates and their associated standard errors will be biased. Correcting for unobserved heterogeneity in event history models is thus seen as a way of correcting for model misspecification, and the frailty approach is a statistical modeling concept that aims to account for heterogeneity caused by unmeasured factors (Guiterrez, 2002).

Although there is considerable research on the effects of unobserved heterogeneity, most of the empirical assessments have been done using data from developed countries. This paper contributes to the discussion by applying frailty models in examining how women's experience with the death of the first birth in infancy associates with the timing of second births using DHS data from Ghana and Kenya. The frailty models discussed are observation-level frailty and not shared frailty models where it is assumed that multiple observations share the same individual frailty. The goal of the paper is methodological. We provide substantive arguments for unobserved heterogeneity in the study which derives from several substantive concerns. First, the absence of some important correlates of the timing of the second birth, including the frequency of intercourse, duration and viability of ova and sperm, induced abortions, and miscarriage in the data necessitates such a model. Also, some women can be thought of as more fecund than others because of unmeasured genetic predisposition while others might be sub-fecund for similar reasons, resulting in differential hazards independent of the observed covariates. Additionally, the probability of a mother experiencing an infant death, the primary independent variable in this study, may be linked to variations in her susceptibility to illness during pregnancy which will ultimately compromise the health of the fetus, and later the infant. To accomplish the objectives of the paper in holistic way, the next two sections provide the context of the population dynamics in both countries as well as the underlying theoretical links between infant mortality and birth timing.

## Population dynamics in Ghana and Kenya: An overview

Kenya and Ghana are among the first group of countries in sub-Saharan Africa to adopt explicit and comprehensive population policies. The major goal of the 1967 Kenyan policy was to reduce the high rate of population growth, and improve the health of mothers and children (NCPD and IRD, 1989). The program however achieved limited successes and its failure was attributed to the over-emphasis on the supply side of the family planning program with little regard to programs aimed at changing norms and attitudes regarding family size. To rectify the weaknesses in the policy, the National Council for Population and Development (NCPD) was established in 1982 not only to formulate population policies and programs but also to coordinate the activities of the various governmental and non-governmental organizations involved with population and family planning issues.

Ghana's attempt at population policy parallels that of Kenya. To achieve policy objectives set up in Ghana's 1969 policy, the National Family Planning Program was launched in May 1970 as a coordinating department within the Ministry of Finance and Economic Planning. Although the policy specified sectors such as education and health within which population programs were to be organized, these components did not receive as much attention as the family planning program. Just like the Kenyan case, there was too much emphasis on the supply side of the family planning component, and poor institutional coordination among organizations engaged in population issues. Against this backdrop and also emerging issues such as HIV/AIDS, the 1986 National Conference on Population and National Reconstruction recognized the need for a revision of the policy. This culminated in a revised policy and the establishment of the National Population Council (NPC) in 1994 as the highest statutory body to advise the government on population related issues (Population Impact Project, 1995).

Notwithstanding the failures of their initial population policies, the two countries are among the few in sub-Saharan Africa to witness substantial decline in fertility and childhood mortality. In Ghana, for example, the infant mortality rate (IMR) declined from 100 per 1000 in 1975 to 57 per 1000 in 1995 while under-five mortality declined from 187 per 1000 to 107 per 1000 during the same period. Similarly, IMR in Kenya declined by 7 percent between 1976 and 1986, although it has been on the rise in recent years partly due to AIDS and AIDS-related mortality (UN-AIDS, 2000). While the 1990s witnessed a remarkable fertility decline in both countries, fertility has stalled in the most recent periods. Between 1978 and 1998, Kenya's total fertility rate declined by 42 percent from 8.1 births to 4.7 births (Brass and Jolly, 1993; NCPD and IRD, 1989; NCPD and MI, 1999). Similarly, Ghana's total fertility rate declined from 6.6 birth in 1982-84 to 4.5 births in 1998 (GSS and MI, 1999). Additionally, the mean number of children ever born to women aged 45-49 years declined from 7.3 to 5.9 between 1988 and 1998 in Ghana, and from 7.6 to 6.9 in Kenya during the same period. The stall in fertility in the last decade or so has been attributed to the 'leveling off or near leveling off' of the demand and use of contraceptives (Bongaarts, 2005). This notwithstanding, there is evidence that both countries have recorded noticeable decline in child mortality and fertility in the last three decades.

## Theoretical context

In examining the timing of the second birth in this study, the survival status of the index child in the first year of life is conceptualized as the primary independent variable. The theoretical pathways through which infant death affects birth timing have been extensively discussed in the literature (Gyimah and Rajulton 2003, 2004; Montgomery and Cohen 1998; Palloni and Rafalimanana 1998). These are mostly a combination of short-term physiological and long-term insurance and replacement effects. The short-term effect operates mainly through the cessation of breastfeeding following the death of the child which triggers the resumption of ovulation. As a result of this biological process, women who have lost a child in infancy are exposed to the risk of pregnancy more quickly and therefore have a quicker pace of a subsequent birth than those whose child had survived the first year of life. In its usual rendition, replacement effect refers to deliberate and conscious effort a couple makes to replace dead children. This is often achieved through lower contraceptive usage, among others, which significantly increases the risk of the next birth.

Besides infant mortality, a number of theoretically relevant demographic, socio-economic, and socio-cultural factors relating to the timing of births are used as controls in this study (see Table 1 for distribution). The cross-sectional nature of our data requires the use of current status information which implicitly requires a Markovian assumption of time invariance which is clearly not the case for all the variables, a limitation that needs to be recognized. The demographic controls include current age, age at first birth, and age at first marriage. The ages at which women marry and have their first birth are of tremendous importance in fertility studies because of their inverse relation to the exposure to the risk of conception (Gyimah 2003; Westoff 1992). They also shed light on characteristics that predispose women to differential timing of births and thus overall fertility. Women who have first births earlier are likely to come from disadvantaged socio-economic background and are thus, more likely to be associated with the higher risk of births than their counterparts whose first birth occurs late[2].

Another demographic factor with significant bearing on reproductive behavior is the age of the woman which may also be indicative of cultural, social-economic and political contexts that shape the life course experiences of individual women. While there are significant differences in the experiences of individual women, the prevailing socio-economic and cultural contexts may lead them to similar life experiences. In this paper, four age groups are identified: less than 25 years, 25-34 years, 35-44, and 45-49 years. Contextually, the younger women became adolescents in a period of more egalitarian gender roles, more efficient contraceptives, higher enrollments in formal education and labor force participation compared with the older generations who were exposed to well-defined gender roles and became adolescents at the time when contraception was less available, and when there was a much lower female enrollment in formal education.

The socio-economic controls included educational attainment, place of residence and household wealth. There is considerable empirical evidence on the effects of maternal education and urban residence on both fertility and child mortality and these are largely negative (Cochrane 1983; Martin, 1995). The pathways through which maternal education affects fertility have been explained through proximate factors such as age at marriage, high contraceptive use, and labour force participation and cultural factors such as beliefs and norms regarding reproduction

**2** There are few exceptions to this pattern. In South Africa, for instance, early age at first birth (often pre-marital) is often followed by long delays to the second birth and increased educational attainment.

(Gyimah, Maxim, and White 2005). Again, there are significant differences in fertility by place of residence which shed light on the differential influence of environmental, socio-economic and cultural factors on reproduction. Finally, past research has shown the relevance of socio-cultural factors on reproduction (Addai 1999; Gyimah 2005). In this study, the socio-cultural variables controlled are ethnicity, religion, type of marriage, and living arrangement of the husband. Both countries have ethnic practices and norms that differentially reflect on duration and intensity of breastfeeding as well as postpartum sexual abstinence which have repercussions for the timing of births. To control for this, five ethnic groups in Ghana (*Akan, Ga-Adangbe, Ewe, Mole-Dagbani, Others)* and six in Kenya (*Kikuyu, Luo, Luhya, Kalenjin, Kamba, Somalis, Others)* are identified. Ethnicity is expected to capture some of the unobservable behavioral and cultural differences that affect the timing of births. The importance of religion on reproductive-related behavior is reflected in differences in attitudes, norms and beliefs regarding birth control and the value of children. In this study, religious affiliation is categorized into the following (1) Christians (2) Others (3) Muslim (4) Traditional and (5) Others. On the basis of the complexity of the relationship between the type of marital union and fertility in sub-Saharan Africa, we classified currently married women into monogamous or polygynous unions.

**Data and Estimation**

The 2003 DHS women's files for Ghana and Kenya are used for the study. The surveys are nationally representative self-weighting samples of women in the reproductive ages of 15-49 years. The respondents for the interview were women who had spent the previous night in the selected households, and the overall response rates were quite high in both countries. In addition to country specific modules in the DHS, standard core questionnaires on reproductive history, family planning knowledge, household and demographic characteristics are used in all countries that participate in the survey, making cross-national comparisons possible. Like other retrospective surveys, a major limitation of the DHS is that the characteristics of the respondents relate to the time of the interview and not to the date of birth of each child. Thus, in the estimation of timing of births, the assumption needs to be made that these maternal characteristics are time invariant. The analysis reported here is based on ever-married women with at least one birth.

Examining the timing of births using retrospective data as this paper does, introduces two types of biases: namely, selectivity and censoring. Selectivity means that the transition from parity $i$ to $i+1$ can only be studied for women with at least parity $i$ at the time of the survey. These women, however, tend to be selected on a number of characteristics and are, therefore, not representative of the entire population. Selectivity in this study is controlled as much as possible by introducing theoretically relevant socio-economic and demographic covariates on the timing of births discussed earlier. Women who have not experienced a second birth at the time of the interview are right censored which require special treatment in estimating exposure time. The transition to the second birth is measured as the duration between first and second birth for women who had second births, and for censored women, duration is measured till the date of interview. Using the nine months gestation period, the start of the period at risk is taken to be ten months after the birth of the first child and ended at the birth of the second birth or date of interview, while exit is specified at 120 months after the first birth. These restrictions yielded an effective sample of 5062 for Kenya and 3681 for Ghana.

An Accelerated Failure Time (AFT) model is used to model the effects of infant mortality on the timing of second births. A major assumption of parametric models is that the underlying timing function follows some known mathematical distribution, and that the specified time-dependent distribution is the right one for the event under study. AFT models thus encompass a variety of sub models that differ in the assumed distribution of the timing function. Several models were explored in this study, and a log-normal distribution which assumes that the log of the timing function follows a normal distribution was chosen. This distribution is suitable for duration models with non-monotonic hazards that increase initially and decrease after some period. The choice of the log-normal distribution in this study derives from its suitability in substantive theory (Gyimah 2005; Richards 1983; Trussell and Richards 1985) and also from empirical methods (Table 1) for discriminating between different distributions. For non-nested parametric models, the preferred model is the one with the lowest value of the Akaike Information Criterion (AIC) defined as $-2LogL+2(k+c+1)$ where $k$ is the number of covariates and $c$ the number of model specific distributional parameters.

| Distribution | Hazard shape | Log Likelihood | AIC |
|---|---|---|---|
| **Kenya DHS 2003** | | | |
| Exponential | constant | -4202.63 | 8451.26 |
| Weibull | monotone | -4112.12 | 8272.24 |
| Gompertz | monotone | -4189.99 | 8427.98 |
| Log Normal | variable | -3891.15 | 7830.30 |
| **Ghana DHS 2003** | | | |
| Exponential | constant | -3191.58 | 6425.16 |
| Weibull | monotone | -3005.05 | 6054.10 |
| Gompertz | monotone | -3120.87 | 6285.74 |
| Log Normal | variable | -2791.39 | 5626.78 |

The log-normal hazard (ht), survival (st), and density (ft) functions are defined as:

$$h(t)=\frac{\frac{1}{t\sigma\sqrt{2\pi}}\exp\left[\frac{-1}{2\sigma^2}\left\{\ln(t)-u\right\}^2\right]}{1-\Phi\left\{\frac{\ln(t)-u}{\sigma}\right\}},$$

survival $S(t)=1-\Phi\left\{\frac{\ln(t)-u}{\sigma}\right\}$

and density $f(t)=\frac{1}{t\sigma\sqrt{2\pi}}\exp\left[\frac{-1}{2\sigma^2}\left\{\ln(t)-u\right\}^2\right]$

where, $\Phi(z)$ is the standard normal cumulative distribution function, $\sigma$ is the standard deviation of the normal distribution, and $u$ is the mean. The model is implemented by setting $u=x\beta$, where $x$ is the covariate and $\beta$ is the coefficient vector. The standard deviation $\sigma$ is an ancillary parameter to be estimated from the data. The focus of parametric models is primarily the timing function where positive coefficients imply longer timing (experiences the event later) and negative coefficients connote early timing (experiences the event sooner). When the coefficients are suitably transformed by exponentiation ($e^{\beta}$), they can be interpreted as *time ratios*. In the present application, time ratios greater than one indicate a longer timing of the second birth while the opposite holds if the time ratio is less than one.

On the basis of the substantive arguments outlined earlier, unobserved heterogeneity is explicitly introduced via frailty models. The frailty $\alpha$ is a random positive quantity and, for the purposes of model identifiability, is assumed to have a mean of one and variance $\theta$ (StataCorp, 2007). Frailty was introduced as an unobservable multiplicative $\alpha$ effect on the hazard function such that $h(t \mid \alpha) = \alpha h(t)$, where $h(t)$ is a non-frailty hazard function of the log-normal model described above. Given frailty, the survival function becomes:

$$S(t \mid \alpha) = \exp\left\{ - \int_0^t h(u \mid \alpha) du \right\} = \exp\left\{ -\alpha \int_0^t \frac{f(u)}{S(u)} du \right\} = \{S(t)\}^{\alpha}$$

where $S(t)$ is the survival function that corresponds to $h(t)$. Also, the density function given frailty, $f(t \mid \alpha) = -S'(t \mid \alpha) = \alpha f(t)\{S(t)\}^{\alpha - 1}$ where $f(t)$ is the probability density function that corresponds to $h(t)$ and $S(t)$. Since $\alpha$ is unobservable, it needs to be integrated out of $f(t|\alpha)$.

Let $g(\alpha)$ be the probability density function of $\alpha$, then $f_{\theta}(t) = \int_0^{\infty} f(t \mid \alpha) g(\alpha) d(\alpha) = \int_0^{\infty} \alpha f(t) \{S(t)\}^{\alpha-1} g(\alpha) d\alpha$ which yields the survival function of frailty model as $S_{\theta}(t) = 1 - \int_0^t f_{\theta}(u) du.$

The choice of the distribution function for unobserved heterogeneity has been the subject of considerable discussion (Heckman and Singer 1982, 1984; Jenkins, 2005; Montgomery and Trussell 1986; Trussel and Richards 1985). The literature, however, suggests two major approaches; a parametric approach with well defined distribution (usually the gamma) for the error term (Clayton 1978; Congdon 1988; Flinn and Heckman, 1982; Guo and Rodriguez 1992; Sastry 1997; Vaupel et al. 1979), and a non-parametric approach that only assumes the existence of finite set of values known as 'support points' for the error component (Heckman and Singer, 1984; Defo 1998; Jenkins, 2005). For convenience, researchers frequently choose parametric representations of frailty that are mathematically tractable. In this paper, the parametric approach was adopted by assuming that frailty is distributed over individuals as gamma $g(\alpha)$. The advantages of gamma distributed frailty, as Sastry (1997) points out, are its flexible shape and mathematical tractability. The gamma distribution with the following probability distribution function was assumed for frailty $g(x) = \dfrac{x^{a-1} e^{-x/b}}{\Gamma(a) b^{a}}$.

To understand the conditions under which the results are likely to be sensitive to the assumption of gamma-distributed frailty, the analysis was also performed under inverse-Gaussian distributed frailty. The results were however, found to be less sensitive to the choice of parametric representation of frailty. As a result, the effect of distributional assumption of frailty on parameter estimates is likely to be minor if not negligible. In sum, the frailty approach discussed is a mixture model in that the conditional distribution is described by the log-normal hazard while the error distribution is described by the gamma.

**Empirical Results**

The characteristics of the respondents associated with second births are presented in Table 2. There is an almost equal proportion of women in both Ghana and Kenya with a second birth, but the median time to the second birth is longer in Ghana (39.6 months) than in Kenya (32.3 months). About 10 percent of the women in Ghana experienced the death of the first child as infants compared with 8 percent among women in Kenya. Although a higher percentage of women in Ghana have no education compared with those in Kenya, the percent with at least secondary education is higher in Ghana. While both countries have large rural populations, the percentage of urban residents is slightly higher in Ghana. Christianity seems to be the predominant religion in both countries although Ghana displays greater diversity in religious affiliation. Christianity, for example, accounts for about 83 percent of all religions in Kenya compared with 69 percent in Ghana. The distribution of women by the type of marital union shows the dominance of polygyny among Ghanaian women where almost a quarter of the sample are in such unions. There is also evidence of early marriage and early birth patterns in both countries consistent with the trend in sub-Saharan Africa (see Westoff, 2002).

[Table 2 about here]

Table 3 examines the median time to the second birth and a log rank test that assesses the equality of survivorship functions across groups. In all cases, the significance of the chi-square values and the associated log rank test indicates a rejection of the null hypothesis that the survivorship functions are the same. For the primary independent variable, Table 3 suggests that the median time to the second birth for women who lost their first child is about 23 months compared with 32 months for women whose first child survived to age one in Kenya. In Ghana, the median time is about 27 months for women whose first child died in infancy compared with 38 months for those whose child survived. Table 3 also indicates significant associations of the control variables with the timing of the second birth. Among the Kenyan sample, for example, the transition time to the second birth is about 38 months for highly educated women compared with 27 months for those with no education. For the Ghanaian sample, the median time is 41 months among women with at least secondary education compared with 35 months among those with no education.

[Table 3 about here]

For the multivariate results, two models are estimated for each country; the standard log normal model and a frailty model that allows unobserved heterogeneity introduced as gamma. Table 4 presents the results of the standard log normal model on the timing of second births and the observed covariates for Kenya and Ghana. The respective log likelihoods and the LR chi-square tests suggest the overall model as statistically significant. The time ratios (exp b) associated with the death of the first child in both samples indicate that the transition to the second birth is significantly faster for women who lost the first child in infancy. For the Kenya sample in particular, the death of the first child in infancy reduces the time to the second birth by 26 percent [(1-0.73)*100] than if the child survives. Substantively, these results are consistent with the physiological and behavioral replacement theses regarding the death of infant children on subsequent fertility. As Gyimah and Rajulton (2003, 2004) have argued, the short-term effect of infant death operates mainly through the cessation of breastfeeding which triggers the resumption of menses and ovulation while replacement effect in its usual rendition refers to deliberate and conscious effort a couple makes to replace dead children.

Of the demographic variables, the results suggest that younger women have a longer transition time to the second birth in both countries. Of the socio-economic factors, higher level of education is associated with longer timing in Ghana although the effects are not statistically significant for Kenya. The inverse association between education and timing of births in the case of Ghana has been explained through a myriad of proximate, economic and cultural factors (Gyimah, Maxim and White 2005). Co-residence with a spouse is significantly associated with earlier timing of the second birth in both countries. Compared with women whose husbands live elsewhere, the time to second birth is 20 percent and 10 percent shorter among women who live with their husbands in Kenya and Ghana respectively. Some significant ethnic differences are also apparent. In Ghana, the *Ga-Adangbe* and *Mole-Dagbani* women have significantly longer times and thus, a lower risk of second birth than *Akan* women. The estimated time ratios suggest the transition time to the second birth is 9 percent longer among the *Ga-Adangbes* and 11 percent longer among the *Mole-Dagbanis* compared with *Akan* women. Among Kenyan women, there is evidence of shorter transition times and thus, a higher risk of second birth for all ethnic groups compared with the *Kikuyu.* Also, women in polygynous unions tend to associate with longer timing of the second birth in Kenya, although the effects are not statistically significant for Ghana.

[Table 4 about here]

But how valid are results of the standard log normal hazard model? As discussed earlier, there is concern that the results of hazard analyses that do not correct for unobserved heterogeneity may be misleading and biased. To test the robustness of the findings described in the preceding section, the models were re-estimated with gamma frailty. The methodological concern here is what happens to the regressors when a gamma term is introduced for unobserved heterogeneity. Table 5 presents the results of the frailty models and the respective frailty variances are reported as theta. For both countries, the null hypothesis that theta=0 can be rejected, signifying that unobserved heterogeneity on the timing of second birth is not negligible. The significance of frailty suggests that the risk of the second birth is not only a function of the observed covariates but also of frailty. The likelihood ratio and the improvements in the log likelihoods suggest preference of the frailty models over the standard AFT models presented in Table 4.

[Table 5 about here]

Focusing on Table 5 in particular, frailty has the expected effects on parameter estimates. Comparing the results in Tables 4 and 5, some significant differences are noticeable. Although the directions of the covariates are consistent in both models, two contrasting patterns on their magnitudes are quite visible. Consistent with prior research (e.g., Jenkins, 1997), the results demonstrate that the standard log- normal models underestimate the degree of positive duration dependence (negative coefficients) but overestimate the degree of negative duration dependence (positive coefficients). That is, positive coefficients become weakened in the frailty models while the reverse is true for regressors with negative associations. For the Kenyan sample, the overestimation of the degree of negative dependence is evident on covariates such as age at survey, current residence, age at first marriage, and type of marital union. For example, the coefficient associated with urban residence reduces by about 50% from 0.08 in the standard AFT model to 0.04 in the frailty model. Similarly, the coefficient associated with age at first marriage attenuates by 35% in the two models. Conversely, there is an underestimation of the degree of positive dependence (negative coefficients) on covariates such as survival status of child, and ethnicity such that their effects become more pronounced in the frailty model. For example, the coefficient associated with the death of the first child increases by about 10% in the two models for Kenya, suggesting a faster pace of second birth in the frailty models.

In Figure 1, we present a graph of the predicted hazards from the standard and frailty models. It is worth noting again that the predicted hazard in the standard model is built on the assumption that all women in the sample are identical other than the theoretically relevant covariates in the model. Such an assumption is restrictive and may seldom be fulfilled in practice. If women differ in unobserved ways inherent in their probability of experiencing the second birth as argued earlier, then the predicted hazard from the standard model may be biased. In examining the graphs, there is evidence that the hazard of second birth from the standard model displays a more gentle rise and fall compared with that of the frailty model. Given the significance of frailty, it is likely that with time, the more frail women fail leaving a more homogenous robust population, hence the faster rise and fall of the hazard in the frailty models. Thus, the standard model overestimates negative dependence because of selection effects as high risk women fail faster and the survivors are increasingly drawn from a low risk group.

[Figure 1]

**Conclusion**

This paper examines the usefulness of frailty models with application to data from sub-Saharan Africa. Log normal AFT models chosen on the basis of substantive theory were used to examine the transition to the second birth in a multivariate context. Concerned about the effects of unobserved heterogeneity in hazard models, frailty models were estimated for the transition to the second birth by including a gamma term for unobserved heterogeneity. A general observation was that the models that accounted for unobserved heterogeneity fitted with respect to overall model fitness than standard models.

For both countries, the test statistic on the frailty variance rejects the null that $\theta=0$, implying that parameter estimates in the standard AFT models are biased. Consistent with prior research, there was evidence that the standard models underestimated the degree of positive dependence and overestimated the degree of negative dependence. Overall, the analysis presented here has several strengths. Substantively, the results provide considerable empirical evidence that infant deaths indeed have significant associations with the timing of subsequent births. Women whose index child died in the first year of life were found to have significantly shorter transitional times, and a faster pace of subsequent birth than women whose first children survived the first year of life. Besides the infant deaths, other covariates were found to have significant association with the timing of second births. The most important among these were age, ages at first marriage and at first birth, education, current place of residence, and ethnicity.Younger women, current urban residents and the highly educated were found to have longer transition times to second birth, and thus a significantly lower risk of second birth. Methodologically, the results underscore the need to explicitly account for unobserved heterogeneity in event history models when there are substantive reasons for it. From a policy perspective, it is obvious that improvement in child survival programs would significantly improve the reproductive life of women by reducing fertility through both biological and behavioural processes.

References

Aalen, OO. (1994). Effects of frailty in survival analysis. *Statistical Methods in Medical Research* 3:227-243.

Addai, I. (1999). Ethnicity and contraceptive use in sub-Saharan Africa: The case of Ghana. *Journal of Biosocial Science*, 31:105-120.

Bachieri, A. and A. Hinde (2007). The proximate determinants of fertility and birth intervals in Egypt: An application of calendar data. Demographic Research, 16(3): 59-96.

Baker, M. and A. Melino. Duration dependence and nonparametric heterogeneity: A Monte Carlo study. Journal of Econometrics 96(2): 357-393.

Blossfeld, H-P. and A. Hamerle (1992). Unobserved heterogeneity in event history models. Quality and Quantity, 26(2): 157-168.

Clayton, D. 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 65: 141-151.

Cochrane, S. (1983). Effects of education and urbanization on fertility. In R.A. Bulatao and R.D. Lee, eds., *Determinants of Fertility in Developing Countries. A Summary of Knowledge, Part B*. New York: Academic Press.

Congdon, P (1988). Occupation mobility in the Irish Republic. In M.D. Uncles, ed., *Longitudinal Data Analysis: Methods and Applications.*

Defo, Kuate, B. (1998). Fertility response to infant and child mortality in Africa with special reference to Cameroon. In Mark R. Montgomery and Barney Cohen, eds., *From Death to Birth: Mortality Decline and Reproductive Change*. Washington, D.C.: National Academy Press.

Flinn, C.J. and J.J. Heckman (1981). Models for the analysis of labor force dynamics. *Advances in Econometrics* 1:35-95.

Guiterrez, R.G. (2002). Parametric frailty and shared frailty survival models. The Stata Journal 2(1): 22-44.

Guo, Guang and German Rodriguez (1992). Estimating a multivariate proportional hazards model for clustered data using EM Algorithm, with an application to child survival in Guatemala. *Journal of the American Statistical Association* 87, 420:969-976.

Gyimah, S. O. (2003). A cohort analysis of timing of the first birth and cumulative fertility in Ghana. *Population Research and Policy Review,* 22 (3):251-266.

Gyimah, S. O. and Rajulton Fernando (2003). The physiological effects of infant death on fertility: A comparative analysis of DHS data from Ghana and Kenya. *Social Biology,* 49(1-2):44-57.

Gyimah, S. O. and Rajulton Fernando (2004). Intentional replacement of dead children in sub-Saharan Africa: Evidence from Ghana and Kenya. *Canadian Studies in Population*, 34(1)33-53.

Gyimah, S. O. (2005). The dynamics of timing and spacing of births in Ghana. *Journal of Comparative Family Studies,* 36(1): 41-60.

Gyimah, S. O., Paul Maxim and Jerry White (2005). Cohort perspectives on educational attainment and timing of parenthood in Ghana. *Journal of Family and Economic Issues,* 26(1) 123-142.

Heckman, J. and B. Singer (1982). Population heterogeneity in demographic models. In K. Land and A. Rogers, eds., *Multidimensional Mathematical Demography*. New York: Academic Press.

Heckman, J. and B. Singer (1984). A method of minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52:271-320.

Heckman, J., J. Hotz and J. Walker (1985). New evidence on the timing and spacing of births. *The American Economic Review, Papers and Proceedings* 75:179-184.

Hougaard, P. (2000). Analysis of Multivariate Data. Springer

Jenkins, S.P. (2005) Survival Analysis. Available at http://www.iser.essex.ac.uk/teaching/degree/stephenj/ec968/pdfs/ec968lnotesv6.pdf

Lancaster, T. (1990). *The Econometric Analysis of Transition Data.* New York: Cambridge University Press.

Manda, S.O.M., Meyer, R. (2005): Age at first marriage in Malawi: a Bayesian multilevel analysis using a discrete time-to-event model. Journal of the Royal Statistical Society. A. 168, 439-455.

Mark R. Montgomery and Barney Cohen, eds. (1998). *From Death to Birth: Mortality Decline and Reproductive Change*. Washington, D.C.: National Academy Press.

Martin, T. (1995). Women's education and fertility: results from 26 Demographic and Health Surveys. *Studies in Family Planning* 26:187-202.

Palloni, A. and H. Rafalimanana (1999). The effects of infant mortality on fertility revisited: New evidence from Latin America. *Demography* 36(1):41-58.

Richards, T. (1983). *Comparative analysis of fertility, breastfeeding and contraception: A dynamic model*. Report 22. Washington: Panel on Fertility Determinants, National Academy of Sciences

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.

StataCorp (2005). *Stata Statistical Software: Release 9.0.* College Station, TX: Stata

Trussell, J., and German Rodriguez (1990). Heterogeneity in demographic research. In

Trussell, J. and T. Richards (1985). Correcting for unmeasured heterogeneity in hazard models using Heckman-Singer procedure. In N.B. Tuma, ed., *Sociological Methodology.*

Trussell, James, E. van de Walle and F. van de Walle (1989). Norms and behavior in Burkinabe fertility. *Population Studies* 43 (3):429-454.

Trussell, James, Linda Martin, Robert Feldman, and James A. Palmore (1985). Determinants of birth interval length in the Philippines: A Hazard Model Analysis. *Demography* 22 (2):145-168.

Van den Berg, G.J. (2001). Duration models: specification, identification and multiple durations. In Handbook of Econometrics Vol. 5 (Eds.) J.J. Heckman and E. Leamer. North-Holland, Amsterdam.

Vaupel, J., K. Manton and E. Stallard (1979). Impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16(3):439-454.

Westoff, C.F. (1992). Age at Marriage, Age at first Birth and Fertility in Africa. *World Bank Technical Paper Number 169.* Washington, D.C.: The World Bank.

Table 2: Descriptive Statistics of Covariates, Kenya and Ghana

| | Kenya (%) | Ghana (%) |
|---|---|---|
| **Survival status of First child** | | |
| Died in the first year | 7.86 | 10.06 |
| Survived | 92.14 | 89.94 |
| **Age of mother** | | |
| under 25 years | 22.31 | 15.34 |
| 25-34 years | 40.10 | 39.68 |
| 45-49 years | 28.73 | 32.23 |
| 35-44 years | 8.95 | 12.75 |
| **Level of education** | | |
| None | 8.69 | 43.48 |
| Primary | 19.68 | 19.37 |
| Secondary (and above Ghana) | 51.64 | 37.14 |
| Higher | 19.99 | - |
| **Place of current residence** | | |
| Urban | 30.16 | 34.38 |
| Rural | 69.84 | 65.28 |
| **Household wealth** | | |
| Rich-Richest | 46.01 | 33.86 |
| Otherwise | 54.00 | 66.14 |
| **Age at first marriage** | | |
| Married afte age 20 | 37.31 | 33.99 |
| Married as a teen | 62.69 | 66.01 |
| **Age at first birth** | | |
| Had first birth after age 25 | 4.61 | 7.06 |
| Had first birth before 25 | 95.39 | 92.94 |
| **Ethncity** | | |
| Luhya | 15.01 | Akan: 41.19 |
| Kalenjin | 8.21 | Ga: 7.38 |
| Luo | 10.61 | Ewe: 11.99 |
| Kamba | 9.05 | Guan/other: 9.53 |
| Somali | 8.43 | Dagbani: 29.92 |
| Other groups | 26.23 | - |
| Kikuyu | 22.47 | - |
| **Religious affiliation** | | |
| Moslem | 14.12 | 18.76 |
| None/Other | 3.03 | 6.86 |
| Christian | 82.85 | 69.37 |
| Traditional | - | 5.02 |
| **Type of marital union** | | |
| Not currently married | 15.25 | 11.59 |
| Polygamous union | 16.53 | 23.15 |
| Monogamous union | 68.22 | 65.26 |
| **Residential pattern of spouse** | | |
| Not applicable | 15.21 | 11.85 |
| co-residence with husband | 67.01 | 65.75 |
| Husband lives elsewhere | 17.77 | 22.4 |
| Median survival time (months) | 32.29 | 39.63 |
| **Censored observations** | 16.92 | 17.87 |

Table 3: Median survial time to the second birth (months), Ghana and Kenya 2003

| | Kenya | | Ghana | |
|---|---|---|---|---|
| | **Median** | **Log rank chi-sqaure test** | **Median** | **Log rank chi-sqaure test** |
| **Survival status of the index child** | | 79.18*** | | 96.69*** |
| Died in the first year | 23 | | 27 | |
| Survived | 32 | | 38 | |
| **Age of mother** | | 57.46*** | | 16.30** |
| under 25 years | 33 | | 39 | |
| 25-34 years | 33 | | 38 | |
| 35-44 years | 27 | | 37 | |
| 45-49 years | 29 | | 35 | |
| **Level of education** | | 116.78*** | | 106.99*** |
| None | 27 | | 35 | |
| Primary | 31 | | 36 | |
| Secondary and above | 35 | | 41 | |
| Higher | 38 | | - | |
| **Place of current residence** | | 86.17*** | | 73.48*** |
| Urban | 33 | | 40 | |
| Rural | 30 | | 36 | |
| **Household wealth** | | 122.42*** | | 95.94*** |
| Rich-Richest | 33 | | 41 | |
| Otherwise | 29 | | 36 | |
| **Age at first marriage** | | 161.83*** | | 7.26*** |
| Married afte age 20 | 35 | | 37 | |
| Married as a teen | 29 | | 38 | |
| **Age at first birth** | | 13.73*** | | 2.1 |
| Had first birth after age 25 | 36 | | 39 | |
| Had first birth before age 25 | 31 | | 37 | |
| **Ethncity** | | 161.37*** | | 30.39*** |
| Luhya | 30 | | | |
| Kalenjin | 29 | | Akan: 37 | |
| Luo | 30 | | Ga: 41 | |
| Kamba | 33 | | Ewe: 38 | |
| Somali | 25 | | Guan/other: 38 | |
| Other groups | 32 | | Dagbani: 37 | |
| Kikuyu | 34 | | | |
| **Religious affiliation** | | 42.35*** | | 39.56*** |
| Moslem | 26 | | 38 | |
| None/Other | 29 | | 34 | |
| Christian | 32 | | 38 | |
| Traditional | - | | 34 | |
| **Type of marital union** | | 25.09*** | | 46.64*** |
| Not currently married | 33 | | 40 | |
| Polygamous union | 31 | | 36 | |
| Monogamous union | 30 | | 38 | |
| **Residential pattern of spouse** | | 11.56*** | | 68.94*** |
| co-residence with husband | 30 | | 36 | |
| Husband lives elsewhere | 32 | | 40 | |

Notes: Statistical significance: ***p<0.001. **p<0.01. *p<0.05. !p<0.10.

Table 4: A log normal model without unobserved heterogeneity on the transition to second births, Kenya and Ghana

| | Kenya | | Ghana | |
|---|---|---|---|---|
| | **b** | **Exp (b)** | **b** | **Exp (b)** |
| **Survival status of the index child** | | | | |
| Died in the first year | -0.30*** | 0.74 | -0.35*** | 0.70 |
| Survived (reference) | - | - | - | - |
| **Age of mother** | | | | |
| under 25 years | 0.24*** | 1.28 | 0.14*** | 1.15 |
| 25-34 years | 0.15*** | 1.16 | 0.06** | 1.06 |
| 45-49 years | -0.06 | 0.95 | -0.01 | 0.99 |
| 35-44 years (reference) | - | - | - | - |
| **Level of education** | | | | |
| Primary | -0.05 | 0.95 | 0.01 | 1.01 |
| Secondary and above | 0.01 | 1.01 | 0.15*** | 1.16 |
| Higher | -0.02 | 0.98 | - | - |
| None (reference) | - | - | - | - |
| **Place of current residence** | | | | |
| Urban | 0.08** | 1.09 | 0.03 | 1.03 |
| Rural (reference) | - | - | - | - |
| **Household wealth** | | | | |
| Rich-Richest | 0.04 | 1.04 | 0.09*** | 1.09 |
| Otherwise (reference) | - | - | - | - |
| **Age at first marriage** | | | | |
| Married afte age 20 | 0.19*** | 1.21*** | 0.04 | 1.04 |
| Married as a teen (reference) | - | - | - | - |
| **Age at first birth** | | | | |
| Had first birth after age 25 | 0.05 | 1.05 | 0.01 | 1.01 |
| Had first birth before 25 (reference) | - | - | - | - |
| **Ethncity** | | | | |
| Luhya | -0.15*** | 0.86 | Ga: 0.09* | 1.09 |
| Kalenjin | -0.17*** | 0.84 | Ewe: 0.05 | 1.05 |
| Luo | -0.10** | 0.90 | Guan/other: 0.08 | 1.08 |
| Kamba | 0.02 | 1.02 | Dagbani: 0.10*** | 1.11 |
| Somali | -0.36*** | 0.70 | Akan (reference) | - |
| Other groups | -0.06* | 0.94 | | |
| Kikuyu (reference) | - | - | | |
| **Religious affiliation** | | | | |
| Moslem | -0.04 | 0.95 | 0 | 1.00 |
| None/Other | 0.09 | 0.94 | -0.05 | 0.95 |
| Christian (reference) | - | - | - | - |
| Traditional | - | - | -0.04 | |
| **Type of marital union** | | | | |
| Not currently married | 0.14*** | 1.15 | 0.07 | 1.07 |
| Polygamous union | 0.10*** | 1.10 | -0.1 | 0.90 |
| Monogamous union (reference) | - | - | - | - |
| **Had co-residence with husband** | -0.20 | 0.80 | -0.10*** | 0.90 |
| CONSTANT | 3.36*** | - | 3.57*** | - |
| SIGMA | 0.59*** | 0.59*** | 0.54*** | 0.54*** |
| NUMBER OF FAILURES | 4329 | | 3118 | |
| LOG LIKELIHOOD | -3891.12 | | -2791.38 | |
| LIKELIHOOD RATIO CHI-SQUARE | 523 | | 308 | |
| PROB > LR CHI-SQUARE | 0.000 | | 0.000 | |

Notes: Statistical significance: ***p<0.001. **p<0.01. *p<0.05. !p<0.10.

Table 5: A log normal model on the timing of second births with Frailty, Kenya and Ghana

| | Kenya | | Ghana | |
|---|---|---|---|---|
| | **b** | **Exp (b)** | **b** | **Exp (b)** |
| **Survival status of the index child** | | | | |
| Died in the first year | -0.33*** | 0.71 | -0.36*** | 0.70 |
| Survived (reference) | - | - | - | - |
| **Age of mother** | | | | |
| under 25 years | 0.22*** | 1.24 | 0.14*** | 1.15 |
| 25-34 years | 0.13*** | 1.14 | 0.06** | 1.06 |
| 45-49 years | -0.07* | 0.93 | -0.01 | 0.99 |
| 35-44 years (reference) | - | - | - | - |
| **Level of education** | | | | |
| Primary | -0.03 | 0.97 | 0.01 | 1.01 |
| Secondary and above | 0.01 | 1.01 | 0.11*** | 1.12 |
| Higher | -0.02 | 0.98 | - | - |
| None (reference) | - | - | - | - |
| **Place of current residence** | | | | |
| Urban | 0.04 | 1.04 | 0.01 | 1.01 |
| Rural (reference) | - | - | - | - |
| **Household wealth** | . | | | |
| Rich-Richest | 0.01 | 1.01 | 0.06 | 1.06 |
| Otherwise (reference) | - | - | - | - |
| **Age at first marriage** | | | | |
| Married afte age 20 | 0.13*** | 1.14 | 0.02 | 1.02 |
| Married as a teen (reference) | - | - | - | - |
| **Age at first birth** | | | | |
| Had first birth after age 25 | 0.04 | 1.04 | -0.01 | 0.99 |
| Had first birth before age 25 (reference) | - | - | - | - |
| **Ethncity** | | | | |
| Luhya | -0.13*** | 0.88 | Ga: 0.06 | 1.06 |
| Kalenjin | -0.13*** | 0.88 | Ewe: 0.02 | 1.03 |
| Luo | -0.09* | 0.91 | Guan/other: 0.10** | 1.11 |
| Kamba | 0.03 | 1.02 | Dagbani: 0.12*** | 1.12 |
| Somali | -0.27*** | 0.77 | Akan (reference) | - |
| Other groups | -0.58* | 0.94 | | |
| Kikuyu (reference) | - | - | | |
| **Religious affiliation** | | | | |
| Moslem | -0.08* | 0.92 | -0.04 | 0.96 |
| None/Other | -0.06 | 0.94 | -0.04 | 0.96 |
| Traditional | - | - | -0.02 | 0.98 |
| Christian (reference) | - | - | - | - |
| **Type of marital union** | | | | |
| Not currently married | 0.07* | 1.07 | 0.03 | 1.03 |
| Polygamous union | 0.07** | 1.07 | 0.02 | 1.02 |
| Monogamous union (reference) | - | - | - | - |
| **Had co-residence with husband** | -0.02 | 0.97 | -0.07 | 0.93 |
| CONSTANT | 3.43*** | | 3.49*** | |
| SIGMA | 0.43*** | | 0.41*** | |
| Frailty Variance | 0.41*** | | 0.43*** | |
| NUMBER OF FAILURES | 4329 | | 3118 | |
| LOG LIKELIHOOD | -3789.12 | | -2716.35 | |
| LIKELIHOOD RATIO CHI-SQUARE | 468.01 | | 274 | |
| PROB > LR CHI-SQUARE | 0.000 | | 0.000 | |
| LR test that theta=0 | 204.04*** | | 150.06*** | |

Figure 1: Predicted Hazard with and without Frailty, Kenya and Ghana