UNIVERSITY OF SOUTHAMPTON

Multilevel Analysis of Health and Family Planning Data

by

Fiona Alison Steele

Doctor of Philosophy

Department of Social Statistics

Faculty of Social Sciences

September 1996

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL SCIENCES
SOCIAL STATISTICS

Doctor of Philosophy

MULTILEVEL ANALYSIS OF HEALTH AND FAMILY PLANNING DATA

by Fiona Alison Steele

Many populations in the social sciences have a hierarchical structure. For example, individuals are often nested within communities which themselves lie within larger units. In such cases, the characteristics of the context in which a person lives are likely to influence their behaviour and thus responses for individuals in the same community will tend to be correlated. Hierarchical data structures can also arise in longitudinal studies where observations over time are nested within an individual, and responses for the same individual may be correlated. One example of longitudinal data is an event history, where an individual is observed until the event of interest occurs or the observation period ends. Multilevel modelling techniques, which take into account these intra-unit correlations, have been developed to analyse hierarchical data. A multilevel approach can also be used as a convenient way of allowing for the effects of omitted covariates, or unobserved heterogeneity, in discrete-time event history models.

In this thesis, multilevel modelling techniques are used to analyse a variety of hierarchical population structures in the areas of health and family planning. Four empirical studies are presented. In the first study, a multilevel multinomial model is used to analyse variations in contraceptive choice in Bangladesh between districts, and within districts between clusters. The analysis shows that a large proportion of the district-level variation in modern method use can be explained by differences in religious practice and literacy. Another study uses a two-level event history model to allow for unobserved heterogeneity in women's risks of contraceptive discontinuation in China. This is extended to a four-level model to analyse the extent of extravariation at the district, cluster and woman level in contraceptive method switching in Bangladesh. The results from these studies provide strong evidence of unobserved heterogeneity between women in contraceptive behaviour.

Multilevel models are also applied in the area of child health to study immunisation uptake in rural Bangladesh. The results show that even after controlling for a range of child-, parental- and household-level characteristics, there remains substantial variation in immunisation rates due to unobserved factors at the household and village level.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank Ian Diamond for supervising this work and for encouraging me to undertake PhD study. Thanks are also due to the other members of the ALCD team, especially Danius Michaelides for his invaluable computing support.

I am grateful to Sajeda Amin for her advice on some of the demographic aspects of this thesis and for her hospitality during my visits to Bangladesh.

The analysis of child immunisation uptake in Chapter 5 would not have been possible without the co-operation of the Bangladesh Field Office of Save the Children (USA). I would like to thank them for allowing me to use their data, their helpful feedback on an initial draft of the work and for organising my visits to the field. The data are available through funding in part to Save the Children (USA) from the United States Agency for International Development (AID/FHA/PHA Child Survival Cooperative Agreements (PDC-0502-A-00-5095 and FAO-0500-A-00-2034-00 and US-AID Cooperative Agreement PDC-0158-A-00-1058-00). The views expressed herein do not necessarily reflect the views of Save the Children or of its funding agencies.

This work was supported by an ESRC research studentship (award R00429334331) under the ESRC Analysis of Large and Complex Data Sets Programme.

Chapter 1

Introduction: Hierarchical Structures and Contextual Effects

1.1 Hierarchical Structures

Many populations in the social sciences have an inherent hierarchical or nested structure. People are nested within local communities which are themselves nested within countries; children are nested within families; pupils are nested within classes which are in turn nested within schools. Where hierarchies exist, characteristics of the context in which a person lives are likely to influence their behaviour. For example, one would expect individuals selected from the same community to share similar attitudes and therefore show behavioural patterns that are more alike than those of individuals selected randomly from the population at large. Children in the same family share certain genetic characteristics and environmental conditions which may lead them to experience similar risks of illness or mortality. Similarly, since pupils in the same class are taught by the same teacher, one would expect the performance of children with similar background characteristics and academic ability to be more alike if they are from the same class than if they were from different classes. Researchers have suspected the importance of contextual effects, or factors operating at one or more levels of aggregation, on individual behaviour for many years. Of particular interest is the extent to which people's behaviour is affected by the social groups to which they belong and how the characteristics of these groups interact with individual-level characteristics. It was not until the early 1980s, however, that appropriate statistical techniques were developed to enable these and other important questions to be addressed. Standard regression techniques ignore hierarchical structures and the clustering of outcomes which can occur at each level. Multilevel approaches model the complex correlation structure between outcomes at each level and as such provide a more faithful reflection of the true nature of the social systems to which individuals belong.

Hierarchical structures can also arise from the study design. In longitudinal studies, observations are made over time on the same individual. This gives rise to a two-level hierarchical structure with repeated observations (level 1) nested within individuals (level 2). Longitudinal data are now collected routinely in developing countries in the form of retrospective data on birth and contraceptive use histories. In Britain, a number of panel studies, such as the National Child Development Study or the British Household Panel Study, follow a sample of individuals through time and collect information in successive waves of the survey. Other examples of prospective longitudinal data are those collected in surveillance systems operated by the International Centre for Diarrhoeal Disease Research and Save the Children USA in rural Bangladesh to monitor the progress of development programmes. Such studies are a rich source of data which can now be analysed using multilevel techniques to take into account the correlations between successive outcomes over time for a given individual.

Another way in which hierarchical structures can arise is as a result of the survey design which in many cases reflects the structure of the population: typically households are selected from clusters or primary sampling units which have themselves been selected from larger administrative areas. Thus the hierarchical structure is an artefact of the multistage sampling techniques which have been used as a convenient means of selecting the sample. In such cases, the use of standard regression techniques, which assume that individual observations are independent and identically distributed, is equivalent to treating individuals as if they were selected using simple random sampling. This problem has been well known in the sample survey literature for many years and appropriate methodology has been developed to take into account clustering or 'design effects' (Skinner et al. 1989). However, while these procedures take proper account of the complex nature of the survey, they usually treat the actual structure of survey data as a nuisance. If the hierarchical structure is of substantive interest, multilevel analysis techniques which model the population structure more explicitly can be used.

1.2 Early Approaches Used to Analyse Contextual Effects in Demography

Analysts used to adopt one of two approaches to analyse hierarchical data: an individual-level analysis using individual data only or an aggregate-level analysis using individual data which have been aggregated to the group level. The individual-level approach remains the most popular way of analysing the determinants of demographic behaviour. However, an individual-level analysis ignores contextual effects altogether and assumes that a person's behaviour can be determined by his or her individual characteristics alone. Such a model implicitly assumes that individuals act independently of each other which is clearly unrealistic in most contexts. Alker (1969) refers to this as the atomistic fallacy. At the other extreme, aggregate-level approaches such as multivariate areal analysis (Hermalin 1979) ignore variability within groups. The ecological fallacy (Robinson 1950) arises when one attempts to make inferences about individual behaviour from an analysis of aggregate data which suppresses within-group variability. Typically, relationships between aggregate responses and aggregate covariates tend to be stronger than relationships between the corresponding individual responses and individual covariates which can make cross-level inferences highly misleading.

It was later recognised that a more complete analysis of individual behaviour should model the effects of characteristics of the individual and of their context simultaneously. During the late 1970s and early 1980s, contextual models came into use which attempted to explain variation in individual-level behaviour in terms of individual and contextual effects. These early models (also called multilevel models in the literature) simply related individual responses to covariates defined at both the individual and aggregate level, using standard regression techniques for single-level data. Interaction terms between individual and contextual variables were often included to examine whether factors operating at higher levels modify or are modified by individual relationships.

In an effort to evaluate the extent to which demographic decisions are influenced by the community or social setting in which people live, the World Fertility Survey (WFS) programme designed and implemented a community module to complement the individual-level data collected in household surveys. Data including measures of accessibility of the community, development indicators and the presence of family planning and health services were collected in rural areas of 17 developing countries. In a WFS seminar in 1983, the results of a number of contextual analyses which utilised these and other community-level data were disseminated (Casterline 1985). Most of these studies sought to examine the joint impact of individual factors and characteristics of the community, such as the availability of family planning services, on fertility and contraceptive use. In general, however, the results were inconsistent and showed little conclusive evidence of contextual effects on individual behaviour.

There are a number of possible explanations for the rather disappointing findings of these early studies of contextual effects. The quality of the WFS community data may have been a contributing factor as it has been suggested that, in general, the community surveys were not well designed or carried out with the same care as the individual surveys (Casterline 1987; Freedman 1985). There were also criticisms of the type of information collected (Bilsborrow 1985; Holt 1985). For example, no data on quality and usage of family planning facilities were gathered; only information on the presence of clinics and other outlets and in some cases their distance from the community was collected. This placed limitations on the contextual analyses which could be performed as only very crude measures of service availability could be incorporated in the models. Further, many of the other factors which might be expected to influence fertility and related behaviour, such as the attitudes of others in the community and local social norms, are not readily measured and therefore only much simplified proxies of these sometimes complex phenomena could be considered in the analyses.

Since the WFS analyses, there have been several studies which have had more success in establishing links between community characteristics and individual behaviour. For example, Entwisle et al. (1984) utilised data from a community survey which was carried out in rural Thailand following a Contraceptive Prevalence Survey (CPS). They found a community-level variable measuring the availability of family planning outlets to have an important influence on the likelihood of contracepting. In a later study using data from another CPS community survey in rural Egypt, Entwisle et al. (1989) found that a number of village-level characteristics had a strong impact on contraceptive behaviour. These included the level of nonagricultural activity in the village economy, the modernisation of agriculture, the level of school participation and, to a lesser extent, the family planning service environment. However, although these studies have been successful in identifying some of the contextual factors which affect individual behaviour, the models used do not have an error term for the community level. Therefore, there is an implicit assumption that all variation between contexts can be explained by a rather limited set of observed community-level variables. As Entwisle et al. (1989) acknowledge in the Egyptian context, it is likely that there remain a number of potential contributing factors which have not been measured, such as the views of the local *imam* concerning contraception and quality (as distinct from quantity) of family planning services. A more realistic approach would be to allow for error at both the individual and community level. In addition, an important advantage of splitting the variation into individual and contextual components is that an estimate of the inter-community variation can be obtained. Using this approach, it is possible to determine the amount of variation that can be explained with observed community-level variables and the proportion that remains unexplained.

1.3 Applications of Multilevel Models in Demography

As described in the previous section, the drawback of the contextual model used in the WFS and other early studies is the assumption that all between-community variation is explained by the community-level variables included in the model. This assumption is likely to be invalid, especially since many contextual factors which influence individual behaviour, such as the attitudes of others in the community, are difficult to measure and are thus likely to remain unobserved. This unexplained component can induce clustering of outcomes among individuals living in the same context. Therefore the assumption of independence between individuals in the contextual model is likely to be invalid. Mason et al. (1983) introduced a hierarchical linear model (also called a multilevel or random effects model) with separate error components for the individual and contextual levels, which takes into account extravariation at both levels of the hierarchy. This method was illustrated in a multilevel analysis of fertility in 15 WFS countries. The multilevel model of the determinants of the number of children ever born included characteristics of the woman (level 1), characteristics of the country (level 2), including the gross national product and a measure of family planning programme effort, and cross-level interactions between the two sets of variables. The linear model was later extended to a hierarchical logistic regression model to allow multilevel analysis of binary response data (Wong and Mason 1985). This methodology was employed to analyse the influence of individual and country effects on ever-use of contraception in the 15 WFS countries used in the earlier fertility analysis (Entwisle et al. 1986; Wong and Mason 1985).

Following the papers by Mason et al. (1983) and Wong and Mason (1985), surprisingly few applications of multilevel modelling appear in the demographic literature. This may have been due in part to the lack of any strong evidence of contextual effects from early studies. It is also possible that in the absence of appropriate software for fitting multilevel models, many researchers may have been discouraged from attempting a full multilevel analysis. Indeed, most studies in demography and sociology which examined contextual effects continued to use standard regression techniques which ignore extravariation at the contextual level. Meanwhile, multilevel models were becoming increasingly popular in the field of educational research. Since the mid-1980s, there have been numerous applications of multilevel techniques to examine school effects on examination results (Aitkin and Longford 1986; Goldstein 1987; Raudenbush and Bryk 1986).

In the early 1990s, multilevel models began to resurface in the demographic literature with applications to familial clustering of child deaths in developing countries. A tendency for child deaths to cluster within families even after controlling for many socioeconomic and biological factors had been noted earlier by several authors (Das Gupta 1990; Pebley and Stupp 1987). These studies used the survival status of siblings as a proxy measure for death clustering due to unobserved family-level characteristics. Guo and Rodriguez (1992) used a multilevel approach and developed a proportional hazards survival model incorporating a family-specific random effect to capture unobserved familial effects on child mortality. Curtis et al. (1991, 1993) and Zenger (1993) used random effects logistic models to explore the extent of familial clustering in neonatal and postneonatal mortality, an approach also adopted by Madise and Diamond (1995) and Curtis and Steele (1996). The results provide strong evidence of death clustering in a number of developing countries. A recent paper by Sastry (1996) uses a three-level proportional hazards model to test for the presence of clustering in Brazil at both the family and community level. He finds that after controlling for characteristics defined at the child, family and community level, there is no evidence of extravariation between communities or between families. In the less developed Northeast region, however, a number of contextual variables are found to have an impact on child mortality risks. These are the type of water supply, the presence of a refuse collection service and public cleaning, sanitation and electricity. Another recent study by Pebley et al. (1996) uses a three-level multilevel logistic model to investigate clustering in other health outcomes, namely prenatal and delivery care and childhood immunisation, also at both the family level and community level. They find a substantial amount of intra-family correlation for all the outcomes considered. In particular, there is strong evidence that families who use or do not seek prenatal care for one pregnancy are very likely to do the same for the next. There is also considerable correlation within communities.

1.4 Applications of Multilevel Models in Event History Analysis

An event history is a longitudinal record of the occurrence of events to an individual over time. In demography, commonly studied events of interest include births, deaths, marriages and discontinuations of contraceptive use. In event history analysis, the time to some event is usually modelled as a function of individual characteristics. The implications of unobserved heterogeneity in such models, that is variation in the risk of event occurrence due to omitted covariates, have long been recognised (Vaupel et al. 1979). As a result, a number of continuous-time survival models which control for this extravariation have been developed (e.g. Blossfeld and Hammerle 1989; Manton et al. 1986). A convenient way of incorporating unobserved heterogeneity into a discrete-time event history model is to use multilevel techniques. To fit a discrete-time model, the data must first be restructured to obtain a binary (if there is only one type of event of interest) response for each time point, indicating whether the event has occurred. Therefore the data have a two-level structure, time units (level 1) nested within individuals (level 2), and the model can be estimated as a multilevel logistic model with a random effect at the individual level.

This discrete-time approach has been used by Egger (1992) to model the length of birth intervals and Davies et al. (1992) to study the time from marriage to a woman's entry into the labour force. An advantage of using multilevel techniques is that the basic model can be extended straightforwardly to account for extravariation at additional hierarchical levels. One obvious such extension is to the analysis of recurrent events. For example, in the above studies, it is likely that a woman will have more than one birth over her reproductive career and individuals may repeatedly move in and out of the labour force. Repeated events can be analysed in a multilevel framework where time units (level 1) are nested within events (level 2) within individuals (level 3). The model can be extended to further levels in order to study, for example, areal variation in the rate of event occurrence.

1.5 An Outline of the Thesis

In this thesis, multilevel techniques for discrete responses are used to analyse hierarchical demographic data. The four studies presented are a series of analyses of family planning and health data which illustrate just some of the considerable potential of multilevel models in these areas of research. The second chapter outlines the fundamentals of the theory of multilevel models for continuous, binary and multinomial response data. Some of the most commonly used estimation procedures are described. Diagnostics, prediction, the estimation of higher-level residuals and their interpretation in multilevel models, and model fitting strategies are also discussed.

The first study, presented in Chapter 3, is an analysis of contraceptive choice in Bangladesh using data from a national fertility survey. A three-level multinomial model is employed to analyse the determinants of current use and the type of method used. Random effects are incorporated in the model to examine inter-district variation, and variation within districts between sampling clusters. After controlling for a range of demographic and socioeconomic characteristics of the woman, district-level variables are added to the model in an attempt to explain some of the between-district variation. A useful feature of multilevel modelling is that the partitioning of the unexplained variation into components for each level in the hierarchy enables the proportion of inter-district variation that can be attributed to these district-level characteristics to be determined.

The other studies all use multilevel discrete-time event history methodology. These techniques and related life table methods are described in Chapter 4. The simplest single-level model for situations where there is only one type of event of interest is presented first. Extensions to multiple kinds of events and repeated events are then discussed. Finally, event history models which incorporate unobserved heterogeneity are described. The estimation of such models can be achieved via the use of multilevel models for discrete response data.

Multilevel discrete-time event history models are first applied in Chapter 5 to analyse the factors affecting child immunisation uptake in several rural regions of rural Bangladesh, using a longitudinal data set collected by Save the Children USA. In this case, there is only one type of event of interest—the completion of a child's immunisation schedule. The main focus of interest is the extent of extravariation at the child, mother, household and village level.

The two remaining analysis chapters use multilevel event history techniques to study two related aspects of contraceptive use dynamics—contraceptive discontinuation and switching. In Chapter 6, a competing risks model is developed to analyse the determinants of contraceptive discontinuation in China by the type of reason for stopping use. Since women may experience more than one discontinuation over the observation period, random effects are incorporated to allow for unobserved heterogeneity between women and for the correlation between the durations of repeated use intervals which may result from this extravariation. A separate random effect is included for each type of discontinuation as the effects of these unobserved woman-level factors may vary by the reason for stopping use. Such a model may be estimated using the multilevel techniques which have been developed for multinomial response data.

The study in Chapter 7 utilises calendar data from the Bangladesh Demographic and Health Survey to analyse changes in contraceptive behaviour among users. After a woman has adopted a method of contraception, she may continue using the same method, switch to another method or abandon use of contraception altogether. Of most interest to family planning policy makers are the factors which prompt a woman to switch from an effective method to an inefficient method or to non-use when she is at risk of experiencing an unintended pregnancy. A competing risks framework is used to examine switching behaviour over the six-year observation period. As in the analysis of discontinuation, there are likely to be unobserved factors that influence use patterns and consequently durations of successive use intervals for the same woman are possibly correlated. Therefore woman-specific random effects for each type of switch are incorporated in the model. In addition, there is the potential for clustering at two higher levels of aggregation: the sampling cluster and district. One might expect women living in the same cluster to exhibit similar patterns of switching behaviour as they have access to the same family planning services and a similar range of method options. There may also be some important district-level characteristics that induce nesting of sampling clusters within districts.

Finally, the main results and substantive conclusions of all four studies are discussed in Chapter 8. In particular, some of the questions which these analyses have raised and ways in which they might be addressed through further work are discussed. Some ideas for future research using multilevel models in demography are also suggested.

Chapter 2

A Review of Multilevel Theory

2.1 Introduction

This chapter provides an overview of the multilevel theory of most relevance to this thesis. More detailed expositions of the theory of multilevel models can be found elsewhere, e.g. Bryk and Raudenbush (1992) or Goldstein (1987,1995). We begin by looking at the problems associated with the single-level approaches commonly used to analyse hierarchical data before the development of multilevel modelling techniques. These are contrasted with a simple two-level model for continuous response data. This is followed by an outline of some possible extensions to the basic random intercepts model, for example to allow for random coefficients or further hierarchical levels. We then review three of the most popular estimation procedures used to fit linear multilevel models: iterative generalised least squares, the EM algorithm and Fisher scoring. Prediction, residual estimation and a method for calculating confidence intervals for higher-level residuals are also discussed. In a multilevel model, residuals can be estimated for each level in the hierarchy. These are useful not only for checking model assumptions, but also for substantive reasons; for instance, they allow comparisons to be made between higher-level units after controlling for important background characteristics.

Finally, nonlinear models for discrete response data are presented. In particular, the multilevel logit model for binary responses and the multilevel multinomial logit model for polychotomous responses are described. Some of the many estimation procedures used to fit multilevel logit models for binary data and an algorithm for estimating the multinomial model are reviewed.

2.2 Single-Level Approaches: Individual versus Aggregate

Before the development of multilevel modelling, researchers were presented with two approaches to analyse their data based on the level of aggregation. For data with a two-level hierarchical structure, there were two alternatives: the individual level or the aggregate level.

To consider the individual level approach first, suppose that there are two hierarchical levels with J clusters at level 2. Let y_{ij} be the response for individual i in cluster j (of size n_j) and let x_{ij} be a covariate. Then the individual model would be

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + e_{ij}, \quad i = 1, \ldots, n_j;\; j = 1, \ldots, J \tag{2.1}$$

where

$$\begin{aligned} &e_{ij} \sim N(0, \sigma^2), \\ &\operatorname{cov}(e_{ij}, e_{i'j'}) = 0, \; ij \neq i'j'. \end{aligned}$$

The problem with analysing data at the individual level is that it is assumed that all individuals are independent, even within the same cluster, which is often an invalid assumption. If the context in which individual behaviour occurs is ignored, one may be open to the atomistic fallacy (Alker 1969).

The alternative is to analyse data at the second level by modelling the aggregate responses for each level 2 unit. This approach gives rise to the following model

$$\bar{y}_j = \beta_0 + \beta_1 \bar{x}_j + \bar{e}_j$$

where

$$\begin{aligned} \bar{y}_j &= \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}, \\ \bar{x}_j &= \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}, \\ \bar{e}_j &= \frac{1}{n_j} \sum_{i=1}^{n_j} e_{ij}. \end{aligned}$$

The problem with this approach is that relationships at the cluster level may not be the same as relationships at the individual level. The ecological fallacy (Robinson 1950) arises when one attempts to make inferences about individual behaviour from an analysis of aggregate data which fails to take into account within-cluster variability. Typically one finds that relationships are stronger at the aggregate level than at the individual level.

Multilevel models (also known as random effects models, random coefficients models, mixed models and hierarchical models) avoid these problems by working at both levels simultaneously. One of the most important advantages of the multilevel approach is that it is more faithful to the true nature of the social system in which individuals live. Whereas in the individual-level model all unexplained variation is assigned to the individual level, in a multilevel model variation is split into separate components corresponding to the levels in the hierarchy. Therefore it is possible to determine not only the individual-level background characteristics which lead to differences in individuals' behaviour, but also the extent to which these differences may be attributed to the context in which they live. This feature of multilevel modelling has proved important in educational research as it enables comparisons in schools' performances to be made after taking into account compositional effects such as the socioeconomic characteristics of pupils within each school (Aitkin and Longford 1986; Goldstein 1987).

In addition to the substantive advantages of a multilevel approach, there are technical reasons for favouring multilevel analysis above standard techniques, even if the substantive aspects are not of direct interest themselves. In a multilevel model, units within a cluster are allowed to be correlated. This is clearly a more realistic assumption than the independence assumption of a single-level analysis, since there are many situations where one would expect individuals in the same cluster to be more similar than individuals in different clusters. A consequence of this is that in a single-level analysis standard errors tend to be underestimated. Goldstein (1995) states that to obtain an estimate of the correct standard error for the estimate of β_1 in (2.1) in the presence of clustering, the ordinary least squares (OLS) estimate needs to be multiplied by a function of the within-cluster correlation and cluster size. This function is equal to 1 if there is no within-cluster correlation or if there is only one individual per cluster, in which case the

OLS estimate of the standard error will be adequate. However, if the within-cluster correlation is non-zero, there will be fewer independent observations per cluster, and as the cluster size increases the OLS estimate increasingly underestimates the true standard error. Therefore, if there is correlation between responses within a cluster, confidence intervals based on OLS estimates will tend to be too short and significance tests will too often reject the null hypothesis. This may lead to incorrect inferences and spurious relationships being detected.

Although the importance of using a multilevel approach to analyse hierarchical data has been recognised for many years, estimation of multilevel models was restricted by computational difficulties and it is only in the last 10 years, with the introduction of appropriate computer software, that these techniques have been viable.

2.3 The Linear Multilevel Model for Continuous Response Data

2.3.1 The Random Intercepts Model

The simplest form of a multilevel model is a random intercepts or variance components model. In this model, regression lines for the relationship between the response y and covariate x can have different intercepts for each cluster, though each line shares the same slope. In other words the average response (represented by the intercept) can differ among clusters, but the effect of the covariate x is constrained to be the same for each cluster. A plot of the regression of y on x would show a set of parallel lines, one for each cluster.

One approach to fitting such a model might be to fit a separate parameter for each cluster. The problem with this fixed effects model, however, is that the number of clusters J is often very large and so there would be a large number of parameters to estimate. The multilevel or random effects approach views the J clusters as a random sample selected from a larger population of clusters. Rather than estimating an intercept parameter for each cluster, the multilevel model assumes that the cluster-specific parameters are from a distribution for which we estimate the mean and variance.

Suppose y_{ij} is the response for individual i in cluster j and x_{ij} is a covariate. The random intercepts model can be written

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + e_{ij}, \tag{2.2}$$

where

$$\beta_{0j} = \beta_0 + u_{0j}.$$

We assume that

$$\begin{aligned} &e_{ij} \sim N(0, \sigma_e^2), \\ &u_{0j} \sim N(0, \sigma_0^2), \\ &\mathrm{cov}(e_{ij}, u_{0j}) = 0. \end{aligned}$$

Observations in the same cluster are not independent since

$$\begin{aligned} \mathrm{cov}(y_{ij}, y_{i'j'}) &= \sigma_0^2 + \sigma_e^2 \qquad && j = j',\ i = i' \\ &= \sigma_0^2 && j = j',\ i \neq i' \\ &= 0 && j \neq j'. \end{aligned}$$

The random intercepts model is often referred to as a 'variance components' model since the variance of the response y_{ij} can be decomposed into the sum of the level 1 and level 2 variances.

The single-equation formulation is

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{ij}.$$

$\beta_0 + \beta_1 x_{ij}$ can be thought of as the fixed part of the model, and $u_{0j} + e_{ij}$ the random part. β_0 (representing the intercept for the 'average' cluster) and β_1 are known as the fixed part parameters and σ_0^2 and σ_e^2 are the random part parameters.

A criterion for measuring the homogeneity of units within clusters compared to between clusters is the intra-class correlation which is defined as

$$\rho = \mathrm{corr}(y_{ij}, y_{i'j}) = \frac{\mathrm{cov}(y_{ij}, y_{i'j})}{\sqrt{\mathrm{var}(y_{ij})\mathrm{var}(y_{i'j})}} = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_e^2}$$

This measures the proportion of the total variation $(\sigma_0^2 + \sigma_e^2)$ which can be attributed to between-cluster variation. Hence if $\sigma_0^2 = 0$, that is, if there is no variation between the second level clusters, responses within a cluster are uncorrelated and (2.2) reduces to a single-level model.

2.3.2 The Random Coefficients Model

A direct extension of the random intercepts model in (2.2) is to allow the slope parameter β_1 to vary across clusters too. One could fit a fixed effects model which would involve fitting a separate regression line for each cluster, but again this is infeasible if there is a large number of clusters. Also some clusters are likely to contain too few individuals to estimate a regression model. It is more efficient to use the multilevel approach which assumes that the cluster-specific intercept and slope parameters come from some distributions for which we estimate the variances. The random coefficients or random slopes model can be written

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}, \tag{2.3}$$

where

$$\begin{aligned} \beta_{0j} &= \beta_0 + u_{0j}, \\ \beta_{1j} &= \beta_1 + u_{1j}. \end{aligned}$$

We assume

$$\begin{aligned} &e_{ij} \sim N(0, \sigma_e^2), \\ &u_{0j} \sim N(0, \sigma_0^2), \\ &u_{1j} \sim N(0, \sigma_1^2), \\ &\text{cov}(u_{0j}, u_{1j}) = \sigma_{01}, \\ &\text{cov}(e_{ij}, u_{0j}) = \text{cov}(e_{ij}, u_{1j}) = 0. \end{aligned}$$

The single-equation formulation is

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + x_{ij}u_{1j} + e_{ij}.$$

The fixed part of the model is $\beta_0 + \beta_1 x_{ij}$ as before, and the random part is $u_{0j} + x_{ij}u_{1j} + e_{ij}$. The random coefficients model is analogous to a single-level model which allows for interactions between covariates. In the random coefficients model, the term $x_{ij}u_{1j}$ can be regarded as an interaction between x_{ij} and the unobserved random effect u_{1j}.

2.3.3 Higher-Level Covariates

So far, we have only considered covariates which are specific to the level 1 units, but level 2 specific covariates which are constant within each level 2 cluster j can also be introduced.

The two-level random slopes model with one covariate x_{ij} at level 1 and one covariate w_j at level 2 can be written

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij},$$

where

$$\begin{aligned}\beta_{0j} &= \beta_0 + \gamma_0 w_j + u_{0j} \\ \beta_{1j} &= \beta_1 + \gamma_1 w_j + u_{1j},\end{aligned}$$

or using the single-equation formulation

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_0 w_j + \gamma_1 x_{ij} w_j + u_{0j} + x_{ij}u_{1j} + e_{ij}.$$

This model is sometimes called a 'slopes-as-outcomes' model (Bryk and Raudenbush 1992) as we have a regression model for both the random intercept and slope parameter. According to this model, part of the variation in the intercepts and slopes across level 2 units is predicted by w_j, but a random component represented by u_{0j} and u_{1j} remains unexplained. One strategy in multilevel modelling is to first include only level 1 covariates in the model and then to add level 2 covariates, in an attempt to explain the variation between level 2 units. It may be that differences between clusters can be partly or completely explained by certain level 2 covariates.

2.3.4 The General Form of the Two-Level Multilevel Model

Suppose we wish to extend the random slopes model in (2.3) to include p covariates in the fixed part and q covariates varying randomly at level 2 (including intercept terms).

The general two-level model for individual i in cluster j may be written

$$y_{ij} = \mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\mathbf{u}_j + e_{ij}, \tag{2.4}$$

where $\mathbf{x}_{ij}$ is a $p-$vector of covariates, defined either at level 1 or level 2; β is the associated vector of fixed parameters; $\mathbf{z}_{ij}$ is a $q-$vector of covariates (usually a subset of $\mathbf{x}_{ij}$) the effects of which vary randomly at level 2; $\mathbf{u}_j$ is a vector of level 2 random effects.

For example, in the case of the two-level random slopes model in (2.3) with one level 1 covariate we have

$$\begin{aligned}
\mathbf{x}_{ij} &= (1, x_{ij}),\\
\beta &= (\beta_0, \beta_1),\\
\mathbf{z}_{ij} &= (1, x_{ij}),\\
\mathbf{u}_j &= (u_{0j}, u_{1j}).
\end{aligned}$$

The model for individuals in cluster j can be written

$$\mathbf{y}_j = \mathbf{X}_j\beta + \mathbf{Z}_j\mathbf{u}_j + \mathbf{e}_j, \tag{2.5}$$

where $\mathbf{y}_j = (y_{1j}, \ldots, y_{n_jj})$; $\mathbf{e}_j = (e_{1j}, \ldots, e_{n_jj})$; $\mathbf{X}_j$ is the $n_j \times p$ design matrix for cluster j for the fixed parameters β; and $\mathbf{Z}_j$ is the $n_j \times q$ design matrix for the random effects $\mathbf{u}_j$ varying across level 2 units. Let $\text{var}(\mathbf{u}_j) = \mathbf{\Omega}_j$.

The matrix form of the general model can be written

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{2.6}$$

where $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_J)$, $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_J)$ and $\mathbf{e} = (\mathbf{e}_1, \ldots, \mathbf{e}_J)$. The design matrices $\mathbf{X}$ and $\mathbf{Z}$ are of dimension $n \times p$ and $n \times Jq$ respectively ($n = \sum_{j=1}^{J} n_j$). The variance-covariance matrix for the random level 2 parameters has a block diagonal structure, with a block for each cluster, i.e., $\text{var}(\mathbf{u}) = \mathbf{\Omega} = \text{diag}(\mathbf{\Omega}_1, \ldots, \mathbf{\Omega}_J)$.

2.3.5 A Three-Level Model

Any of the two-level models described above can be extended to three levels or more. For example, extending the two-level random intercepts model in (2.2) to a three-level model where intercepts vary across both level 2 and level 3 units gives

$$y_{ijk} = \beta_{0jk} + \beta_1 x_{ij} + e_{ijk},$$

where

$$\beta_{0jk} = \beta_0 + u_{0jk} + v_{0k}.$$

u_{0jk} and v_{0k} are random effects corresponding to level 2 and level 3 respectively. The single-equation formulation is

$$y_{ijk} = \beta_0 + \beta_1 x_{ijk} + u_{0jk} + v_{0k} + e_{ijk}.$$

We assume

$$\begin{aligned} e_{ijk} &\sim N(0, \sigma_e^2), \\ u_{0jk} &\sim N(0, \sigma_{u0}^2), \\ v_{0k} &\sim N(0, \sigma_{v0}^2) \end{aligned}$$

For a three-level model, we can calculate the intra-class correlation for level 3 units as well as for level 2 units. For the three-level random intercepts model above, the total variation is $\sigma_e^2 + \sigma_{u0}^2 + \sigma_{v0}^2$. Therefore, the proportion of the total variance due to variation among level 2 units is $(\sigma_{u0}^2 + \sigma_{v0}^2)/(\sigma_e^2 + \sigma_{u0}^2 + \sigma_{v0}^2)$ and the proportion due to level 3 units is $\sigma_{v0}^2/(\sigma_e^2 + \sigma_{u0}^2 + \sigma_{v0}^2)$.

2.4 Estimation Procedures for the Linear Multilevel Model

In this section, the algorithms most commonly used to estimate linear multilevel models of the type described in Section 2.3 are described. The three main algorithms are iterative generalised least squares (Goldstein 1986), the EM algorithm (Dempster et al. 1981; Mason et al. 1983), and a Fisher scoring algorithm (Longford 1987), all of which have been implemented in available software.

2.4.1 Iterative Generalised Least Squares

Consider the general two-level model in (2.6)

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where

$$E(\mathbf{u}) = E(\mathbf{e}) = \mathbf{0},$$
$$\text{var}(\mathbf{u}) = \mathbf{\Omega},$$
$$\text{var}(\mathbf{e}) = \sigma_e^2\mathbf{I}_n.$$

Let ϵ equal $\mathbf{y} - \mathbf{X}\beta = \mathbf{Zu} + \mathbf{e}$, the random part of the model. Also let $\mathbf{\Sigma}$ equal $\text{var}(\mathbf{y}) = \text{var}(\epsilon) = \mathbf{Z\Omega Z}' + \sigma_e^2\mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix.

If $\mathbf{\Sigma}$ were known, β could be estimated using generalised least squares (GLS) to give the estimator

$$\hat{\beta} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{y}.$$

Usually, however, neither $\mathbf{\Sigma}$ or β are known. Goldstein (1986) uses a variation on GLS, the iterative generalised least squares (IGLS) procedure, to estimate the linear multilevel model. Essentially, this involves alternating between the estimation of $\mathbf{\Sigma}$ and β, updating the estimate of one with the current estimate of the other at each iteration until convergence is reached.

Let $\mathbf{a}$ be the $\frac{1}{2}n(n+1) \times 1$ vector formed by stacking the squares and crossproducts of distinct elements of ϵ, e.g., the component of $\mathbf{a}$ corresponding to cluster j is

$$\mathbf{a}_j = ({\epsilon_{1j}}^2, \epsilon_{1j}\epsilon_{2j}, {\epsilon_{2j}}^2, \epsilon_{1j}\epsilon_{3j}, \epsilon_{2j}\epsilon_{3j}, {\epsilon_{3j}}^2, \ldots, {\epsilon_{n_jj}}^2).$$

Then $E(\mathbf{a}) = \mathbf{W}\gamma$ where γ is a vector containing the variances and covariances of the random parameters at levels 1 and 2, i.e., the elements of Ω and σ_e^2; $\mathbf{W}$ is a design matrix which depends on $\mathbf{Z}$. We can then write the model

$$\mathbf{a} = \mathbf{W}\gamma + \xi,$$

where $E(\xi) = \mathbf{0}$.

Let $\mathbf{\Gamma} = \text{var}(\mathbf{a}) = \text{var}(\xi)$. $\mathbf{\Gamma}$ turns out to be a function of the variances and covariances of $\mathbf{u}$ and $\mathbf{e}$, i.e., $\mathbf{\Gamma} = \mathbf{\Gamma}(\gamma)$. If $\mathbf{\Gamma}$ is known, the GLS estimator of γ is

$$\hat{\gamma} = (\mathbf{W}'\mathbf{\Gamma}^{-1}\mathbf{W})^{-1}\mathbf{W}'\mathbf{\Gamma}^{-1}\mathbf{a}.$$

The IGLS algorithm consists of three main steps:

Step 0 The initial estimate of $\mathbf{\Sigma}$.

The initial estimate of $\mathbf{\Sigma}$ is $\mathbf{\Sigma}^{(0)} = \mathbf{I}_n$, where the starting values for the variances and covariances of the level 2 random parameters are all zero and the initial estimate for the level 1 variance is $\sigma_e^2 = 1$.

Step 1 Estimating the fixed part parameters β.

At iteration m, the estimate of β is

$$\beta^{(m)} = \left(\mathbf{X}'\mathbf{\Sigma}^{(m-1)^{-1}}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{\Sigma}^{(m-1)^{-1}}\mathbf{y},\ m = 1, 2, \ldots.$$

The first estimate of β is obtained by substituting $\mathbf{\Sigma}^{(0)}$ to give $\beta^{(1)} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$, which is the ordinary least squares estimator for the single-level model.

Step 2 Estimating the random part parameters γ.

We first derive $\mathbf{a}^{(m)}$ from $\epsilon^{(m)}$, where $\epsilon^{(m)} = \mathbf{y} - \mathbf{X}\beta^{(m)}$.

Since $\mathbf{\Gamma}$ is a function of the elements of $\mathbf{\Sigma}$, an estimate of $\Gamma^{(m-1)}$ can be calculated from $\Sigma^{(m-1)}$. Then the estimate of γ at iteration m can be computed as

$$\gamma^{(m)} = \left(\mathbf{W}'\mathbf{\Gamma}^{(m-1)^{-1}}\mathbf{W}\right)^{-1}\mathbf{W}'\mathbf{\Gamma}^{(m-1)^{-1}}\mathbf{a}^{(m)}.$$

A new estimate of $\mathbf{\Sigma}$, $\mathbf{\Sigma}^{(m)}$, is then computed from $\gamma^{(m)}$. This is used in a repeat of Step 1 to calculate an updated estimate of β, $\beta^{(m+1)}$, at the next iteration.

Steps 1 and 2 are repeated alternately for iterations $m = 1, 2, \ldots$ until convergence is achieved.

The IGLS algorithm for the estimation of the multilevel linear model is implemented in the software MLn (Rasbash and Woodhouse 1995). However, Goldstein (1995) remarks that for small samples IGLS may produce biased estimates of the random parameters because it does not take into account the sampling variation of the fixed parameters. In such cases, a modification of IGLS, restricted IGLS (RIGLS), is recommended to obtain unbiased estimates (Goldstein 1989).

2.4.2 The EM Algorithm

Dempster et al. (1981) were among the first to apply the EM algorithm (Dempster et al. 1977), originally developed for incomplete data problems, to the estimation of linear multilevel models. To illustrate the basic idea of this approach, consider the two-level model in (2.6)

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where

$$\mathbf{e} \sim N(\mathbf{0}, \sigma_e^2 \mathbf{I}_n),$$

and

$$\boldsymbol{\Omega} = \text{diag}(\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_J),$$

where $\boldsymbol{\Omega}_j = \boldsymbol{\Omega}_u$ say, $j = 1, \ldots, J$.

The random effects $\mathbf{u}$ are regarded as missing data. Therefore the actual observed data $\mathbf{y}$ are viewed as the 'incomplete' data and $(\mathbf{y}, \mathbf{u})$ is the 'complete' data. The likelihood function for the complete data can be derived from the joint distribution of $(\mathbf{y}, \mathbf{u})$ and is of the form

$$L(\mathbf{y}, \mathbf{u}) = f(\mathbf{y}|\mathbf{u}).f(\mathbf{u}).$$

Bryk and Raudenbush (1992) show that the complete data log-likelihood, $l(\mathbf{y}, \mathbf{u}) = \log L(\mathbf{y}, \mathbf{u})$, is proportional to

$$\begin{aligned} l(\mathbf{y}, \mathbf{u}|\beta, \sigma_e^2, \boldsymbol{\Omega}_u) &\propto -n \log(\sigma_e^2) - J \log(\det \boldsymbol{\Omega}_u) \\ &- \mathbf{e}'\mathbf{e} - \mathbf{u}'\boldsymbol{\Omega}_u^{-1}\mathbf{u} \end{aligned}$$

The log-likelihood can be maximised with respect to β, σ_e^2 and $\boldsymbol{\Omega}_u$ to give the following estimates

$$\begin{aligned} &\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{Z}\mathbf{u}), \\ &\hat{\sigma}_e^2 = n^{-1}\mathbf{e}'\mathbf{e}, \\ &\hat{\boldsymbol{\Omega}}_u = J^{-1}\mathbf{u}\mathbf{u}'. \end{aligned}$$

These estimates are in terms of their sufficient statistics, i.e., the sufficient statistics for β, σ_e^2 and $\boldsymbol{\Omega}_u$ are respectively $\mathbf{X}'(\mathbf{y} - \mathbf{Z}\mathbf{u})$, $\mathbf{e}'\mathbf{e}$ and $\mathbf{u}\mathbf{u}'$.

Each iteration of the EM algorithm consists of two main steps called the expectation (E) step and the maximisation (M) step.

In the E-step, the expectations of the sufficient statistics for β, σ_e^2 and $\mathbf{\Omega}_u$, conditional on $\mathbf{y}$ and the current parameter estimates are calculated. Formulae for these conditional expectations can be found in Bryk and Raudenbush (1992). In the first iteration, initial values of β and σ_e^2 are obtained from OLS regression and $\mathbf{\Omega}_u$ is set to $\mathbf{0}$.

In the M-step, the sufficient statistics in the expressions for $\hat{\beta}$, $\hat{\sigma}_e^2$ and $\hat{\mathbf{\Omega}}_u$ are replaced by their conditional expectations computed in the previous E-step. This produces updated estimates of β, σ_e^2 and $\mathbf{\Omega}_u$.

The E- and M-steps are repeated alternately until convergence is achieved.

The EM algorithm has been widely used for the estimation of multilevel models, for example see Mason et al. (1983) and Raudenbush and Bryk (1986). It has been implemented in the HLM program (Bryk et al. 1988) which can handle hierarchical data with up to three levels. The most attractive feature of the EM algorithm is that it always converges for any set of reasonable initial values. However, convergence can be very slow, especially for more complex models or when the random effect variances are small (Longford 1987).

2.4.3 The Fisher Scoring Algorithm

Longford (1987) describes how a fast Fisher scoring algorithm may be used for maximum likelihood estimation of linear multilevel models. To illustrate this procedure, consider again the two-level model in (2.6)

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where

$$E(\mathbf{u}) = E(\mathbf{e}) = \mathbf{0},$$
$$\text{var}(\mathbf{u}) = \mathbf{\Omega},$$
$$\text{var}(\mathbf{e}) = \sigma_e^2\mathbf{I}_n.$$

Let $\mathbf{\Sigma}$ equal $\text{var}(\mathbf{y}) = \text{var}(\mathbf{Z\Omega Z}' + \sigma_e^2\mathbf{I}_n)$, and let θ be a vector containing the random parameters in $\mathbf{\Sigma}$.

Then the log-likelihood can be written

$$l(\mathbf{y};\beta,\theta) = -\frac{1}{2}\{n\log(2\pi) + \log(\det\boldsymbol{\Sigma}) + (\mathbf{y}-\mathbf{X}\beta)'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\beta)\}.$$

The first and second derivatives with respect to β are

$$\frac{\partial l}{\partial \beta} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\beta),$$

and

$$\frac{\partial^2 l}{\partial\beta\partial\beta'} = -\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}.$$

Longford (1987) uses the Newton-Raphson procedure (Thisted 1988) to solve the score equations $\frac{\partial l}{\partial \beta} = 0$ to obtain the following estimate of β at iteration j:

$$\beta^{(j)} = \beta^{(j-1)} + \left[\left(-\frac{\partial^2 l}{\partial\beta\partial\beta'}\right)^{-1}\frac{\partial l}{\partial\beta}\right]_{\beta=\beta^{(j-1)}}.$$

In the *Fisher* scoring algorithm, $\frac{\partial^2 l}{\partial\beta\partial\beta'}$ is replaced by its expected value.

Expressions for the derivatives of the log-likelihood with respect to θ can be calculated to obtain estimates of the random parameters θ. There are two main stages in each iteration of the algorithm. In the first stage, an estimate of β is calculated using the current estimate of θ; in the second stage, θ is updated using the current estimate of β. The OLS estimates of β and σ_e^2 are used as starting values for β and σ_e^2 and the components of $\boldsymbol{\Omega}$ are initialised at zero.

The Fisher scoring algorithm has been implemented in the program VARCL (Longford 1988) which is available in two versions: VARL3, which can estimate random coefficient models to data with up to three levels, and VARL9, which can analyse data with up to nine levels but allows only a simple random intercepts structure.

2.5 Residuals and Predicted Values in Multilevel Analysis

2.5.1 Residual Estimation

Consider first the two-level random intercepts model in (2.2) with one covariate x_{ij}

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{ij},$$

where u_{0j} is the error term for level 2 cluster j, and e_{ij} is the level 1 error term for individual i in cluster j.

The predicted value of y_{ij}, given x_{ij} is

$$\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{ij},$$

and the total residual for individual i in cluster j is

$$r_{ij} = y_{ij} - \hat{y}_{ij} \simeq u_{0j} + e_{ij}.$$

However, in a multilevel model we require separate estimates of u_{0j} and e_{ij}. We let the predicted value of u_{0j} be proportional to the mean of r_{ij} in group j such that

$$\hat{u}_{0j} = \kappa \bar{r}_j = \frac{\kappa}{n_j} \sum_{i=1}^{n_j} r_{ij},$$

where κ is a constant chosen to minimise $E(\hat{u}_{0j} - u_{0j})^2$. $\hat{u}_{0j}$ is called a 'shrunken' residual, and κ is the 'shrinkage' factor which is always less than or equal to 1 (Goldstein 1995).

The optimum value of κ for the two-level random intercepts model can be shown to be

$$\kappa_{opt} = \frac{\sigma_0^2}{\sigma_0^2 + \frac{\sigma_e^2}{n_j}} = \frac{\mathrm{cov}(u_{0j}, \bar{r}_j)}{\mathrm{var}(\bar{r}_j)}.$$

Since $\kappa_{opt} \to 1$ as $n_j \to \infty$, u_{0j} is predicted by $\bar{r}_j$ for large cluster sizes. If n_j is small, however, $\hat{u}_{0j}$ will shrink towards zero (hence the term 'shrunken' residual). Thus, for a small cluster, the residual estimate reflects the relative lack of information in that unit and places the cluster mean close to the overall population value as predicted by the fixed part of the model. Note that for $n_j = 1$, $\kappa_{opt} = \rho$, the intra-class correlation. If all clusters are of size 1, there will be no intra-cluster correlation, and $\hat{u}_{0j}$ will equal zero which gives rise to a single-level model with error only at level 1.

Now consider the general two-level model in (2.4) with p covariates in the fixed part of the model and q covariates varying randomly at level 2,

$$y_{ij} = \mathbf{x}_{ij}'\beta + \mathbf{z}_{ij}'\mathbf{u}_j + e_{ij},$$

where $\text{var}(\mathbf{u}_j) = \mathbf{\Omega}_j$ and $\text{var}(e_{ij}) = \sigma_e^2$.

The predicted value of y_{ij} is $\hat{y}_{ij} = \mathbf{x}_{ij}'\hat{\beta}$, and $r_{ij} = y_{ij} - \hat{y}_{ij} \simeq \mathbf{z}_{ij}'\mathbf{u}_j + e_{ij}$ is the composite residual.

Let $\bar{\mathbf{r}}_j = (\bar{r}_{1j}, \ldots, \bar{r}_{n_j j})$, then it can be shown that the estimates of the level 2 residuals are

$$\hat{\mathbf{u}}_j = \hat{\text{cov}}(\mathbf{u}_j, \bar{\mathbf{r}}_j)\hat{\text{var}}(\bar{\mathbf{r}}_j)^{-1}\bar{\mathbf{r}}_j,$$

where $\hat{\text{cov}}(\mathbf{u}_j, \bar{\mathbf{r}}_j)$ and $\hat{\text{var}}(\bar{\mathbf{r}}_j)$ are matrices whose elements are functions of $\hat{\mathbf{\Omega}}_j$, $\hat{\sigma}_e^2$ and $\mathbf{z}_{ij}$.

The level 1 residuals e_{ij} are estimated by

$$\hat{e}_{ij} = r_{ij} - \mathbf{z}_{ij}'\hat{\mathbf{u}}_j.$$

As for a single-level model, the residuals from a multilevel model can be used for diagnostic purposes. The level 2 residuals can be standardised to take account of differences in cluster size n_j by dividing the elements of $\hat{\mathbf{u}}_j$ by their estimated standard errors. Normal plots of both sets of residuals can be used to check the normality assumptions. To test the assumptions of constant variance, residuals at each level can be plotted against further omitted covariates to see whether any more can be usefully added to the model. Another use of residuals is to check for the existence of any outlying observations. In the case of level 2 residuals, outliers are of also of substantive interest and may identify clusters for further, perhaps qualitative, research.

Level 2 residuals can also be used to predict 'cluster effects'. In their multilevel analysis of educational performance of pupils nested within schools, Aitkin and Longford (1986) used school-level residuals to predict 'school effects', that is the effect of a child being in a particular school on their O-level results, after controlling for their intellectual ability at the time of starting secondary school. The school residuals were then used to obtain adjusted rankings of schools according to their average exam performance. Raw rankings of school performance are not sensitive to important compositional factors such as the socioeconomic background of pupils. This often results in schools in deprived areas having low rankings in conventional school league tables. A multilevel analysis controls for such factors to allow fairer comparisons to be made between schools. However, some caution should be exercised when interpreting such rankings of level 2 units as they are sensitive to model misspecification and typically have large standard errors (Goldstein and Spiegelhalter 1996). It is possible that the addition of an important covariate to the model can completely change the rankings (Goldstein 1987). For example Aitkin and Longford (1986) note that if information on the social class composition of each school were available and included in the model, some of the observed between-school differences may have disappeared which would affect the overall rankings.

2.5.2 Confidence Intervals for Higher-Level Residuals

In order to test for significant differences between higher-level units, confidence intervals can be constructed for each higher-level residual to allow for the uncertainty in the estimates. Consider a simple two-level random intercepts model with random effect u_j. Let $\hat{u}_j$ be the estimated residual for cluster j and σ_j its standard error. Then the conventional 95% confidence interval for normally distributed u_j is given by

$$\hat{u}_j \pm 1.96\ \sigma_j.$$

Typically, however, comparisons across all level 2 units are of interest and, therefore, a slight adjustment to the conventional interval is needed. This is because when multiple comparisons are made the significance level $\alpha = 0.05$ is reduced which makes the confidence intervals wider than the required 95% intervals. Goldstein and Healy (1995) have proposed a procedure for the construction of simultaneous confidence intervals to test for differences between any pair of clusters, while ensuring that the average type I error over all possible pairwise comparisons remains at the specified value α. The criterion used to determine whether any two clusters are significantly different is to examine whether their respective confidence intervals overlap. If they do not overlap, the differences are statistically significant at the chosen level α.

To describe this approach in brief, suppose there are J independently normally distributed estimates of cluster-level residuals $\hat{u}_j$, $j = 1, \ldots, J$, with known standard errors σ_j. Suppose we wish to compare a pair of clusters j and k, with estimated residuals $\hat{u}_j$ and $\hat{u}_k$ and standard errors σ_j and σ_k respectively. The simultaneous 95% confidence interval for the j^{th} cluster residual u_j is given by

$$\hat{u}_j \pm z\, \sigma_j,$$

where z is selected so that the average significance level over all pairs of contrasts (j,k) is 5%. An approximation to z is the average of $1.96\ \sigma_{jk}/(\sigma_j + \sigma_k)$ over all pairs of clusters (j,k), where $\sigma_{jk}^2 = \text{var}(\hat{u}_j - \hat{u}_k) = \sigma_j^2 + \sigma_k^2$.

If there are J clusters, there are $\frac{1}{2}J(J-1)$ possible pairwise contrasts. Therefore

$$z \simeq 1.96 \times \frac{2}{J(J-1)} \sum_{j<k} \frac{\sigma_{jk}}{\sigma_j + \sigma_k}.$$

The process can be extended to enable simultaneous comparisons between more than two clusters.

Goldstein and Spiegelhalter (1996) illustrate the importance of displaying not only residual estimates, but also their confidence intervals to allow for the uncertainty in the estimates. They give several examples to show how confidence intervals can be used to make comparisons between institutions in the areas of education and health. For example, in a multilevel regression of exam performance on intake scores, there initially appears to be large variations in the school-level residuals. However, if a simultaneous 95% confidence interval is constructed for each school residual estimate, they find that of all possible comparisons between pairs of schools, 2/3 are not statistically significant. Only a few schools at the extremes can be isolated as having significantly higher or lower exam scores than the others. In a health example, Goldstein and Spiegelhalter (1996) present estimates of rankings of health boards in terms of teenage conception rates. Using a Markov chain Monte Carlo method, Gibbs sampling, variances of these rank estimates are computed which allow confidence intervals to be constructed for each health board. These show that once the uncertainty in the rankings is considered, it is not possible to reach any firm conclusions about the position of any health board compared to another.

2.6 The Multilevel Logit Model for Binary Response Data

So far, we have considered only linear multilevel models for continuous response data. In the remaining sections of this chapter, nonlinear multilevel models for binary and polychotomous response data are described. In this section, the multilevel logit model for clustered binary response data is presented, and in the next section some alternative estimation approaches are discussed.

Let y_{ij} be the binary response for individual i in cluster j. We assume that y_{ij} are independent Bernoulli random variables with 'success' probability $\pi_{ij} = \Pr(y_{ij} = 1)$.

Then the multilevel logit model can be written

$$y_{ij} = \pi_{ij} + e_{ij},$$

where $\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \eta_{ij} = \mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\mathbf{u}_j$.

Alternatively, the model can be written in matrix notation as

$$\mathbf{y} = \pi + \mathbf{e},$$

where

$$\log\left(\frac{\pi}{1-\pi}\right) = \mathbf{X}\beta + \mathbf{Z}\mathbf{u}.$$

As for the linear model, $\mathbf{u}$ is assumed to have a multivariate normal distribution with $\mathrm{var}(\mathbf{u}) = \Omega = \mathrm{diag}(\Omega_1, \ldots, \Omega_J)$, where $\Omega_j = \mathrm{var}(\mathbf{u}_j)$.

The conditional likelihood function for this model has the form

$$L(\beta|\mathbf{u}) = \prod_{j=1}^{J}\prod_{i=1}^{n_j}\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right)^{y_{ij}}(1-\pi_{ij}). \tag{2.7}$$

In order to estimate β and Ω, one needs to multiply (2.7) by the likelihood for the random effects $\mathbf{u}$ and then 'integrate out' the random effects $\mathbf{u}$ to obtain the following unconditional (marginal) likelihood

$$L(\beta, \Omega) = \int \ldots \int L(\beta|\mathbf{u})\Phi(\mathbf{u})\, d\mathbf{u}, \tag{2.8}$$

where $\Phi(.)$ is the probability density function of the multivariate normal.

However, this expression is intractable and some approximation must be used.

2.7 Estimation Procedures for the Multilevel Logit Model

There is a vast literature on the estimation of random effects models for clustered binary data. Many approaches have used some form of numerical integration to obtain the unconditional likelihood in (2.8), see for example Anderson and Aitkin (1985) or Im and Gianola (1988). The EM algorithm has also been used (Stiratelli et al. 1984), while other researchers, including Goldstein (1991) and Longford (1994), have opted for a maximum likelihood approach. Another approach is to use a Bayesian method, Gibbs sampling (Zeger and Karim 1991). Alternatively, it is possible to avoid the calculation of the integral in (2.8) by choosing a different link function from the usual logit. Conaway (1990) uses a log-log link and assumes the conjugate log-gamma distribution for the random effects which leads to a tractable integral. However, since these methods cannot be directly extended to other generalised linear models, only those approaches which assume normally distributed random effects are considered here.

In this section, we begin by looking at ways in which the IGLS algorithm (Section 2.4.1) can be adapted to handle binary response data. This is the estimation approach adopted in the later analysis chapters of this thesis. There are a number of other possible approaches, however, and some of the more popular of these are also discussed.

2.7.1 Marginal Quasi-Likelihood Estimation via Iterative Generalised Least Squares

Goldstein (1991) estimates the multilevel logit model by first linearising the model, using a first-order Taylor series expansion, and then applying IGLS or iterative reweighted least squares (IRLS: see McCullagh and Nelder 1989) as for the linear case.

Goldstein uses a first-order Taylor series expansion, expanding about trial values $\beta = \beta^{(0)}$ and $\mathbf{u} = \mathbf{0}$ to give the following approximation (the subscripts ij are omitted for convenience)

$$\pi \simeq \pi(\eta^{(0)}) + \sum_{k=1}^{p} x_k(\beta_k - \beta_k^{(0)}) \frac{\partial \pi}{\partial x_k \beta_k}(\eta^{(0)}) + \sum_{k=1}^{q} z_k u_k \frac{\partial \pi}{\partial z_k u_k}(\eta^{(0)}), \qquad (2.9)$$

where $\eta^{(0)} = \mathbf{x}'\beta^{(0)}$, and $\frac{\partial \pi}{\partial x_k \beta_k}(\eta^{(0)})$ and $\frac{\partial \pi}{\partial z_k u_k}(\eta^{(0)})$ are the first derivatives of π with respect to $x_k\beta_k$ and $z_k u_k$ respectively, evaluated at $\eta^{(0)}$.

For the logit link,

$$\frac{\partial \pi}{\partial x_k \beta_k}(\eta^{(0)}) = \frac{\partial \pi}{\partial z_k u_k}(\eta^{(0)}) = \pi^{(0)}(1-\pi^{(0)}) = w^{(0)} \text{ say,}$$

where $\pi^{(0)} = \pi(\eta^{(0)})$.

Therefore the approximate model is

$$y = \pi + e \simeq \pi^{(0)} + w^{(0)}\mathbf{x}'(\beta - \beta^{(0)}) + w^{(0)}\mathbf{z}'\mathbf{u} + e,$$

which can be written in matrix notation as

$$\mathbf{y} \simeq \pi^{(0)} + \mathbf{W}^{(0)}\mathbf{X}(\beta - \beta^{(0)}) + \mathbf{W}^{(0)}\mathbf{Z}\mathbf{u} + \mathbf{e}.$$

where $\mathbf{W}^{(0)} = \text{diag}(\pi^{(0)}(1-\pi^{(0)}))$, and $\mathbf{X}$ and $\mathbf{Z}$ are the design matrices for β and $\mathbf{u}$ respectively.

Rearranging this expression yields the approximate model

$$\mathbf{y}^* = \mathbf{X}^*\beta + \mathbf{Z}^*\mathbf{u} + \mathbf{e},$$

where $\mathbf{y}^* = \mathbf{y} - \pi^{(0)} + \mathbf{W}^{(0)}\mathbf{X}\beta^{(0)}$, the working dependent variable, and $\mathbf{X}^* = \mathbf{W}^{(0)}\mathbf{X}$ and $\mathbf{Z}^* = \mathbf{W}^{(0)}\mathbf{Z}$ are the working design matrices for the fixed and random part parameters respectively.

Approximating var(e) by $\mathbf{W}^{(0)}$ gives $\text{var}(\mathbf{y}) \simeq \mathbf{Z}^*\boldsymbol{\Omega}\mathbf{Z}^{*\prime} + \mathbf{W}^{(0)}$.

The structure of this model is that of a linear multilevel model and provides an approximation to the nonlinear multilevel logit model. Therefore, the IGLS procedure can be applied, as described in Section 2.4.1 for the continuous response model, to obtain estimates of β and $\boldsymbol{\Omega}$.

Alternatively, rather than applying the linearisation using initial values and then iterating to convergence, Goldstein (1991) proposes the use of an iterative reweighted least squares (IRLS) algorithm. This involves updating $\mathbf{W}^{(0)}$ and therefore $\mathbf{y}^*$, $\mathbf{X}^*$ and $\mathbf{Z}^*$ at each iteration. In this case the approximation at iteration m is

$$\pi \simeq \pi^{(m-1)} + \sum_{k=1}^{p} x_k(\beta_k - \beta_k^{(m-1)})\frac{\partial \pi}{\partial x_k \beta_k}(\eta^{(m-1)}) + \sum_{k=1}^{q} z_k u_k \frac{\partial \pi}{\partial z_k u_k}(\eta^{(m-1)}). \quad (2.10)$$

A series of macros which recalculate the working dependent variable, working covariates and iterative weights at each iteration fit this model in MLn.

When fitting the multilevel logit model in MLn an extra covariate z'_{ij} is declared at level 1 to ensure that y_{ij} has the required Bernoulli variance, $\pi_{ij}(1-\pi_{ij})$. The model can be written

$$y_{ij} = \pi_{ij} + e_{ij}z'_{ij}$$

where $z'_{ij} = \sqrt{\pi_{ij}(1-\pi_{ij})}$ and $\mathrm{var}(e_{ij}) = \sigma_e^2$.

When fitting the model, the level 1 variance is usually constrained to be purely binomial, in which case $\sigma_e^2 = 1$. Alternatively, the constraint on σ_e^2 can be relaxed to allow for extra-binomial variation (Goldstein 1991; Williams 1982).

2.7.2 Penalised Quasi-Likelihood Estimation via Iterative Generalised Least Squares

Rodriguez and Goldman (1995) found that estimation of the multilevel logit model using Goldstein's IGLS approach or Longford's approximate likelihood approach (see Section 2.7.4) can lead to a substantial downward bias in the estimation of the random effects. Their simulation study fitting random intercepts models to data with a range of different structures, including rectangular and non-rectangular structures and two- and three-level hierarchical structures, showed that the random effect variance was severely underestimated whenever it was 'large enough to be interesting'. Furthermore, they discovered that the fixed effects were also underestimated and differed very little from those obtained from a standard logit model which ignores the hierarchical structure.

A possible reason for the downward bias of the random effects towards zero could be due to the fact that the approaches used by Goldstein (1991) and Longford (1994) both use approximations based on a Taylor series expansion around $\mathbf{u} = \mathbf{0}$. Another factor could be cluster size since, in the simulation studies, the situation was found to be worse when the average cluster size was small.

Rodriguez and Goldman (1995) also consider the quadratic approximation proposed by Goldstein (1991) in which a second-order Taylor series expansion is used. The approximation in (2.10) is extended to incorporate second-order terms corresponding to the random effects, while omitting second-order terms involving the fixed effects, to give the following approximate model at iteration m

$$\begin{aligned}
y &= \pi^{(m-1)} + \sum_{k=1}^{p} x_k(\beta_k - \beta_k^{(m-1)})\frac{\partial \pi}{\partial x_k \beta_k}(\eta^{(m-1)}) + \sum_{k=1}^{q} z_k u_k \frac{\partial \pi}{\partial z_k u_k}(\eta^{(m-1)}) \\
&+ \frac{1}{2}\sum_{k=1}^{q}\sum_{l=1}^{q} z_k z_l u_k u_l \frac{\partial^2 \pi}{\partial z_k u_k \partial z_l u_l}(\eta^{(m-1)}) + e, \qquad (2.11)
\end{aligned}$$

where, for the logit link,

$$\frac{\partial^2 \pi}{\partial z_k u_k \partial z_l u_l} = \pi(1-\pi)\frac{(1-\exp(\mathbf{x}'\beta + \mathbf{z}'\mathbf{u}))}{(1+\exp(\mathbf{x}'\beta + \mathbf{z}'\mathbf{u}))}.$$

However, although the results using the quadratic approximation showed considerable improvement in the estimation of the fixed effects, the random effects still showed a strong downward bias.

Based on the penalised quasi-likelihood (PQL) approach of Breslow and Clayton (1993), Goldstein (1995) adapted the second-order approximation by expanding around $\mathbf{u} = \mathbf{u}^{(m-1)}$ rather than $\mathbf{u} = \mathbf{0}$, where $\mathbf{u}^{(m-1)}$ contains the current estimates of the level 2 residuals. In this case the approximation in (2.11) becomes

$$\begin{aligned}
y &= \pi^{(m-1)} + \sum_{k=1}^{p} x_k(\beta_k - \beta_k^{(m-1)})\frac{\partial \pi}{\partial x_k \beta_k}(\eta^{(m-1)}) + \sum_{k=1}^{q} z_k(u_k - u_k^{(m-1)})\frac{\partial \pi}{\partial z_k u_k}(\eta^{(m-1)}) \\
&+ \frac{1}{2}\sum_{k=1}^{q}\sum_{l=1}^{q} z_k z_l (u_k - u_k^{(m-1)})(u_l - u_l^{(m-1)})\frac{\partial^2 \pi}{\partial z_k u_k \partial z_l u_l}(\eta^{(m-1)}) + e,
\end{aligned}$$

where $\eta^{(m-1)} = \mathbf{x}'\beta^{(m-1)} + \mathbf{z}'\mathbf{u}^{(m-1)}$.

At each iteration in the IGLS or IRLS procedure, the current estimate of $\mathbf{z}'\mathbf{u}^{(m-1)}$ is added to the fixed predictor $\mathbf{x}'\beta^{(m-1)}$ and is used to compute the first and second derivatives of π.

Goldstein (1995) and Goldstein and Rasbash (1995) have compared simulation results for the first-order MQL and second-order PQL approximations and found that the PQL estimates are much closer to the true values than those obtained using MQL. In particular, PQL also performs well for very unbalanced designs with small cluster sizes, the situations in which Rodriguez and Goldman (1995) found MQL to produce the most severely biased estimates. Both MQL and PQL with either first- or second-order approximations have since been implemented in MLn. However, Goldstein and Rasbash (1995) also note that while the first-order PQL approximation nearly always converges, second-order PQL does not always do so.

2.7.3 The Iterative Bootstrap

Goldstein (1996) proposes using an iterated bootstrap procedure in situations where the MQL approximation (Section 2.7.1) is inadequate and when the improved second-order PQL approximation (Section 2.7.2) fails to converge. This is based on a procedure developed by Kuk (1995) to obtain unbiased parameter estimates in generalised linear models with random effects and can be applied to any type of nonlinear multilevel model. The bootstrap (Efron and Tibshirani 1993) is a method used to obtain point estimates and standard errors via simulation. It is commonly used to provide robust standard errors and confidence intervals in situations where there is a departure from the distributional assumptions or when standard errors cannot be directly obtained from the estimation procedure, for example when the EM algorithm is used. The bootstrap involves generating a random sample of responses, known as a 'bootstrap sample', from a fitted model and then refitting the same model to the simulated data set to obtain a new set of parameter estimates. This procedure is repeated a large number of times, M say, to give M sets of parameter estimates. The 'bootstrap estimates' can then be calculated by taking the mean of each parameter estimate over the M samples. Their variances are estimated as the variances of the parameter estimates over all bootstrap samples.

In the iterative bootstrap procedure proposed by Goldstein (1996), the bootstrap is used to estimate the biases in the parameter estimates at each iteration. The aim is to then use these bias estimates to correct the original parameter estimates in order to obtain estimates which are asymptotically unbiased. To illustrate the method, consider the following two-level logistic model

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \mathbf{x}'\beta + \mathbf{z}'\mathbf{u}, \tag{2.12}$$

where

$$\mathbf{u} \sim N(\mathbf{0}, \mathbf{\Omega})$$
$$y_{ij} \sim \text{Bernoulli}(\pi_{ij}).$$

Iteration 0 In the first stage of the procedure, initial estimates of the parameters β and $\mathbf{\Omega}$ are obtained. Goldstein (1996) recommends using the estimates from the first-order MQL approximation as starting values as MQL nearly always converges and is considerably faster then PQL. Denote the initial estimates by

$$\hat{\beta}^{(0)}, \hat{\mathbf{\Omega}}^{(0)}. \tag{2.13}$$

Iteration 1 In the first iteration, a set of M_1 bootstrap samples are generated by sampling repeatedly from a distribution with the parameters in (2.13). Binomial and normal distributions are assumed for the level 1 and 2 residual terms respectively. The model in (2.12) is refitted to each of the bootstrap samples. We then calculate the average of each of the M_1 estimates of β and $\mathbf{\Omega}$ to get a set of bootstrap estimates

$$\tilde{\beta}^{(0)}, \tilde{\mathbf{\Omega}}^{(0)}. \tag{2.14}$$

To determine the number of bootstrap samples required at each iteration, Goldstein (1996) keeps a running mean for the estimates of β and $\mathbf{\Omega}$. In other words, the bootstrap estimates are updated after every resample and convergence is accepted when two consecutive running means differ by less than a prespecified tolerance level.

The bootstrap estimates of the bias in $\hat{\beta}^{(0)}$ and $\hat{\mathbf{\Omega}}^{(0)}$ are obtained by subtracting (2.13) from (2.14). These bias estimates are then subtracted from the initial parameter estimates as a first adjustment to give the following bias-corrected estimates

$$\begin{aligned} \hat{\beta}^{(1)} &= \hat{\beta}^{(0)} - (\tilde{\beta}^{(0)} - \hat{\beta}^{(0)}) \\ \hat{\mathbf{\Omega}}^{(1)} &= \hat{\mathbf{\Omega}}^{(0)} - (\tilde{\mathbf{\Omega}}^{(0)} - \hat{\mathbf{\Omega}}^{(0)}) \end{aligned} \tag{2.15}$$

Iteration 2 In the second iteration, a new set of M_2 bootstrap samples are generated from the model based on the current estimates in (2.15). Averaging over the estimates obtained from each of the M_2 resamples gives a new set of bootstrap estimates

$$\tilde{\beta}^{(1)}, \tilde{\mathbf{\Omega}}^{(1)}. \tag{2.16}$$

The estimates in (2.15) are subtracted from (2.16), and this is subtracted from the initial estimates to obtain a new set of bias-corrected estimates

$$\hat{\beta}^{(2)} = \hat{\beta}^{(0)} - (\tilde{\beta}^{(1)} - \hat{\beta}^{(1)})$$
$$\hat{\Omega}^{(2)} = \hat{\Omega}^{(0)} - (\tilde{\Omega}^{(1)} - \hat{\Omega}^{(1)})$$

This procedure is repeated until convergence is reached. For example, at the kth iteration, the bias-corrected estimates are given by

$$\hat{\beta}^{(k)} = \hat{\beta}^{(0)} - (\tilde{\beta}^{(k-1)} - \hat{\beta}^{(k-1)})$$
$$\hat{\Omega}^{(k)} = \hat{\Omega}^{(0)} - (\tilde{\Omega}^{(k-1)} - \hat{\Omega}^{(k-1)}).$$

When the procedure has converged, the final step is to estimate the standard errors of the bias-corrected estimates. These can also be estimated using the bootstrap. When convergence is achieved, a new set of bootstrap samples is generated from the final parameter estimates. The standard errors are estimated as the variances of the parameter estimates over all samples.

At convergence, Kuk (1995) demonstrates that the bias-corrected estimates are asymptotically consistent, i.e. they tend to the true parameter values in large samples, and unbiased. However, although the procedure is relatively straightforward to implement, it is highly computationally intensive. For example, Goldstein (1996) found that in a simulation study with a data set of 100 observations, an average of 81 bootstrap samples per iteration and 14 iterations were required to achieve convergence.

2.7.4 Fisher Scoring

Longford (1994) uses an approximation to the likelihood function in (2.8) based on a second-order Taylor series expansion of the logarithm of the conditional likelihood in (2.7) about $\mathbf{u} = \mathbf{0}$,

$$\log\ L(\beta|\mathbf{u}) \simeq \log\ L(\beta|\mathbf{0}) + (\mathbf{y} - \pi^{(0)})'\mathbf{Z}\mathbf{u} - \frac{1}{2}\mathbf{u}'\mathbf{Z}'\mathbf{W}^{(0)}\mathbf{Z}\mathbf{u},$$

where $\mathbf{W}^{(0)} = \text{diag}(\pi^{(0)}(1 - \pi^{(0)}))$.

Using this approximation, the unconditional likelihood in (2.8) can be integrated analytically. As for the continuous response case (Section 2.4.3), expressions for the first and second derivatives of the log-likelihood with respect to β and θ (the parameters in $\boldsymbol{\Omega}$) can be obtained. The Fisher scoring algorithm is then applied, alternating between the estimation of β and $\boldsymbol{\Omega}$. This approximation has been implemented in VARCL. However, Rodriguez and Goldman (1995) show that this procedure is in fact equivalent to Goldstein's MQL approach and can produce severe downward biases in the parameter estimates.

2.7.5 Numerical Integration and the EM Algorithm

A number of researchers have used Gaussian quadrature for the evaluation of the likelihood function in (2.8). This is feasible for relatively simple models such as a two-level model (Anderson and Aitkin 1985) or for a three-level model where there are no covariates defined at the first level (Im and Gianola 1988). The package SABRE (Barry et al. 1989) uses numerical integration via Gaussian quadrature to estimate simple random intercepts models. Alternatively, it is possible to approximate the logistic-normal model by a logistic-binomial model, where the random effect is assumed to follow a symmetric standardised binomial distribution (Mauritsen 1984). This approach has been implemented in the program EGRET (Statistics and Epidemiology Research Corporation 1991) which can fit more general random coefficient models but only for two-level structures. The problem with numerical integration approaches is that they can be highly computationally intensive for more complex models.

The EM algorithm can also be used to estimate binary logistic models (Stiratelli et al. 1984; Wong and Mason 1985). Again, however, this procedure rapidly becomes extremely computer-intensive for complex models, since numerical integrations are required at every iteration in order to calculate the conditional expectations in the E-step. Also, since usually many iterations are required, convergence can be very slow.

2.7.6 Generalised Estimating Equations

Another estimation approach is that of generalised estimating equations (GEE), introduced by Liang and Zeger (1986). GEE was originally developed for longitudinal data analysis to take into account correlation between observations over time on an individual. However, the procedure can be used to estimate any generalised linear model with a complex error structure. Zeger et al. (1988) make a distinction between two classes of model for hierarchically structured data with discrete or continuous outcomes, both of which can be estimated via GEE: the population-averaged model and the subject- or unit-specific model. For illustrative purposes, let us consider a two-level logit model for binary response data y_{ij}. The population-averaged (PA) model can be written

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \mathbf{x}'_{ij}\beta^*,$$

where

$$\begin{aligned}
\pi_{ij} &= E(y_{ij}),\\
\mathrm{var}(y_{ij}) &= \alpha\pi_{ij}(1-\pi_{ij}),\\
\mathrm{var}(\mathbf{y}_j) &= \Sigma_j.
\end{aligned}$$

In the PA model, the correlation between responses y_{ij} is represented by a parameter α which is chosen to yield a structure for Σ_j specific to each application. In the above model, an exchangeable or equicorrelation structure is assumed where observations within a level 2 unit are equally correlated over time. More complex correlation structures can be specified, for example, an autocorrelation structure where an individual's response at one point in time depends on their responses at one or more previous time points. An advantage of the PA approach is that no parametric assumptions about the level 2 heterogeneity distribution are needed. However, in the PA model, primary interest is focused on obtaining reliable estimates of β^* and the level 2 variance structure is regarded as a nuisance. Therefore the PA approach does not provide estimates of the variance components, only of the correlation parameter α. In this sense, PA models are not multilevel models at all. Further, since the PA model provides no specific information about higher level variation, one can only make inferences about average population effects.

An alternative specification is the unit-specific (US) model, which is equivalent to the multilevel logit model described in Section 2.6

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\mathbf{u}_j,$$

where

$$\pi_{ij} = E(y_{ij}),$$
$$\text{var}(y_{ij}) = \pi_{ij}(1 - \pi_{ij}),$$
$$\text{var}(\mathbf{y}_j) = \mathbf{\Sigma}_j.$$

$\mathbf{u}_j$ is usually assumed to follow a normal distribution, $\mathbf{u}_j \sim N(\mathbf{0}, \mathbf{\Omega_j})$.

In the US model, the dependency between individuals within a level 2 unit arises solely from the shared level 2 random effects $\mathbf{u}_j$. Since the hierarchical structure is specified explicitly in the model, the US model allows us to estimate the change in response probability corresponding to a change in $\mathbf{x}$ for any given level 2 unit. The PA model only allows us to estimate the change in response probability for the population as a whole.

Both the PA and US model can be estimated using GEE. For example, the generalised estimating equations for the US model are

$$\left(\frac{\partial \pi}{\partial \beta}\right)' \mathbf{\Sigma}^{-1}(\mathbf{y} - \pi) = \mathbf{0}, \tag{2.17}$$

which are solved for β.

For continuous response data with multivariate normal random effects, these are equivalent to the least squares estimating equations.

An iterative estimation procedure is used. At each iteration, an estimate of Σ is obtained which is used to update the estimate of β via (2.17). However, for the logit link, no closed-form expression for $\mathbf{\Sigma}$ is available. Zeger et al. (1988) propose an approximation to $\mathbf{\Sigma}$ based on a first-order Taylor series expansion about $\mathbf{u}_j = \mathbf{0}$. They find that in general the algorithm converges except when the linear approximation used to calculate $\hat{\mathbf{\Sigma}}$ becomes inaccurate. This occurs when the response probabilities are extreme, either too large or too small, when there are few individuals per level 2 unit, or when the level 2 variances become large. Therefore, Zeger et al. (1988) note that the GEE estimates of $\mathbf{\Omega}$ obtained for nonlinear models are only 'rough' approximations and as such should be interpreted with caution.

2.7.7 Gibbs Sampling

Gelfand and Smith (1990) give an overview of the potential of Gibbs sampling, a Markov chain Monte Carlo method, in a number of applications. In particular, this approach is recommended for problems where there is no exact solution and where standard approximate solutions have difficulties. Zeger and Karim (1993) describe how the Gibbs sampler can be used to estimate a two-level logit model with random coefficients. To illustrate this technique in brief, consider the following two-level logit model with random intercepts (omitting subscripts ij)

$$\log\left(\frac{\pi}{1-\pi}\right) = \mathbf{x}'\beta + u,$$

where $\beta = (\beta_1, \ldots, \beta_k)$ and $\mathrm{var}(u) = \sigma_u^2$.

The Bayesian approach to model estimation views β as random rather than as a set of fixed parameters as in the frequentist approach. Suppose β, σ_u^2 and u have joint probability density $f(\beta, \sigma_u^2, u|y)$ and let $f_1(\beta|\sigma_u^2, u, y)$, $f_2(\sigma_u^2|\beta, u, y)$ and $f_3(u|\beta, \sigma_u^2, y)$ denote the conditional distributions of β, σ_u^2 and u respectively. Although $f(\beta, \sigma_u^2, u|y)$ is intractable, the conditional distributions have simple forms.

The Gibbs sampler generates a sample from the joint distribution, without the need to calculate $f(\beta, \sigma_u^2, u|y)$, as follows. At iteration j, $\beta^{(j)}$ is sampled from $f_1(\beta|\sigma_u^{2(j-1)}, u^{(j-1)}, y)$. Next $\sigma_u^{2(j)}$ is sampled from $f_2(\sigma_u^2|\beta^{(j)}, u^{(j-1)}, y)$. Finally $u^{(j)}$ is sampled from $f_3(u|\beta^{(j)}, \sigma_u^{2(j)}, y)$ to obtain the sample $(\beta^{(j)}, \sigma_u^{2(j)}, u^{(j)})$.

It can be shown that for large j the distribution of $(\beta^{(j)}, \sigma_u^{2(j)}, u^{(j)})$ tends to the joint distribution of the unknown quantities, $f(\beta, \sigma_u^2, u|y)$. An initial run of B iterations, known as 'burn in', is performed until convergence is thought to have been reached. Gelfand and Smith (1990) advocate that the initial B iterations are carried out M times, while only the B^{th} sample in each run is retained, to ensure that the M samples are independent. However, Zeger and Karim (1993) suggest that this may not be necessary and that one run is sufficient in which a further M iterations are carried out after the 'burn in' period and only the first B samples are discarded. The estimate of β_i, $i = 1, \ldots, k$, can then be approximated by the mean of the estimates of β_i across all M

samples

$$\hat{\beta}_i = \frac{1}{M} \sum_{j=B}^{B+M} \beta_i^{(j)},$$

and their standard errors can be approximated by the standard deviations

$$s_{\hat{\beta}_i} = \sqrt{\frac{1}{M-1} \sum_{j=B}^{B+M} (\beta_i^{(j)} - \hat{\beta}_i)^2}.$$

Estimates of σ_u^2 and its standard error can be obtained in a similar way.

The Gibbs sampler has been implemented in the BUGS software (Spiegelhalter et al. 1994) which can be used for a wide range of complex problems, including linear and nonlinear multilevel models, measurement error and missing data. In the formulation of the model, the user specifies prior distributions for all the parameters of interest. A common practice is to specify 'non-informative' or flat priors which have very large variances for β and σ_u^2, while u is assumed to follow a normal distribution with mean 0 and variance σ_u^2.

An advantage of the Gibbs sampler is that it is relatively easy to implement. It is a technique with considerable potential in the estimation of complex multilevel models. However, it is highly computer-intensive and judging convergence can be problematic.

2.8 A Multilevel Model for Polychotomous Response Data

In this section, a nonlinear model for hierarchical polychotomous response data is described. The multilevel multinomial model is presented, followed by a description of the data structure required to fit the model in MLn and an outline of how the IGLS algorithm may be modified to estimate the model.

2.8.1 The Multilevel Multinomial Logit Model

Let y_{ij} be the polychotomous response for individual i in cluster j. Suppose the response has s possible categories and let $\pi_{rij} = \Pr(y_{ij} = r)$, $r = 1, \ldots, s$, such that $\sum_{r=1}^{s} \pi_{rij} = 1$.

In addition, define s dummy variables y_{rij},

$$y_{rij} = \begin{cases} 1 & \text{if } y_{ij} = r \\ 0 & \text{else, } r = 1, \ldots, s, \end{cases}$$

so that $\pi_{rij} = \Pr(y_{rij} = 1)$.

Then taking the final category s as the baseline, the two-level multinomial logit model can be written

$$y_{rij} = \pi_{rij} + e_{rij},$$

where

$$\log\left(\frac{\pi_{rij}}{\pi_{sij}}\right) = \eta_{rij} = \mathbf{x}_{rij}{}'\beta_r + \mathbf{z}_{rij}{}'\mathbf{u}_{rj}, \; r = 1, \ldots, s-1 \tag{2.18}$$

As before, $\mathbf{x}_{rij}$ is a vector of covariates in the fixed part of the model with associated parameters β_r, and $\mathbf{z}_{rij}$ is a vector of covariates random at level 2 with associated random effects $\mathbf{u}_{rj}$. Since the effect of covariates may vary across response categories, separate β's and $\mathbf{u}$'s are estimated for each contrast with the baseline. Although in general the covariates $\mathbf{x}$ and $\mathbf{z}$ will be the same set and of the same functional forms for each contrast, there is no such restriction in the model. The $\mathbf{u}_{rj}$ are assumed to follow a multivariate normal distribution with zero expectation and variance $\mathbf{\Omega}_{rj}$. The random effects for the same cluster but for two separate contrasts, $\mathbf{u}_{rj}$ and $\mathbf{u}_{r'j}$ $r \neq r'$ say, can be correlated.

The probability of response category r for individual i in cluster j can be written

$$\pi_{rij} = \frac{\exp(\eta_{rij})}{1 + \sum_{k=1}^{s-1} \exp(\eta_{kij})}, \; r = 1, \ldots, s-1,$$

and the probability of response category s (the baseline) is

$$\pi_{sij} = \frac{1}{1 + \sum_{k=1}^{s-1} \exp(\eta_{kij})}.$$

Goldstein (1995) describes how the multilevel multinomial logit model can be formulated as a multilevel multivariate model, where $\mathbf{y}_{ij} = (y_{1ij}, \ldots, y_{sij})$ is the multivariate response for individual ij. Using this approach, the two-level model in (2.18) is fitted as a three-level model with the multivariate response at level 1, the individual now at level 2 and the cluster now at level 3.

2.8.2 The Multivariate Data Structure

In order to fit the multilevel multinomial logit model, the data set must first be restructured to obtain multiple records for each individual, corresponding to the elements of the multivariate binary response vector $\mathbf{y}_{ij}$.

To illustrate the way in which a data set must be restructured, let us consider a simple example. Suppose the multinomial response y_{ij} has three categories and that we wish to fit a simple two-level random intercepts model with one explanatory variable AGE. Suppose the third category of y_{ij} is taken as the baseline. Then the model can be written

$$\log\left(\frac{\pi_{rij}}{\pi_{3ij}}\right) = \alpha_r + \beta_r AGE_{ij} + u_{rj}, \ r = 1, 2. \tag{2.19}$$

Consider the following sample data set

Individual i	Cluster j	y_{ij}	Constant	AGE
1	1	1	1	20
2	1	3	1	25
3	2	2	1	30

The data need to be restructured to obtain two records, or $s - 1$ in the general case, for each individual (the record corresponding to the baseline category is redundant). In addition, one needs to define a new set of covariates for each response category in order to estimate separate intercept and slope parameters for each contrast with the baseline category.

In the above example, the restructured data set would be

r	Individual i	Cluster j	y_{rij}	AGE	C1	C2	C1*AGE	C2*AGE
1	1	1	1	20	1	0	20	0
2	1	1	0	20	0	1	0	20
1	2	1	0	25	1	0	25	0
2	2	1	0	25	0	1	0	25
1	3	2	0	30	1	0	30	0
2	3	2	1	30	0	1	0	30

where C1 and C2 are dummy variables such that

$$C1 = \begin{cases} 1 & \text{if } r = 1 \\ 0 & \text{else} \end{cases}$$

$$C2 = \begin{cases} 1 & \text{if } r = 2 \\ 0 & \text{else} \end{cases}$$

C1, C2, C1*AGE and C2*AGE are fitted as covariates in the model and their coefficients are α_1, α_2, β_1 and β_2 in (2.19) respectively. The original multinomial response has been expanded to two binary responses per individual, y_{1ij} and y_{2ij}. The level 1, 2 and 3 identifiers needed to specify the hierarchical structure are respectively r, 'Individual' and 'Cluster'.

2.8.3 Estimation of the Multilevel Multinomial Logit Model

As for the multilevel binary logit model, the expression for the unconditional likelihood formed by 'integrating out' the random effects $\mathbf{u}_{rj}$ is intractable and an approximation must be used. Again a linear approximation for π_r may be obtained using a Taylor series expansion (Goldstein 1995) and then IGLS or IRLS is applied to the linear model. At iteration m, expanding about $\beta_r = \beta_r{}^{(m-1)}$ and $\mathbf{u}_r = 0$, the approximation is (omitting subscripts ij)

$$\begin{aligned}\pi_r &\simeq \pi_r^{(m-1)} + \sum_{k=1}^{p} x_{rk}(\beta_{rk} - \beta_{rk}^{(m-1)})\frac{\partial \pi_r}{\partial x_{rk}\beta_{rk}}(\eta_r{}^{(m-1)}) \\ &+ \sum_{k=1}^{q} z_{rk}u_{rk}\frac{\partial \pi_r}{\partial z_{rk}u_{rk}}(\eta_r{}^{(m-1)}),\end{aligned} \tag{2.20}$$

where $\eta_r{}^{(m-1)} = \mathbf{x}_r'\beta_r^{(m-1)}$.

For the logit link,

$$\frac{\partial \pi_r}{\partial x_{rk}\beta_{rk}} = \frac{\partial \pi_r}{\partial z_{rk}u_{rk}} = \pi_r(1-\pi_r).$$

Therefore writing $\pi_r^{(m-1)}(1-\pi_r^{(m-1)}) = w_r{}^{(m-1)}$, the approximate model is

$$y_r = \pi_r + e_r \simeq \pi_r{}^{(m-1)} + w_r{}^{(m-1)}\mathbf{x}_r'(\beta_r - \beta_r^{(m-1)}) + w_r{}^{(m-1)}\mathbf{z}_r'\mathbf{u}_r + e_r.$$

Rearranging this expression yields

$$y_r^* = \mathbf{x}_r^{*\prime}\beta_r + \mathbf{z}_r^{*\prime}\mathbf{u}_r + e_r,$$

where

$$\begin{aligned}y_r^* &= y_r - \pi_r{}^{(m-1)} + w_r{}^{(m-1)}\mathbf{x}_r{}'\beta_r{}^{(m-1)}, \\ \mathbf{x}_r^* &= w_r{}^{(m-1)}\mathbf{x}_r, \\ \mathbf{z}_r^* &= w_r{}^{(m-1)}\mathbf{z}_r.\end{aligned}$$

$y_r{}^*$, $\mathbf{x}_r^*$ and $\mathbf{z}_r^*$ are the working dependent variable and covariates in the fixed and random part respectively.

For a multinomial model, $\mathrm{var}(y_{rij}) = \pi_{rij}(1-\pi_{rij})$ and $\mathrm{cov}(y_{rij}, y_{r'ij}) = E(y_{rij}y_{r'ij}) - E(y_{rij})E(y_{r'ij})$. Since only one of y_{rij} and $y_{r'ij}$ can equal one for any individual, $E(y_{rij}y_{r'ij}) = 0$. Hence, $\mathrm{cov}(y_{rij}, y_{r'ij}) equals - \pi_{rij}\pi_{r'ij}$.

To produce this multinomial covariance structure, an extra set of covariates needs to defined at level 1 and 2. The model can be written

$$y_{rij} = \pi_{rij} + \gamma_{1ij}z_{1rij}' + \gamma_{2ij}z_{2rij}' + e_{rij}z_{3rij}',$$

where $z_{1rij}' = \frac{\pi_{rij}}{\sqrt{2}}$; $z_{2rij}' = -\frac{\pi_{rij}}{\sqrt{2}}$; $z_{3rij}' = \sqrt{\pi_{rij}}$; $E(\gamma_{1ij}) = E(\gamma_{2ij}) = E(e_{rij}) = 0$; $\mathrm{var}(\gamma_{1ij}) = \mathrm{var}(\gamma_{2ij}) = 0$; $\mathrm{var}(e_{rij}) = 1$; $\mathrm{cov}(\gamma_{1ij}, \gamma_{2ij}) = 1$; and $\mathrm{cov}(\gamma_{1ij}, e_{rij}) = \mathrm{cov}(\gamma_{2ij}, e_{rij}) = 0$.

To allow for extra-multinomial variation, $\text{cov}(\gamma_{1ij}, \gamma_{2ij})$ and $\text{var}(e_{rij})$ can be different from 1, but constrained to be equal.

As for the binary logit model, the MQL approximation in (2.20) can be extended to a PQL approximation using a second-order Taylor series expansion about $\beta_r = \beta_r^{(m-1)}$ and $\mathbf{u}_r = \mathbf{u}_r^{(m-1)}$. The second-order terms involve the calculation of the second derivatives of π_r which for the multinomial logit model are

$$\frac{\partial^2 \pi_r}{\partial(x_{rk}\beta_{rk})^2} = \frac{\partial^2 \pi_r}{\partial(z_{rk}u_{rk})^2} = \pi_r^2(1 - \pi_r).$$

MLn macros are available for fitting the multilevel multinomial logit model using MQL or PQL estimation procedures with either first- or second-order approximations. The macros set up the multinomial covariance structure and, at each iteration, recalculate first and second derivatives of π_r and carry out the necessary data transformations to obtain the working variables y^*, $\mathbf{x}^*$ and $\mathbf{z}^*$. MLn can also perform the data manipulations necessary to create the multivariate structure described in Section 2.8.2.

2.9 Hypothesis Testing and Model Selection in Nonlinear Multilevel Models

This section begins with a discussion of hypothesis testing for both fixed and random parameters in multilevel models. This is followed by some practical guidelines for fitting multilevel models to discrete response data, including suggestions for speeding up convergence and overcoming convergence problems.

Suppose we wish to test the null hypothesis $\text{H}_0 : \beta = 0$ against the alternative $\text{H}_1 : \beta \neq 0$, where β is a parameter in the fixed part of a linear or nonlinear model. A common approach is to carry out a likelihood ratio test in which the likelihood ratio or deviance statistic is calculated as

$$\text{D}_{01} = -2\log(\lambda_0/\lambda_1),$$

where λ_0 is the likelihood under H_0, and λ_1 is the likelihood under H_1. D_{01} is compared with a χ^2 distribution on r degrees of freedom, where r is the difference in the number of parameters fitted under the two models (in the above example, r=1).

Sometimes, however, we are interested in linear combinations of parameters. For example, to test whether two parameters β_1 and β_2 are equal, we would test the hypothesis $H_0 : \beta_1 - \beta_2 = 0$. Goldstein (1995) describes a more general hypothesis testing procedure that allows such contrasts to be tested. Let $\mathbf{C}$ be a $r \times p$ 'contrast' matrix, so that the hypothesis test for the linear combinations of the model parameters can be written in the form

$$H_0 : \mathbf{C}\beta = \mathbf{k}.$$

For example, suppose $\beta = (\beta_0, \beta_1, \beta_2)$ and we wish to test $H_0 : \beta_1 = \beta_2$ (i.e. $H_0 : \beta_1 - \beta_2 = 0$). Then this test can be written in the form

$$H_0 : \mathbf{C}\beta = \mathbf{0},$$

where $C = (0, 1, -1)'$. To carry out the test, we form the test statistic

$$R = (\mathbf{C}\hat{\beta} - \mathbf{k})'[\mathbf{C}(\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{C}']^{-1}(\mathbf{C}\hat{\beta} - \mathbf{k}).$$

This is compared with a χ^2 distribution on r degrees of freedom, where r is the number of linear contrasts to be tested. Recall that $\hat{\Sigma}$ is the estimated covariance matrix for the response vector $\mathbf{y}$ and $(\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X})$ is the estimated covariance matrix for the fixed parameters β. This test can be performed by MLn. The user specifies $\mathbf{C}$ and $\mathbf{k}$ and MLn calculates R with the corresponding p-value.

For continuous response models, likelihood ratio tests may be used to test the significance of the random effect variances and covariances. However, Goldstein (1995) states that likelihood ratio tests are unreliable for binary response data and recommends the use of an approximate chi-squared test for linear contrasts of the form

$$H_0 : \mathbf{C}\theta = \mathbf{k},$$

where θ is a vector containing the variances and covariances of the random effects. Again, the appropriate test statistic and p-value can be computed in MLn.

In the remainder of this section, some strategies for model selection and model fitting using MLn are discussed, with particular reference to discrete response data. For continuous response data, convergence tends to be very fast and convergence problems are rarely encountered. Convergence is usually much slower for nonlinear models, especially for the multinomial logit model (Section 2.8.3). However, the number of iterations required may be considerably reduced by specifying alternative starting values to the OLS estimates used by MLn. A straightforward way of obtaining suitable starting values is to first fit the corresponding single-level model. These estimates can then be used as initial values for the fixed parameters in the multilevel model. The strategy of fitting single-level models prior to carrying out a full multilevel analysis can also be a useful aid to initial model screening, since each iteration of the multilevel multinomial model estimation procedure is very slow.

In some situations, particularly if the second-order PQL approximation (Section 2.7.2) is used, convergence may not be achieved unless suitable starting values are used. If convergence problems occur, one strategy which is often successful is to begin by fitting a first-order MQL model (Section 2.7.1), using single-level estimates as starting values. The first-order approximation yields initial values for both the fixed and random parameters. This model can then be extended to a first-order PQL or second-order MQL approximation, which can be further extended to the second-order PQL model.

Chapter 3

Contraceptive Choice in Bangladesh

3.1 Introduction

Since the mid 1970s, there has been a rapid decline in fertility levels in Bangladesh. The total fertility rate (TFR) dropped from 6.3 in 1971-75 to 5.1 in 1984-88, and has continued to fall to 3.4 in 1991-93 (Mitra et al. 1994). The major determinant of the fertility decline is the large increase in contraceptive use over the past 20 years. Contraceptive prevalence rose fourfold between 1975 and 1989, from 8% to 31%, and the most recent estimate for 1993-94 is 45% (Mitra et al. 1994).

Early demographic transition theories (e.g. Notestein 1945) asserted that a country's fertility can only start to decline after socioeconomic development has led to lower infant mortality. Only then would couples wish to limit their family size and adopt contraception. The remarkable decline in fertility observed in Bangladesh, however, has taken place in an extremely unfavourable setting: Bangladesh remains a very poor, underdeveloped and predominantly rural country. Most indicators of socioeconomic development thought to erode the demand for children, including levels of school enrolment, women's employment and women's status, have been fairly static during the period of fertility decline (Cleland and Streatfield 1992). Therefore, in the absence of any large changes in the other determinants of fertility such as age at marriage, Cleland et al. (1994) conclude that the only convincing explanation for the sharp rise in contraceptive use and commensurate fertility decline is the government's strong family planning programme.

The first official government family planning programme in Bangladesh began in 1960 and was mainly clinic-based and implemented through existing health care facilities on a limited scale. This was followed in 1965 by a field-orientated programme with a strong information and education component. These early campaigns were interrupted by the 1971 war of liberation and the economic crisis which followed a spate of natural disasters in the early 1970s. In 1975, the government identified the control of the country's rapid population growth as its top priority and initiated the first integrated health and family planning programme. One of the most important components of the programme was the introduction of female family welfare assistants (FWA) at the grassroots level who visit each household every two months to discuss family planning and deliver pills and condoms. This feature of the programme has proved particularly successful since the low level of female autonomy and mobility in Bangladesh prevents many women from attending clinics to obtain contraception. In addition, family welfare centres (FWC) were established in each *union* (a cluster of villages) where injectables and IUD insertions are available. Sterilisations are performed at *thana* (sub-district) health complexes. Since 1981, the family planning programme has been further strengthened with the involvement of non-governmental organisations. More recently, in 1993, the number of FWAs was increased to achieve wider coverage, particularly in rural areas.

Despite the large decline in fertility, substantial regional variation in contraceptive use and fertility within Bangladesh persists. Most studies have focused on the individual-level factors affecting contraceptive adoption, although many have observed differences between the four administrative divisions (Kamal and Sloggett 1993; Shahidullah and Chakraborty 1993). In particular, Chittagong division shows consistently low levels of contraceptive use and high fertility. Rashid and Ali (1993) attempt to explore this further with a district-level analysis of contraception in Chittagong. They compare contraceptive prevalence rates for each district and conclude that there is little district-level variation within Chittagong. They then carry out an individual-level regression analysis to determine whether the low level of contraceptive use in Chittagong can be explained by individual-level demographic, socioeconomic and cultural characteristics, but find that none significantly reduce the variation between Chittagong and other divisions.

Contraceptive behaviour is influenced by a range of factors acting at individual, village and higher levels. In this chapter, a multilevel approach is used to examine the impact of variables both at the individual level and at higher levels of aggregation on contraceptive choice. Two higher levels are considered: community (represented by sampling cluster in this analysis) and district. Those living in the same community are likely to have similar access to family planning services because of their geographical closeness. Other factors operating at this level which are likely to have an impact on contraceptive behaviour are the attitudes of local leaders and other members of the community towards family planning. Negative attitudes as well as discouraging the use of family planning can also have a detrimental effect on the programme, which exacerbates the low levels of contraceptive use. For instance, the social conservatism in districts in Chittagong makes it difficult to recruit and retain good family planning staff and, therefore, women in these areas are less likely to be visited by a family planning worker (Cleland and Streatfield 1992). Districts have a different level of homogeneity that is more cultural in nature. A cultural identity is shared by the people of the same district that is manifest in behavioural patterns. An example of where district cohesion is strongly expressed is in marriage markets—families prefer to find grooms within the district, presumably because of an assumption of cultural similarity. In this analysis, a three-level model is used to examine the extent of clustering in contraceptive behaviour at the community and district levels.

3.2 The Bangladesh Fertility Survey 1989

The data for the analysis are from the Bangladesh Fertility Survey (BFS) of 1989 (Huq and Cleland 1990)—a nationally representative sample survey of ever-married women aged less than 50. Respondents were selected using a two-stage cluster sample design. In the first stage, 270 clusters—100 urban and 170 rural—were selected with probability proportional to size. Clusters are primary sampling units (PSU) defined by the National Census of 1981, and correspond approximately to villages in rural areas and *mohalla* (neighbourhoods) in urban areas. In the next stage, households within each cluster were listed and a sample was selected with probabilities inversely proportional to the first-stage selection probabilities to yield a sample which was self-weighting within urban and rural stratum. Five out of 64 districts were excluded from the sampling scheme—Meherpur, Natore, Khagrachori, Rangamati and Bandarban—mainly due to access problems. A total of 11,905 women were successfully interviewed.

The survey, in addition to the standard set of questions on fertility and related behaviour and demographic and socioeconomic characteristics, included questions on religious practice and women's status. Two questions on religious practice were asked—whether the respondent prayed every day and how strictly they observed religious practices compared to other households in the locality. Several other aspects of women's lifestyles are captured in a series of questions concerning women's independence, their mobility and their role in family decision-making.

The analysis sample consists of 9,777 currently married women. All women who were pregnant or postpartum amenorrheic at the time of the survey are excluded from the analysis.

3.3 Methodology

3.3.1 A Multilevel Multinomial Model of Contraceptive Choice

A multilevel multinomial logit model is used to analyse the determinants of current contraceptive choice. Contraceptive method is categorised as 1) sterilisation (female or male), 2) modern reversible (pill, condom, IUD or injectables), 3) traditional, and 4) none. A multinomial formulation is used in order to test for differences in the effects of covariates and in the extent of cluster- and district-level variation on the use of one type of method as opposed to another.

Let y_{ijk} be the multinomial response for individual i in cluster j in district k. As described in Section 2.8.2, this is converted to a series of three binary responses, $\{y_{rij}\}, r = 1,2,3$, one for each method type. The fourth category 'non-use', is selected as the baseline. The general three-level multinomial model can be written

$$y_{rijk} = \pi_{rijk} + e_{rijk}$$

where

$$\log\left(\frac{\pi_{rijk}}{\pi_{4ijk}}\right) = \mathbf{x}'_{rijk}\beta_r + \mathbf{z}'_{urijk}\mathbf{u}_{rjk} + \mathbf{z}'_{vrijk}\mathbf{v}_{rk},\ r = 1,2,3.$$

π_{rijk} is the probability that individual i in cluster j in district k uses method type r and π_{4ijk} is the probability of non-use. The set of covariates $\mathbf{x}_{rijk}$ is usually the same for each contrast of a method type with non-use, though there is no such restriction in the general model. The above model is an example of a random coefficients model: $\mathbf{z}_{urijk}$ and $\mathbf{z}_{vrijk}$ are sets of covariates, usually subsets of $\mathbf{x}_{rijk}$, the coefficients of which are permitted to vary randomly across clusters and districts respectively. The corresponding random effects vectors are $\mathbf{u}_{rjk}$ ($\sim N(0,\Omega_{ru})$) and $\mathbf{v}_{rk}$ ($\sim N(0,\Omega_{rv})$), representing unobserved factors operating at the cluster- and district-level. Cluster-level random effects for different contrasts of a method with non-use, $\mathbf{u}_{rjk}$ and $\mathbf{u}_{r'jk}$, $r \neq r'$, can be correlated as can pairs of method-specific district-level random effects. The individual error term is represented by e_{rijk} which is assumed to follow a multinomial distribution. The estimation of this model is described in Section 2.8.3.

3.3.2 Indices for Woman's Empowerment and Economic Status

Women's empowerment and household wealth are likely to be important predictors of fertility behaviour. However, these are quantities which are difficult to measure directly. Two aspects of woman's empowerment, mobility and role in household decision-making, are measured indirectly through a series of 13 questions in the BFS. Although wealth is a more easily defined concept than woman's empowerment, it is often difficult to obtain an accurate measure of income in a society such as Bangladesh where wages are paid both in cash and in kind. Also, in Bangladesh, a single measure of income recorded at the time of the survey will not reflect the seasonal variation in employment opportunities which leads to fluctuations in wages. Therefore, proxies for household wealth are obtained from more easily measured variables such as the number and type of possessions owned by the household. Rather than consider all 13 variables from the questions relating to women's empowerment, one would usually like to summarise this information in fewer variables which capture the important interrelationships between the variables. Some previous attempts to develop indices for empowerment and wealth have simply totalled the number of positive responses to the relevant questions (e.g. Kamal and Sloggett 1993). The problem with this approach is that each variable is given equal weight in the construction of the score, even though some variables are more powerful discriminators than others and should be weighted accordingly. A more appropriate methodology is latent class analysis, a technique applied predominantly in the social sciences to study the relationship between an observed set of categorical indicators and some underlying concept that is difficult to measure. Bartholomew (1987) and Andersen (1990) give detailed expositions of latent class models and more general latent-structure models.

Suppose there is a set of p observed binary variables $x_1, \dots, x_p$ which are thought to be indirect measures of the underlying concept of interest, represented by a single latent variable θ. In latent class models, θ is assumed to be a discrete variable with K classes. The number of classes, K, is chosen so that within each of the K groups, $x_1, \dots, x_p$ are mutually independent. In other words, we assume that the mutual association between the observed variables can be accounted for by a single categorical variable θ, so that $x_1, \dots, x_p$ are independent given θ. In this case, all the observed indicators are binary, but Bartholomew (1987) describes how latent class models can be extended to polychotomous indicators.

Let

$$\begin{aligned} \omega_{ij} &= \Pr(\text{positive response on variable } i \text{ for an individual in latent class } j) \\ &= \Pr(x_i = 1 | \theta = j), \quad i = 1, \dots, p; j = 1, \dots, K, \end{aligned}$$

and

$$\eta_j = \Pr(\theta = j), \quad j = 1, \dots, K,$$

the prior probability that a randomly chosen individual is in latent class j.

The joint probability density function of $\mathbf{x} = (x_1, \ldots, x_p)$ is

$$f(\mathbf{x}) = \sum_{j=1}^{K} \eta_j \prod_{i=1}^{p} \omega_{ij}^{x_i} (1 - \omega_{ij})^{1-x_i}$$

and the posterior probability that an individual with indicator vector $\mathbf{x}$ is in class j is thus

$$h(j|\mathbf{x}) = \frac{\eta_j \prod_{i=1}^{p} \omega_{ij}^{x_i} (1 - \omega_{ij})^{1-x_i}}{f(\mathbf{x})}.$$

Bartholomew (1987) describes how the EM algorithm (Dempster et al. 1977) can be used to obtain maximum likelihood estimates of η_j and ω_{ij}. Based on their response pattern $\mathbf{x}$, an individual can then be placed in the latent class for which the posterior probability of class membership $h(j|\mathbf{x})$ is greatest.

Latent class models can be fitted in the software PANMARK (Van de Pol et al. 1991) which uses the EM algorithm. The data are required to be in the form of a contingency table with the observed frequencies for every possible response pattern $\mathbf{x}$. The number of latent classes K which can be fitted depends on the number of cells in the contingency table; for identifiability of the model, the number of model parameters must be smaller than the number of cells (Andersen 1990). To test for goodness of fit, standard likelihood ratio or Pearson chi-squared tests can be used to compare the observed frequencies with those predicted by the model. However, some cells may need to be amalgamated to justify the distributional assumptions of the tests. As a guideline, Bartholomew (1987) recommends that the expected frequencies should be above five.

Having selected a model which fits the data, the next step is to interpret or 'label' the classes. By examining the estimates of ω_{ij}, the probability of a positive response for variable i for a person in class j, one can determine the most likely response pattern for individuals in class j. These response patterns can then be used to help to characterise each class.

Latent class analysis was used to construct indices for the two dimensions of women's empowerment measured in the survey, mobility and role in decision-making. To obtain a measure of a woman's overall mobility, the responses to a series of questions on whether she is able to perform the following activities alone were considered: walking around her neighbourhood, going outside the neighbourhood, talking to a man she does not know, going to the cinema, visiting a club or health centre, attending a political meeting or shopping. However, using all eight items led to a number of very small cell frequencies in the 2^8 table. This was mainly because for some variables the response was highly dependent on the response given for one or more of the other variables. For example, if a woman is not permitted to walk around her neighbourhood alone, it is extremely unlikely that she will be allowed to go outside her neighbourhood. For this reason, three of the indicators—going outside the neighbourhood, talking to an unknown man and attending a political meeting—were excluded. A model with four latent classes gave a very good fit (likelihood ratio (LR) test statistic=6.59, 8 d.f.). On examination of the estimates of ω_{ij}, it was found that a positive response to all five items was the most likely response pattern in one of the latent classes. This class was labeled 'high mobility'. Women in another class tended to respond negatively to all questions and this class was labeled 'very low mobility'. The remaining two classes, with a mixture of positive and negative responses, were labeled 'low mobility' and 'medium mobility'.

The BFS contains five questions concerning who makes the decisions on children's education, visits to relatives and friends, purchases, family planning and health. There were three response categories for each variable: wife only, husband only or joint. Attempts to fit a latent class model for decision-making were less successful than for mobility. To reduce the proportion of cells with zero counts, the response categories for wife only and joint were collapsed. However, the test statistic remained highly significant. This suggests that the observed indicators cannot be summarised by a single categorical variable which retains the interrelationships between the variables. Therefore it was decided not use an index constructed using latent class analysis, but to consider each variable separately in the model of contraceptive choice.

There are eight binary indicators on whether the household owns the following items: chair, bed, wardrobe, working radio, boat, cart or bicycle, electricity, pitcher or plate. Possession of a bed or plate was omitted as they are highly correlated with other indicators. Electricity was also excluded because it was highly dependent on whether the woman lives in a rural or urban area. Latent class analysis was performed using the remaining five items. The best model yielded three latent classes (LR statistic=9.61, 14 d.f.). On examination of the $\hat{\omega}_{ij}$ to determine the most likely response patterns in each class, it was possible to identify the overall economic status of each group and thus construct a covariate with three categories representing 'low', 'medium' and 'high' degrees of relative wealth.

3.4 Preliminary Analysis

Table 3.1 summarises the use of contraception by a number of socioeconomic and demographic characteristics. These variables have been chosen to reflect the factors that are likely to influence both the decision to practice contraception and the choice of method. Religious practice is represented by a variable with three categories: strict Muslims (Muslims who pray every day), non-strict Muslims (those who pray less frequently) and Hindus. The question on frequency of prayer was used in the analysis as this is a more objective measure of religiosity than the question on how strictly religious practices are observed relative to others in the locality. The social independence and possessions variables were constructed using the latent class analysis described in Section 3.3.2.

As a measure of family planning supply, a cluster-level variable indicating the presence of a FWC was constructed from the responses to the question "Is there a FWC in your union?". The original woman-level variable is potentially endogenous since it is possible that those using contraception, especially users of clinical methods, are more likely to know of the existence of a FWC than non-users. Non-users who have no interest in family planning may respond "no" or "don't know" even if there is a FWC in the area. If the variable were a true indicator of the existence of a FWC in the union, all women in a given PSU should respond in the same way, with perhaps some "don't know"s. In fact, there is substantial variation in responses within a large number of PSUs. Therefore it was decided to create a cluster-level variable from the individual-level responses as a measure of family planning service availability. After some exploration of different categorisations, a dichotomous variable was constructed: 1) 50% or more of women in a cluster responded "yes" when asked if there was a FWC in the locality, and 2) 50% or more responded "no" or "don't know". For clusters in the first category, it is reasonable to assume that there is a FWC in the locality. In the remaining clusters, it may be assumed that either there is no FWC in the area, or there is a FWC but it has not been successfully promoted since a large proportion of respondents do not know of its existence.

An inverted U-shaped relationship is found with increasing age in that young women and those older than 45 are least likely to use any method of contraception while those in their late twenties and early thirties are most likely to use a method. Sterilisation is most common among those aged 35-44, and modern reversible methods are most frequently used among those aged 25-34. This finding gives an indication of some use for spacing among modern method users, an indication that is supported by the results for number of living children which show modern reversible method use as reasonably high (16%) for women with only one child.

The association between socioeconomic characteristics and contraceptive use is as expected—urban dwellers are more likely to use contraceptives than their rural counterparts, educated women are particularly likely to use modern methods but relatively less likely to choose sterilisation, as are those women who have high levels of social independence and who are in the 'high' possessions group. Little effect is seen of religious practice among Muslims on contraceptive use, though there is some evidence of higher sterilisation and traditional use among Hindus. Finally, use of modern reversible methods is higher in clusters where the majority of respondents know of the existence of a FWC in the locality.

Areal-level factors can influence contraceptive use in a number of ways. For example, economic characteristics may play a role; women living in relatively wealthy areas may have an increased chance of using contraceptives beyond that which would be expected from knowledge of their individual characteristics. Wealthy areas may have better health care facilities and easier means of transport to them than poorer areas. Alternatively, the social and cultural characteristics of an area may act to influence the social norms within an area. For example, in a very conservative area, contraception may be less acceptable culturally and barriers may exist to the successful delivery of a family planning programme.

Table 3.1: Percentage of women who practice contraception, by selected individual characteristics, according to type of method used

Variable	Method type				n
	Sterilisation	Modern	Traditional	None	
Residence					
Rural	11.0	12.9	8.9	67.2	6988
Urban	9.7	32.6	9.5	48.2	2789
Age (years)					
< 24	2.3	15.7	6.3	75.6	3553
25-34	14.3	18.4	9.9	57.4	3436
35-44	20.1	9.9	12.2	57.8	2084
45+	11.6	2.2	8.0	78.2	704
No. living children					
0	1.5	6.2	4.1	88.2	1136
1	2.9	16.5	6.6	73.9	1727
2+	14.4	15.2	10.3	60.1	6914
Education					
None	12.7	9.5	7.7	70.0	6110
Lower primary	9.3	17.2	9.9	63.7	1303
Upper primary	6.6	22.3	13.6	57.4	868
Secondary +	4.6	37.0	11.8	46.7	1496
Religious practice					
Non-strict Muslim	11.4	14.5	7.0	67.1	4909
Strict Muslim	9.0	15.0	10.0	66.0	3531
Hindu	13.8	12.4	13.4	60.4	1337
Social independence					
1 (low)	7.8	10.1	8.6	73.6	5550
2	17.4	15.9	9.1	57.7	2127
3	15.7	24.5	9.0	50.8	1047
4 (high)	12.7	35.9	11.4	40.0	1053
Possessions					
1 (low)	13.4	9.5	7.3	69.7	5087
2	8.1	17.5	10.8	63.7	2856
3 (high)	6.1	29.5	12.2	52.3	1834
Family welfare centre in locality[a]					
No	9.3	11.2	9.0	70.6	2311
Yes	11.5	15.6	8.9	64.0	7466
Overall	10.9	14.4	8.9	65.8	9777

[a] A cluster-level variable constructed from individual responses. See text for details

Figures 3.1 and 3.2 show the distribution of contraceptive use across districts (the data for three districts—Narail, Thakurgaon and Nowabanj—were too sparse to permit reliable estimates and were merged with adjacent districts). The box plots show considerable inter-district variation in contraceptive prevalence for each method type, particularly for modern reversible methods. For example, sterilisation varies from 2% to 25%; and modern reversible-method use varies from 2% to 39%.

Figure 3.1: Box plots of district-level percentage contraceptive use by method type

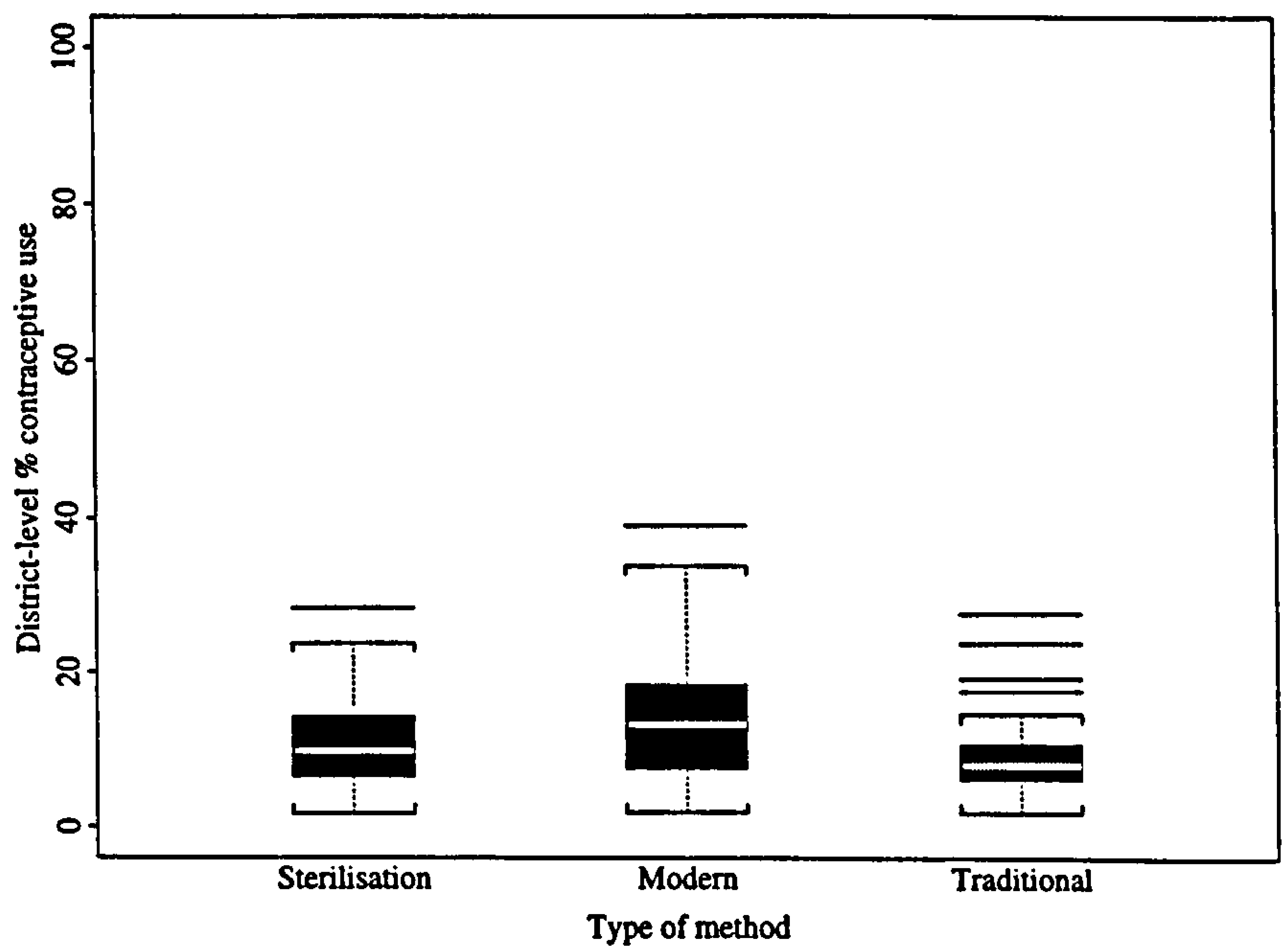
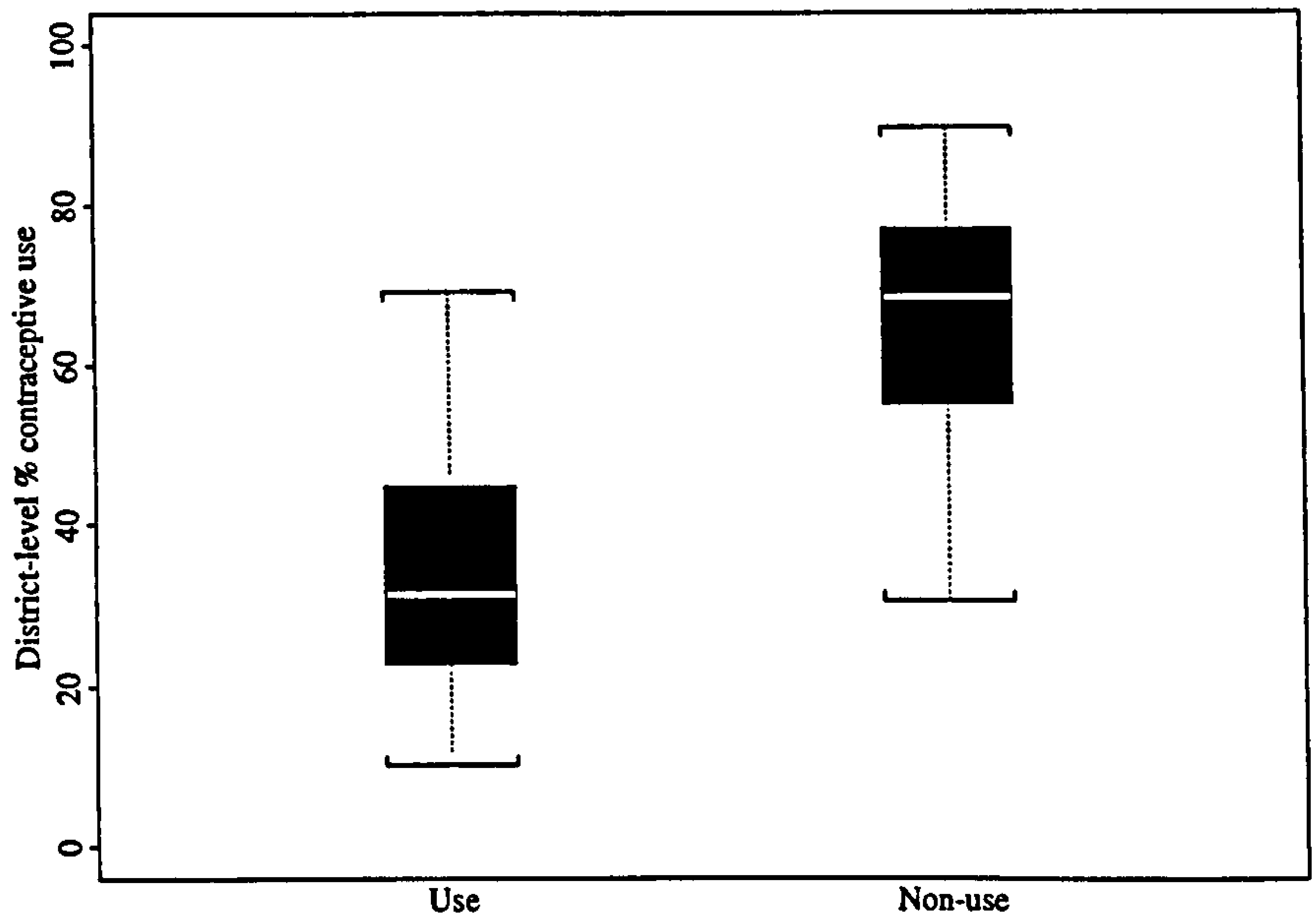


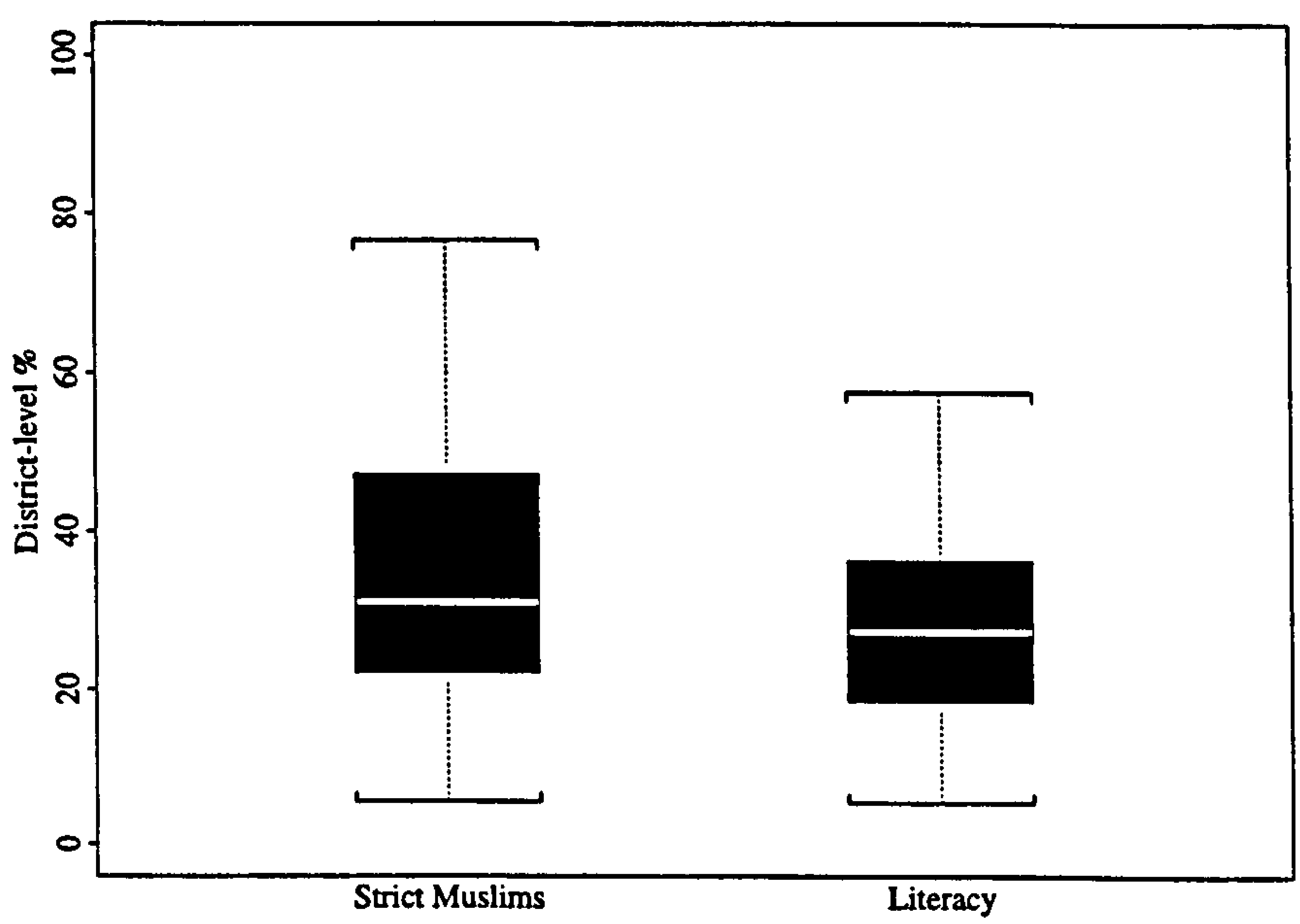
Figure 3.2: Box plots of district-level percentage contraceptive use



3.5 Results from the Multilevel Multinomial Model of Contraceptive Choice

The strategy adopted in the multilevel analysis was first to consider all the individual-level variables described in Table 3.1 as predictors of contraceptive use, controlling for homogeneity among women in the same cluster or district. A range of district-level variables were then introduced systematically into the model. The aim of this part of the analysis is to account for the unexplained variation at the district level. A number of district-level characteristics were considered: religious practice (as measured by the proportion of Muslims who pray every day), literacy, female autonomy (as measured by the proportion of women in the 'high' category of the mobility variable), toilet facilities, source of drinking water, and economic status (as measured by the proportion of households in the 'low' category of the possessions variable). These were constructed by aggregating individual-level data across districts. Only religion and literacy were found to be influential predictors of contraceptive use. Figure 3.3 shows the district-level variation in the proportion of Muslims who pray every day and the proportion of women who are literate. A large amount of between-district variation can be seen, particularly in the proportion of strict Muslims which tends to be highest in districts in Chittagong division. In general, a negative relationship is observed between religious practice in a district and contraceptive use: the higher the proportion of Muslims who pray every day, the lower is the level of contraceptive use. A positive relationship is seen between literacy and contraceptive use.

Figure 3.3: Box plots of district-level percentages of strict Muslims and literacy



After adding the significant individual- and district-level variables to the fixed part of the model, each selected covariate was tested for inclusion in the random part of the model to test whether its coefficient varied randomly across PSUs or districts. A random coefficient can be thought of as an interaction between an observed covariate and unobserved factors operating at the PSU or district level. For instance, one might expect the influence of female independence to be weaker in clusters with a strong doorstep delivery programme which successfully overcomes the difficulties in obtaining contraception experienced by women with restricted mobility. By allowing the coefficient of the female independence variable to vary randomly across PSUs, it is possible to test for such effects. However, in this case, there is no evidence to suggest that this or any other coefficient varies randomly across PSU or districts. Therefore, it is possible to simplify the general model presented in Section 3.3 to the following random intercepts model

$$\log\left(\frac{\pi_{rijk}}{\pi_{4ijk}}\right) = \mathbf{x}'_{rijk}\beta_r + u_{rjk} + v_{rk}, \ r = 1, 2, 3.$$

where $u_{rjk} \sim N(0, \sigma^2_{ru})$ and $v_{rk} \sim N(0, \sigma^2_{rv})$.

In addition, none of the random effects at the PSU- or district-level were found to be correlated across different method types. In other words, neither $\text{cov}(u_{rjk}, u_{r'jk})$ or $\text{cov}(v_{rk}, v_{r'k})$ were significantly different from zero for $r \neq r'$. These nonsignificant covariance estimates are not presented in the subsequent results.

Table 3.2 displays the PSU- and district-level variances in contraceptive use for the most important models. The estimates of the individual-level covariates are almost unaffected by the addition of district-level variables and are therefore not presented. The first row of Table 3.2 gives the unexplained variation at the PSU- and district-levels after controlling for the individual-level characteristics alone. Subsequent rows show the unexplained variation after controlling for the effects of religious practice and literacy. As the district-level variables are introduced, there is a marked decline in the magnitude of the district-level variation for both the contrast between sterilisation and non-use and between modern reversible method use and non-use, particularly when religious practice is introduced into the model. Indeed, when both religious practice and literacy are added, the unexplained district-level variation for the contrast between sterilisation and non-use is halved.

Table 3.2: The effect of district-level variables, literacy and religious practice, on the PSU- and district-level variation

Characteristics included in model	Sterilisation vs. none		Modern vs. none		Traditional vs. none	
	PSU	District	PSU	District	PSU	District
	σ^2_{1u} (SE)	σ^2_{1v} (SE)	σ^2_{2u} (SE)	σ^2_{2v} (SE)	σ^2_{3u} (SE)	σ^2_{3v} (SE)
Individual only	0.52 (0.09)	0.34 (0.11)	0.31 (0.05)	0.27 (0.08)	0.20 (0.05)	0.15 (0.06)
Individual + literacy	0.52 (0.09)	0.33 (0.11)	0.31 (0.05)	0.26 (0.08)	0.20 (0.05)	0.15 (0.06)
Individual + religious practice	0.51 (0.09)	0.21 (0.08)	0.30 (0.05)	0.21 (0.07)	0.19 (0.05)	0.13 (0.05)
Individual + literacy + religious practice	0.52 (0.09)	0.15 (0.07)	0.30 (0.05)	0.18 (0.06)	0.20 (0.05)	0.12 (0.05)

Table 3.3 displays the parameter estimates and standard errors for the model including the district-level variables, which corresponds to the final row of Table 3.2. To aid interpretation, estimated probabilities are also presented (Table 3.4). These are calculated for each covariate in turn while holding all other covariates and the PSU and district random effects, u_{rjk} and v_{rk}, at their mean values. Since all covariates are categorical variables, the mean value corresponds to the proportion in a each category. The district-level variables measuring religious practice and literacy are fixed at national average proportions.

After controlling for number of living children, the impact of age on sterilisation is considerably reduced, though there remains a large increase in the likelihood of being sterilised after age 25. The high rate of sterilisation among women aged between 25 and 44 is probably a consequence of the policy of cash payments made to sterilisation clients for a period of more than 10 years before the time of survey. Reversible modern method use shows a negative relationship with age, and is particularly low among women aged 45 or more who have or perceived themselves to have a low risk of becoming pregnant.

Table 3.3: Parameter estimates (and standard errors) from the multilevel multinomial model of method choice, including district-level variables

Variable	Method type					
	Sterilisation/ None		Modern/ None		Traditional/ None	
	Est.	(SE)	Est.	(SE)	Est.	(SE)
Constant	-3.17	(0.29)	-1.81	(0.24)	-2.49	(0.22)
Region (base=rural)						
Urban	0.20	(0.14)	0.64***	(0.11)	0.03	(0.11)
Age (years) (base=<24)						
25-34	1.84***	(0.14)	0.002	(0.07)	0.42***	(0.10)
35-44	2.16***	(0.15)	-0.62***	(0.09)	0.64***	(0.11)
45+	1.24***	(0.19)	-2.35***	(0.23)	-0.35**	(0.17)
No. living children (base=2+)						
0	-2.12***	(0.29)	-1.95***	(0.24)	-1.25***	(0.16)
1	-1.10***	(0.16)	-0.61***	(0.08)	-0.53***	(0.12)
Education (base=none)						
Lower primary	-0.09	(0.11)	0.40***	(0.09)	0.10	(0.11)
Upper primary	-0.19	(0.15)	0.68***	(0.10)	0.59***	(0.12)
Secondary +	-0.31**	(0.15)	1.03***	(0.09)	0.61***	(0.12)
Religious practice (base=non-strict Muslim)						
Strict Muslim	0.05	(0.09)	0.04	(0.07)	0.26***	(0.09)
Hindu	0.43***	(0.12)	0.07	(0.11)	0.67***	(0.12)
Social independence (base=1 (low))						
2	0.98***	(0.09)	0.53***	(0.08)	0.22**	(0.09)
3	1.03***	(0.11)	0.76***	(0.09)	0.33***	(0.12)
4 (high)	1.21***	(0.13)	0.85***	(0.09)	0.49***	(0.12)
Possessions (base=1 (low))						
2	-0.44***	(0.09)	0.34***	(0.07)	0.31***	(0.09)
3 (high)	-0.61***	(0.13)	0.61***	(0.09)	0.48***	(0.11)
District-level variables						
Proportion who pray every day	-2.88***	(0.58)	-2.10***	(0.52)	-1.22***	(0.45)
Proportion of literate women	2.03***	(0.78)	1.92***	(0.71)	0.70	(0.62)
Random effect variances						
PSU-level σ^2_{ru}	0.51***	(0.09)	0.31***	(0.05)	0.19***	(0.05)
District-level σ^2_{rv}	0.17**	(0.08)	0.18***	(0.06)	0.12***	(0.05)

* p=0.10; ** p=0.05; *** p=0.01

Table 3.4: Estimated probabilities of contraceptive use by method type and individual- and district-level variables

Variable	Method type Sterilisation	Modern	Traditional	None
Residence				
Rural	0.056	0.127	0.097	0.719
Urban	0.061	0.214	0.088	0.637
Age (years)				
< 25	0.018	0.202	0.074	0.706
25-34	0.100	0.178	0.099	0.623
35-44	0.140	0.098	0.126	0.636
45+	0.074	0.023	0.062	0.841
No. living children				
0	0.014	0.039	0.048	0.902
1	0.033	0.125	0.078	0.764
2+	0.082	0.187	0.108	0.623
Education				
None	0.066	0.119	0.085	0.731
Lower primary	0.057	0.167	0.088	0.688
Upper primary	0.046	0.201	0.130	0.623
Secondary +	0.038	0.263	0.123	0.576
Religious practice				
Non-strict Muslim	0.055	0.149	0.080	0.716
Strict Muslim	0.056	0.150	0.101	0.693
Hindu	0.076	0.143	0.141	0.641
Social independence				
1 (low)	0.039	0.120	0.089	0.752
2	0.090	0.173	0.094	0.642
3	0.089	0.206	0.099	0.606
4 (high)	0.101	0.214	0.111	0.575
Possessions				
1 (low)	0.076	0.124	0.081	0.719
2	0.047	0.165	0.105	0.683
3 (high)	0.037	0.203	0.117	0.642
Overall average	0.058	0.149	0.095	0.698

Note: probabilities are calculated for each covariate in turn while holding all other covariates at average values and the PSU and district random effects at zero. The district-level variables for religious practice and literacy are fixed at national average proportions.

The impact of the number of living children on contraceptive use is as expected: for all method types, the probability of use increases with family size. The sex of surviving children was also considered to test whether the number of sons influences contraceptive use. After controlling for other individual-level characteristics, however, there was no evidence of an effect of the sex composition of children or of a difference in use between women with two children and those with more than two.

The effects of the socioeconomic variables, type of region of residence, education and economic status, on modern reversible method use are in the expected directions: urban women, educated women and those who have household items that indicate relative wealth, are all more likely to use a modern method. There is also evidence of positive relationships between education and economic status and traditional method use. The lower rates of sterilisation among relatively wealthy women has been observed by others (Cleland and Mauldin 1991; Kamal and Sloggett 1993). The introduction in 1976 of compensation payments to sterilisation acceptors is likely to have made male or female sterilisation a particularly attractive contraceptive option to those living in extreme poverty. During the period prior to the BFS, payments equivalent to one week's wage for a unskilled labourer were made to clients in addition to a *saree* for women or a *lunghi* for men. Cleland and Mauldin (1991) conclude from their survey of sterilisation clients that the financial payment was a contributing factor to the decision to become sterilised in a large majority of cases, though a dominant motive for only a small minority, and that this is the most plausible explanation for the link between poverty and sterilisation. The decrease in the probability of sterilisation among educated women is likely to be due to similar reasons.

There is a pronounced increase in use of sterilisation and reversible modern method use with social independence, and a slight though significant increase in traditional use. Women in the 'low' category have severely restricted mobility and are unlikely to be permitted to attend a family planning clinic unaccompanied. The system of purdah which promotes the seclusion of women constrains the movements and activities of women. These social norms prohibit women from travelling to FWCs or commercial outlets for family planning unless accompanied by a close male relative. Although the programme strategy of doorstep delivery of contraceptives was devised especially to address this issue, several studies have revealed very low visitation rates in some rural areas (Koenig and Simmons 1992; Rahman et al. 1993).

The effects of the religion variable show that Hindus experience significantly higher probabilities of sterilisation and traditional method use than Muslims. However, there is no significant difference in either sterilisation or reversible modern method use between strict and non-strict Muslims which implies that the extent to which an individual practices religion is not a major barrier to the acceptance of family planning.

In the initial models, women living in clusters in which more than 50% of respondents reported that there was a FWC in the area had an increased chance of using modern reversible methods. However, after controlling for individual-level demographic, socioeconomic and cultural characteristics, the cluster-level variable indicating the awareness of the presence of a FWC in the locality is no longer significant. This is consistent with the findings of Kamal (1995b) who found that the presence of a FWC in a cluster had no significant effect on modern method use. A more influential family planning supply factor is likely to be the frequency of visits of outreach workers. Kamal (1995b) found that the presence of a FWA in the cluster explained almost one third of the inter-cluster variation in contraceptive use. A high proportion of women are reliant on these home visits for obtaining contraceptive supplies since, as mentioned earlier, constraints on their mobility make visits to a static family planning clinic infeasible for many women.

The results in Table 3.3 show that a district's religious practice is strongly negatively correlated with all types of use, particularly sterilisation and modern reversible methods. The level of literacy in a district is positively related to modern method use, but not significantly correlated with traditional use. These findings are further illustrated in Table 3.5 which shows the expected probabilities of contraceptive use for a series of districts chosen to represent a range of religious practice and literacy. These have been calculated for average values of the individual-level covariates and the PSU and district effects, u_{rjk} and v_{rk}, are fixed at their average of zero. Therefore the estimated probabilities reflect purely the effect of district-level religious practice and literacy on contraceptive choice. The importance of both literacy and religious practice can be seen clearly. Moulvibazar is a district with low literacy and high religious practice and this is reflected in the very low rates of sterilisation and modern reversible method use. Comparisons of Chandpur and Nilphumari with Moulvibazar show the impact of higher levels of literacy and less frequent praying respectively. In these districts, both literacy and religious practice have a strong influence on levels of contraceptive use. For example, the probability of modern reversible method use in Nilphumari is almost three times greater than in Moulvibazar. Since there is no district with a combination of high literacy and low religious practice, Dhaka with a medium level of religious practice is chosen to represent the most favourable scenario from the estimated model which can be observed. The probabilities for Dhaka show that the combined effect of high literacy and medium religious practice leads to a further increase in contraceptive use for all types of method.

Table 3.5: Estimated probabilities of contraceptive use for selected districts with different levels of literacy and religious practice, by method type

	Method type			
District	Sterilisation	Modern	Traditional	None
Moulvibazar (Low literacy, high religious practice)	0.022	0.059	0.062	0.857
Chandpur (High literacy, high religious practice)	0.035	0.138	0.087	0.739
Nilphumari (Low literacy, low religious practice)	0.054	0.146	0.097	0.704
Dhaka (High literacy, medium religious practice)	0.077	0.207	0.104	0.611

Note: Individual characteristics are fixed at average values and PSU and district effects are fixed at 0. No district has a high level of literacy and a low level of religious practice.

The joint effect of literacy and religious practice on district-level rates of contraceptive use is further illustrated in Figures 3.4 to 3.7. These figures show estimates of the district-level residuals v_{rk} for sterilisation (Figures 3.4 and 3.5) and reversible modern method use (Figures 3.6 and 3.7), with 95% confidence bars to allow for the uncertainty in these estimates. Simultaneous confidence intervals have been constructed to allow for multiple comparisons between pairs of districts. If the 95% confidence intervals for any pair of districts overlap, then there is no evidence at the 5% level of any difference in use between those two districts. The estimates for a district with an 'average' level of use would lie on the horizontal line at v_{rk}=0, while 'above average' and 'below average' districts would lie above and below the line respectively. This approach was proposed by Goldstein and Healy (1995) and is described in more detail in Section 2.5.2. Figure 3.4 shows the estimated district-level effects for the contrast of sterilisation with non-use when only individual-level characteristics are controlled. In this model, there is clearly a substantial amount of unexplained inter-district variation in sterilisation rates. Most districts in Chittagong division have below average sterilisation rates, but there is no evidence of any district-level variation within Chittagong as all confidence intervals overlap. Most inter-district variation occurs in the other divisions, Dhaka, Khulna and Rajshahi. This variation is considerably reduced in Figure 3.5 which displays the estimated district effects and 95% confidence intervals after controlling for district-level religious practice and literacy. After the addition of the district-level variables, there is no significant inter-district variation in Chittagong and Dhaka divisions and only two districts in Khulna have significantly different sterilisation rates. The inter-district variation in Rajshahi, though decreased, is still apparent. Figures 3.6 and 3.7 show similar graphs for modern reversible method use before and after controlling for district-level religious practice and literacy. They show an effect that is less marked but still notable.

Figure 3.4: District effects v_k for sterilisation with simultaneous 95% confidence intervals, by division: individual characteristics only

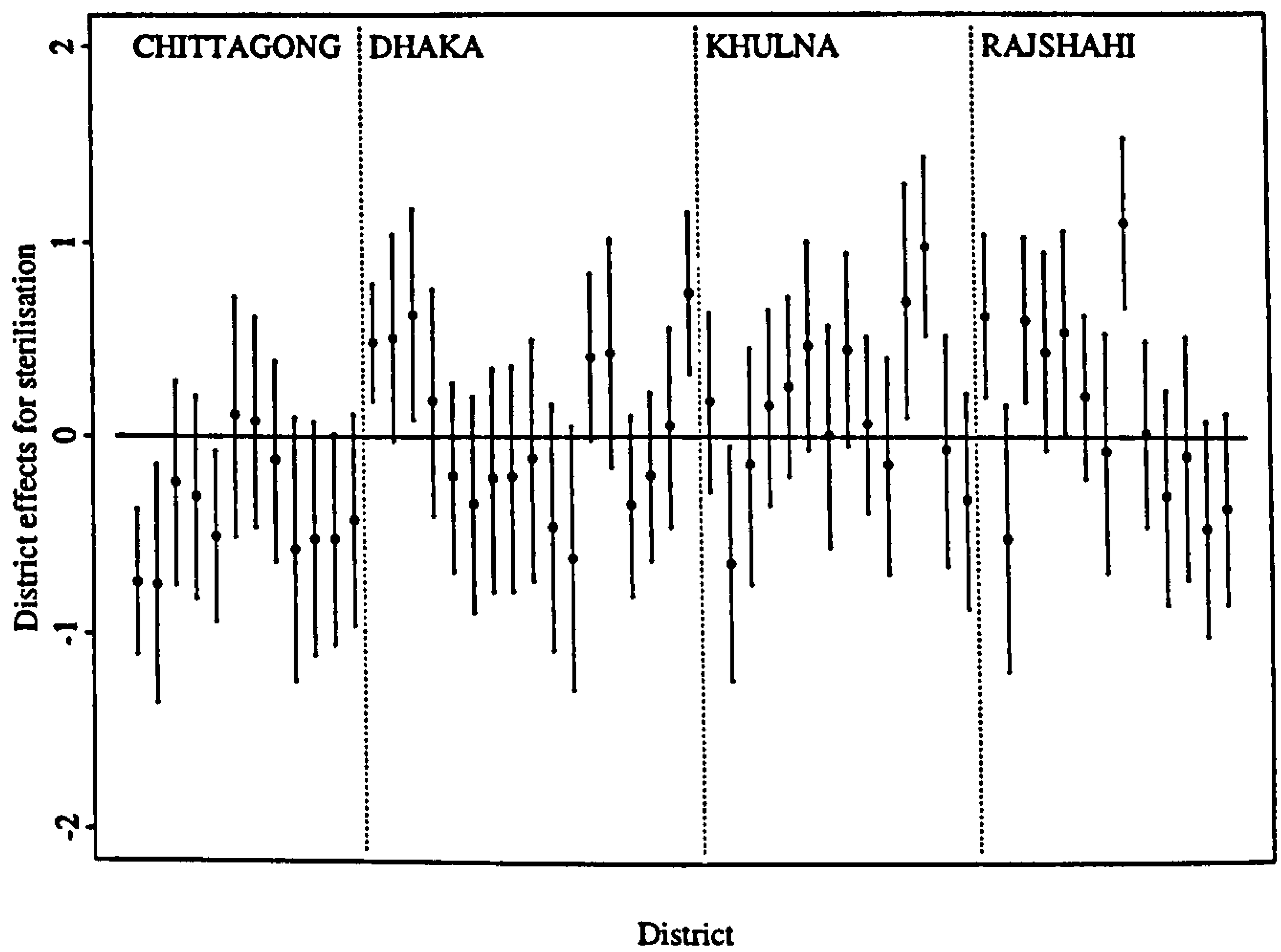


Figure 3.5: District effects v_k for sterilisation with simultaneous 95% confidence intervals, by division: individual and district-level characteristics

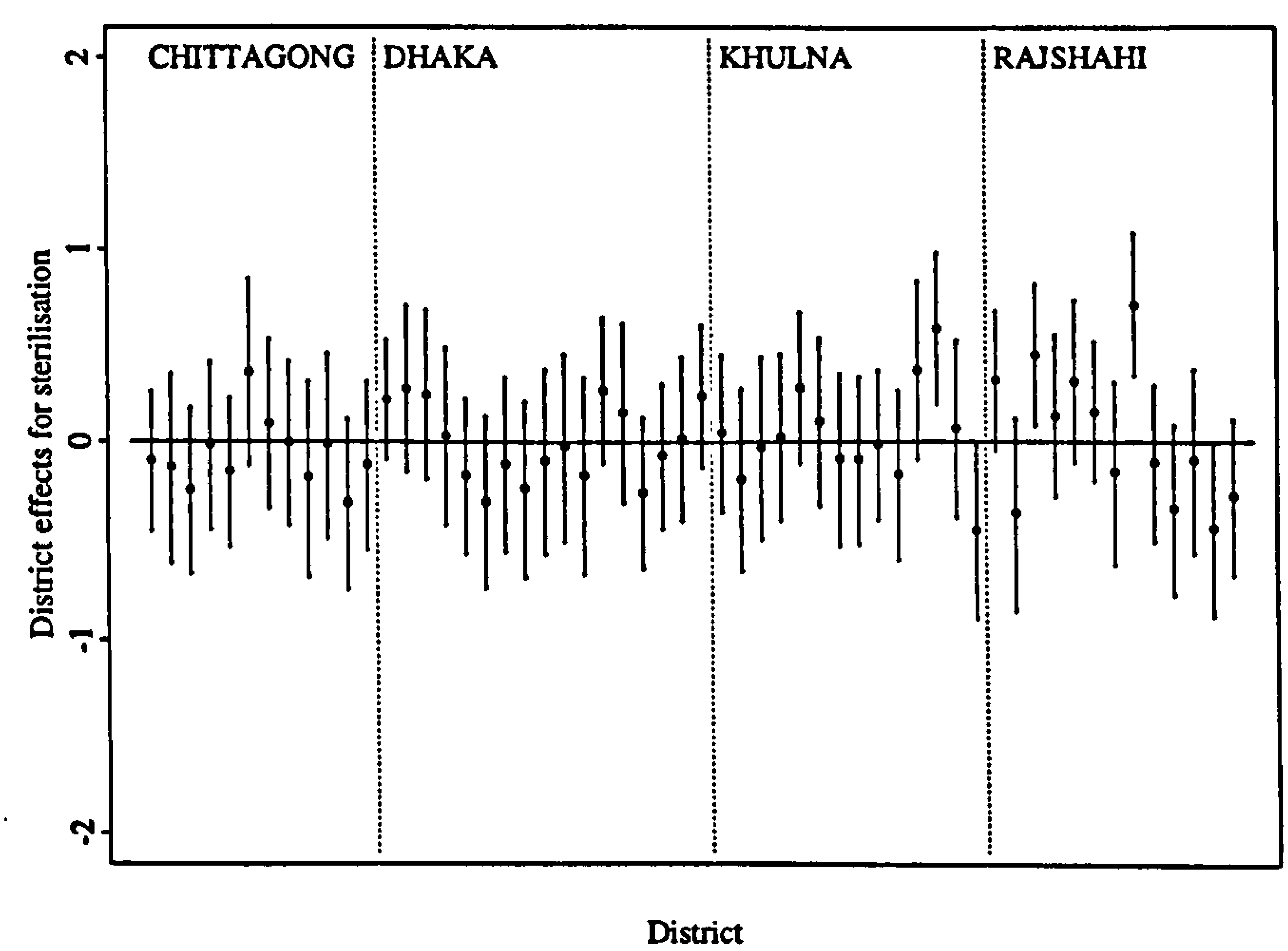


Figure 3.6: District effects v_k for modern methods with simultaneous 95% confidence intervals, by division: individual characteristics only

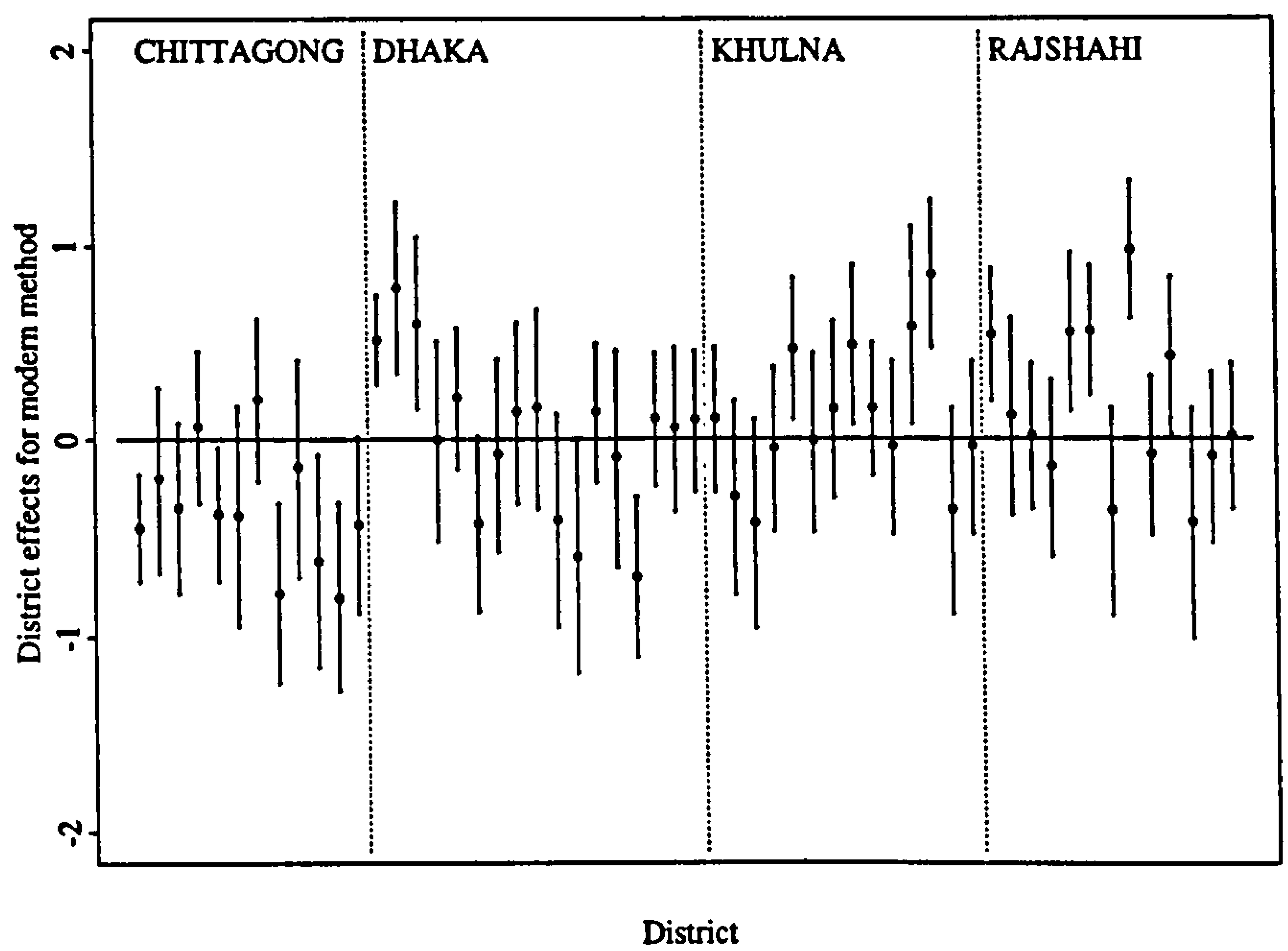
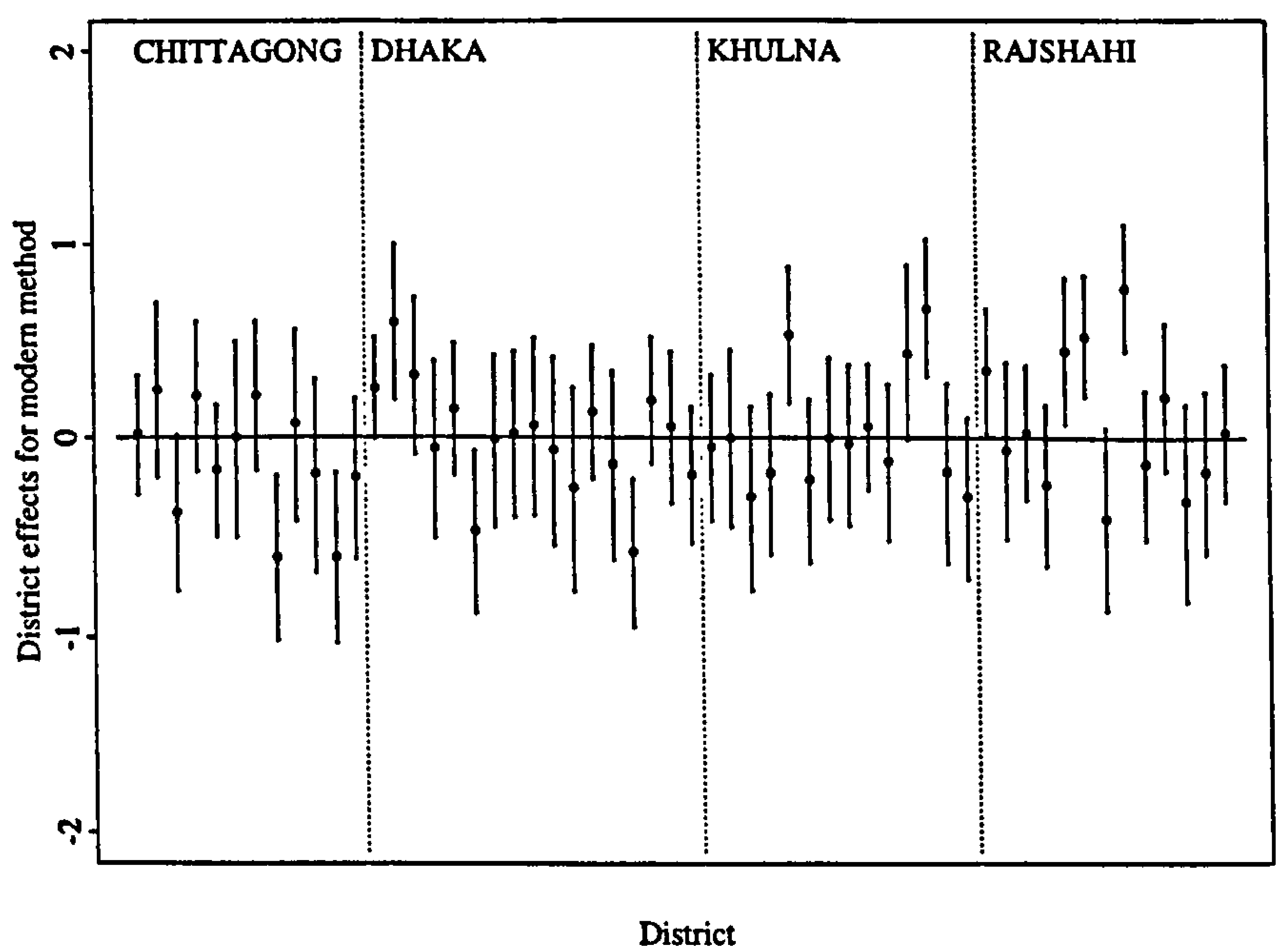


Figure 3.7: District effects v_k for modern methods with simultaneous 95% confidence intervals, by division: individual and district-level characteristics



An important inference from this analysis is that religious practice among Muslims operates at the district level as opposed to the individual level as a barrier to contraceptive use. This inference is now explained. The Quran is ambiguous on the issue of contraception, and does not forbid its use explicitly. On such issues, Muslim scholars will usually turn to analogic reasoning or consensus building and this leads to a variety of interpretations among religious leaders (Mussallam 1983). For example, some scholars have argued that because withdrawal is allowed, contraceptive drugs and abortion should also be accepted. Sterilisation is the most controversial method: some hold the view that if long-term use of temporary methods is allowed then permanent methods should also be permitted, but many view sterilisation as unacceptable under any circumstances (Suchedina 1990). This would explain the large impact of district-level religious practice on sterilisation rates. Since consensus building operates at the level of the group rather than of the individual, the opinion and behaviour of the larger group determines individual choice rather than the individual's own beliefs. In Bangladesh, where the use of contraception is driven by a programme that relies on the diffusion of ideas, opposition from religious leaders can be a barrier to programme effectiveness. In an alternative setting where the diffusion of ideas, and hence changes in contraceptive behaviour, is driven by developing and changing economic circumstances, religious practice would likely play a lesser role at the group level, while remaining influential at the individual level.

The influence of religion may also have an indirect effect on contraceptive use by working through women's status. There is a tendency for increasing numbers of religious communities to reinforce the practice of female seclusion which inhibits women's freedom of movement within the community, maintaining their low status. Evidence exists that individual women's status within the household, in terms of their autonomy, mobility, role in decision-making and authority within the family, are strong influences on fertility and other behaviour (Ali 1993; Amin et al. 1993; Cleland et al. 1994; Kamal and Sloggett 1993). Indeed, one of the findings of the present analysis is that women with a low level of mobility are considerably less likely to be users of sterilisation and other modern methods of contraception.

In Bangladesh, the family planning programme follows a strategy of co-opting religious leaders in terms of motivation and education. This strategy appears to have been successful in overcoming initial opposition to contraception. However, with evidence of increasing opposition to the practice of contraception in several conservative Islamic countries, including Saudi Arabia, Iran and Sudan (Amin and Hossain 1995), it may be more difficult to influence local leaders in the future. A recent religious interpretation by a Bangladeshi Islamic scholar, although supported by the government family planning programme to enlist the support of religious leaders, stated his reservation about the widespread availability of contraception (Amin and Hossain 1995). Even if religious leaders are currently amenable to family planning, their acceptance may wane after the connections between contraceptive use and women's changing roles become more evident.

Education has an impact on contraceptive use both at the individual level and at the district level. The strong effect of the proportion of literate women in a district provides support for what Caldwell (1980) has termed the influence of mass education on fertility decline. Traditionally, children are employed at an early age and make a valuable contribution to the household economy. If a child attends school, however, their potential to work inside and outside the home is reduced and the flow of wealth is from the parents to the child. In a relatively well educated society, the high cost of raising children and equipping them for school leads to a lower demand for large families and an increase in the level of contraceptive use in order to limit family size. Caldwell (1980) also hypothesises that mass education increases the rate of cultural change as education steers children away from traditional beliefs. Thus an educated society is more amenable to innovative ideas such as contraception. In the present analysis, the impact of adult female literacy rather than children's education is explored. This is a measure of second generation mass education. Educated parents are likely to educate their children too and, because of the increased costs of rearing school children, cannot afford to have large families. Therefore, in an area with a high proportion of literate women, there is a shift in fertility desires and small families become the norm. In such an area, these ideas are likely to spread to less educated women whose attitudes and behaviour are influenced by that of their neighbours.

Individual- and areal-level socioeconomic and cultural factors are likely to affect not only the demand for contraception but also the supply of family planning. For example, in an area where religious conservatism is high or the status of women is low, it can be difficult to recruit and retain fieldworkers (Cleland and Streatfield 1992; Koenig and Simmons 1992). In these areas, female family planning workers travelling on their own to outlying areas and interacting with male workers are perceived to be violating the norms of behaviour for women and can experience harassment from others in the community. In remote rural areas with a low level of development, the lack of infrastructure makes regular household visits infeasible and female workers may be concerned about their safety. For these reasons, many posts remain unfilled—Koenig and Simmons (1992) report that in Chittagong division nearly half of the new family planning worker positions created in 1992 remained vacant. The level of female education in an area can also affect the implementation of the family planning programme. Family planning workers are required to possess a minimum of 10 years school education and, in an area of low literacy, the pool of suitably qualified applicants may be restricted.

There is also evidence of a large amount of unexplained variation within districts between PSUs. Factors operating at this level may include those relating to family planning supply. Although the family planning programme is organised at the district level, service delivery takes place locally in the village or neighbourhood. Family planning policy may be the same across a district, but implementation is likely to vary greatly within a district. There will be variation not only in the type of facilities available and the services they provide but in the quality of service provision. For instance, the frequency of doorstep delivery is likely to vary across villages, especially in remote areas. Kamal (1995b) found that a large proportion of the PSU-level variation in sterilisation and modern reversible method use could be explained by differences in the frequency of household visits made during the six-month period prior to the survey. In addition, although the knowledge that there was a FWC in the locality was found to be non-significant, factors relating to accessibility of the centre, the range of methods available and the quality of services are all potential sources of inter-cluster variation.

3.6 Summary

In this chapter, a multilevel multinomial logit model was used to analyse the factors affecting contraceptive choice in Bangladesh. Of particular interest is the extent of district- and community-level variation in acceptance rates for sterilisation, modern reversible and traditional methods. To allow for the possibility of unobserved district and community factors which could lead to clustering of outcomes for women within each level of the hierarchy, random effects corresponding to the district and cluster were incorporated in the model. A number of individual-level background characteristics were considered and of these the woman's age, number of living children, type of region of residence, education, religion, economic status and level of social independence were all found to have a significant impact on the decision to use a method versus none.

The main focus of the analysis is the substantial amount of variation between districts, and within districts between clusters, that remains unexplained by the covariates included in the model. In an attempt to account for some of this inter-district variation, a range of district-level characteristics were added to the model. Two of these—the proportion of Muslims who pray every day and the proportion of literate women—were found to be significant. Since the individual-level variable measuring the frequency of prayer among Muslims did not have a significant effect on modern method use, the importance of the district-level religious practice variable would indicate that the religiosity of a woman's neighbours and the attitudes of religious leaders in the area are more powerful predictors of her chance of accepting a modern method than her own level of religiosity. The joint impact of district-level religious practice and literacy leads to a large reduction in the unexplained district-level variation for the contrasts between sterilisation and non-use and between modern reversible method use and non-use. However, there remains considerable extravariation within districts between clusters, especially for modern methods. One likely explanation for this is unobserved differences in the availability and quality of family planning services which are implemented at a local level, roughly corresponding to clusters in this analysis.

Chapter 4

Discrete-Time Methods for the Analysis of Event Histories

4.1 Introduction

This chapter provides an outline of discrete-time methods for the analysis of event history data. We begin with a discussion of event history data and some of the problems encountered in their analysis. A description of the life table approach is then presented. This is followed by a section on discrete-time event history models, an extension of life table methodology which allows covariates to be incorporated into the analysis. In event history data, a number of complications can arise. The event of interest may occur more than once to an individual over time; there may be more than one type of event of interest; or some unobserved variation may exist because of omitted covariates. To cope with these situations, models for repeated events, competing risks and unobserved heterogeneity have been developed and are described in this chapter. In particular, multilevel models which account for both competing risks and unobserved heterogeneity at the individual and higher levels are presented.

4.2 Event History Data

In its simplest form, an event history is a longitudinal record of when events occurred to an individual. The aim of event history analysis is to study the patterns and causes of these events. In event history analysis, we analyse the length of time to the occurrence of some event of interest. Examples of event history data in the social sciences include birth histories, where we study the length of a birth interval until the event of interest 'giving birth'; contraceptive use histories, where we are interested in the duration of a use interval until the event 'discontinuation of use'; and employment histories, where we might wish to look at the length of an unemployment spell, that is, the time interval between becoming unemployed and the event 'entering employment'. In the medical sciences, the focus is usually on the survival time until a patient becomes ill or dies and the analysis of such event histories is called survival analysis. In either the social or medical sciences situation, the occurrence of an event indicates a transition from one state to another.

The simplest approach for the analysis of event history data would be to model whether the event of interest has occurred during a prespecified time period. For example, in a study of contraceptive discontinuation, one could use simple logistic regression techniques to model whether or not discontinuation has occurred within 12 months of starting use. The problem with this approach is that dichotomising the response variable in this way tends to be fairly arbitrary and valuable information on the timing of events is wasted. Therefore, techniques have been developed to model explicitly the actual length of time until the event, or the probability of the event occurring during a time interval. However, two particular features of event history data—censoring and time-varying covariates—make the use of standard methodology such as multiple linear regression inappropriate.

One characteristic feature of duration data is the presence of censored observations, that is, cases where the time to the event of interest is not observed. Censoring can occur for a number of reasons. For event history data collected in a survey, a time interval may have been interrupted by the interview date and therefore, it is not known when the event will occur in the future. In clinical trials, some patients may withdraw from the study because they move away or because they experience side-effects as a result of the treatment, or in some cases the trial may end before the event has been observed to occur. The above situations are examples of right-censoring and all that can be observed is the time to censoring. Usually censoring is assumed to be random or uninformative where the censoring mechanism is independent of the duration. Much less common, and more difficult to deal with, is left-censoring. This occurs when observation on an individual starts after entry into the risk period and the time of the start of the duration interval is, therefore, unknown.

Another feature of event history data is the presence of time-varying covariates. Since event histories are longitudinal in nature, it is likely that some of an individual's background characteristics, for example, income, marital status or employment status, will change over the observation period. Often covariates are measured only once, usually at the start of observation, and in such cases the analyst has no choice but to assume that they remain constant over time, even though this assumption may be unrealistic. If measurements are made at more regular intervals, however, it is important to incorporate this information into the model in order to obtain accurate estimates of the effects of these variables over time.

4.3 The Single-Decrement Life Table

A simple approach for the analysis of event history data, which takes into account right-censored observations, is to use life table methods. Life tables were originally used by demographers and actuaries to study mortality, but can be used to study the length of time to the occurrence of any type of event. To construct a life table, duration is broken down into intervals $[t, t+1)$, $t = 1, 2, 3, \ldots$, and for each interval the number of individuals at risk, number of events and number of censored observations is recorded. These can then be used to calculate quantities such as the probability of event occurrence in a given time interval and the probability that no event occurs before a specified time. In this section, the single-decrement life table is described where there is only one type of event or 'decrement' of interest. For consistency with the notation used in later sections of this chapter, event history analysis notation is used rather than conventional life table notation.

Let

$$\begin{aligned} n_t &= \text{no. at risk at start of } [t, t+1), \\ d_t &= \text{no. events in } [t, t+1), \\ c_t &= \text{no. censored in } [t, t+1). \end{aligned}$$

Then

$$\begin{aligned} n_1 &= n = \text{no. individuals in the population} \\ n_j &= n_{j-1} - d_{j-1} - c_{j-1}, \; j = 2, 3, 4, \ldots \end{aligned}$$

A fundamental concept in event history analysis is the hazard rate which is the probability of an event occurring in the interval $[t, t+1)$ given that an event has not already occurred. This is defined as

$$h_t = \Pr(\text{event occurs in } [t, t+1) | T \geq t),$$

where T is a discrete random variable taking nonnegative integer values, representing the time of event occurrence.

The hazard rate is estimated as

$$\hat{h}_t = \frac{d_t}{\text{no. exposed to risk in } [t, t+1)}.$$

However, since the exact timings of events and censorings within an interval $[t, t+1)$ are not usually known, the number of individuals exposed to the risk of an event occurrence must be estimated. It is usually assumed that censored cases were at risk for half the length of the interval in which they were censored. This is equivalent to the assumption that, conditional on their occurrence in the interval, the event time and censoring time are independently and uniformly distributed through the interval. This yields the estimate

$$\hat{h}_t = \frac{d_t}{n_t - 0.5c_t}.$$

The probability of 'surviving' to time t is

$$\begin{aligned} S_t &= \Pr(\text{event does not occur before } t) \\ &= \Pr(T \geq t) \\ &= \prod_{j=1}^{t-1} (1 - h_j), \end{aligned}$$

and the unconditional probability of event occurrence is

$$f_t = \Pr(\text{event occurs in } [t, t+1)).$$

h_t, S_t and f_t are related by the identity

$$h_t = \frac{f_t}{S_t}.$$

The cumulative probability that an event occurs at any time before t is

$$\begin{aligned} F_t &= \Pr(T < t) \\ &= \sum_{j=1}^{t-1} f_j \\ &= 1 - S_t. \end{aligned}$$

Since S_t, f_t and F_t can all be expressed in terms of h_t, they can be estimated by substitution of $\hat{h}_t$.

4.4 The Discrete-Time Event History Model

Although the life table approach is a very useful and simple technique, we often wish to determine the influence of covariates on the hazard rate. It is possible to examine differentials between subgroups of a population by calculating separate life tables for each subgroup. However, this approach quickly becomes infeasible if we wish to look at the effects of a large number of covariates simultaneously. This is because the number of observations within the subgroups defined by each combination of covariate values can be very small. In this section, an outline of a multivariate extension of the life table, the discrete-time event history model, is described. Allison (1982) and Yamaguchi (1991) provide more detailed reviews of discrete-time models.

4.4.1 Discrete-Time versus Continuous-Time Models

For many event histories, a discrete-time analysis approach is appropriate. In a continuous-time model, time is assumed to be measured as a continuous variable which can take any nonnegative value. However, in some situations, the duration of interest is measured in discrete time units, for example the number of menstrual cycles to conception. In such cases, a discrete-time model is clearly more suitable. Even if the underlying process operates in continuous time, durations are often recorded in discrete intervals of time, such as the nearest month or year; thus it is more natural to assume a model which reflects a discrete-time measurement than one which assumes that durations are measured exactly.

One feature of data which are grouped into reasonably broad discrete-time intervals is that they are likely to contain a large number of ties, which arise when two or more individuals experience an event at the same time. In the case of the widely used Cox proportional hazard model (Cox 1972), estimated via partial likelihood, the presence of ties can lead to serious biases since one of the assumptions of the continuous-time model is that the probability of the occurrence of more than one event at a time point is zero. Although modifications can be made which allow for ties (e.g. Kalbfleisch and Prentice 1980; Sinha et al. 1994), an attractive and easily implemented alternative is to use a discrete-time formulation.

Another advantage of the discrete-time approach is that it is straightforward to incorporate time-varying covariates in the model. As we will discuss in the next section, the data are structured as a series of Bernoulli responses, one for each discrete time point, and, therefore, covariates can take on different values from one time point to the next. Finally, discrete-time models can be fitted using any model for binary response data and as such can be estimated in a wide range of statistical packages.

4.4.2 The Discrete-Time Model

Suppose there are n individuals, indexed by $i = 1, \ldots, n$. For individual i, observation starts at time $t = 1$ and continues until time $t = t_i$ when the event of interest or censoring occurs. In the discrete-time model, time is assumed to take only integer values. Let c_i be the censoring indicator such that

$$c_i = \begin{cases} 1 & \text{if individual } i \text{ is uncensored} \\ 0 & \text{if individual } i \text{ is censored} \end{cases}$$

In addition, suppose there is a vector of possibly time-varying covariates $\mathbf{x}_{ti}$ associated with individual i.

The discrete-time hazard rate h_{ti} is defined as

$$h_{ti} = \Pr(T_i = t | T_i \geq t, \mathbf{x}_{ti}),$$

where T is a discrete random variable for the time at which an event occurs.

We next need to specify the form of the relationship between h_{ti}, time and the covariates $\mathbf{x}_{ti}$. The general form can be written

$$h_{ti} = g^{-1}(\alpha_t + \mathbf{x}'_{ti}\beta),$$

where $g(.)$ is the link function, α_t is some function of time and β is a vector of model parameters.

The most common choice of $g(.)$, and one which we will adopt in the later analyses, is the logit link

$$g(h_{ti}) = \log\left(\frac{h_{ti}}{1 - h_{ti}}\right) = \alpha_t + \mathbf{x}'_{ti}\beta.$$

Other possible specifications include the complementary log-log link

$$g(h_{ti}) = \log[-\log(1 - h_{ti})],$$

and the probit

$$g(h_{ti}) = \Phi^{-1}(h_{ti}),$$

where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution.

The function of time α_t can be specified in a number of ways. The simplest model would assume no time dependency, or a constant hazard, $\alpha_t = \alpha$. Other choices include a linear function of time, $\alpha_t = \alpha_0 + \alpha_1 t$; a quadratic function, $\alpha_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$; or log-linear relationship, $\alpha_t = \alpha_0 + \alpha_1 \log(t)$. Alternatively, one could opt for a non-parametric approach which at its most complex entails fitting a separate parameter for each time point. However, it is often possible to group together time points into wider intervals of time. This leads to a piecewise-constant hazards model where the hazard is assumed to be constant within each time interval. For this model, time is treated as a categorical covariate.

For any link function $g(.)$, the likelihood can be written

$$L = \prod_{i=1}^{n} [\Pr(T_i = t)]^{c_i} [\Pr(T_i > t)]^{1-c_i}, \tag{4.1}$$

where

$$\Pr(T_i = t) = h_{ti} \prod_{j=1}^{t-1} (1 - h_{ji}) = \frac{h_{ti}}{1 - h_{ti}} \prod_{j=1}^{t} (1 - h_{ji}) \tag{4.2}$$

and

$$\Pr(T_i > t) = \prod_{j=1}^{t} (1 - h_{ji}). \tag{4.3}$$

Substitution of (4.2) and (4.3) into (4.1) and taking logarithms gives the following log-likelihood

$$l = \sum_{i=1}^{n} c_i \log \left(\frac{h_{ti}}{1 - h_{ti}} \right) + \sum_{i=1}^{n} \sum_{j=1}^{t} \log(1 - h_{ji}).$$

Now suppose we define a new dummy variable

$$y_{ti} = \begin{cases} 1 & \text{if individual } i \text{ experiences event at time } t \\ 0 & \text{otherwise.} \end{cases}$$

Substituting y_{ti} for c_i, the log-likelihood can be rewritten as

$$l = \sum_{i=1}^{n} \sum_{j=1}^{t} \left[y_{ji} \log \left(\frac{h_{ji}}{1 - h_{ji}} \right) + \log(1 - h_{ji}) \right]$$

which is recognisable as the log-likelihood for binary response data, where the binary response is y_{ti} and $h_{ti} = \Pr(y_{ti} = 1)$. Therefore, it follows that this model can be estimated using any software which fits generalised linear models to binary data. For example, if we had chosen the logit link, the model can be fitted in any package that performs logistic regression.

For any individual i with covariate vector $\mathbf{x}_{ti}$, an estimate of the hazard h_{ti} can be obtained directly from the fitted model. For example, if the logit link is used, h_{ti} is estimated as

$$\hat{h}_{ti} = \frac{\exp(\hat{\alpha}_t + \mathbf{x}'_{ti}\hat{\beta})}{1 + \exp(\hat{\alpha}_t + \mathbf{x}'_{ti}\hat{\beta})}.$$

As in the life table method, $\hat{h}_{ti}$ can then be used to derive estimates of S_{ti}, the probability that no event occurs before time t, and F_{ti}, the cumulative probability of an event occurring before t. The estimates of S_{ti} and F_{ti} are given by

$$\hat{S}_{ti} = \prod_{j=1}^{t-1} (1 - \hat{h}_{ji}),$$

and

$$\hat{F}_{ti} = 1 - \hat{S}_{ti}.$$

In order to fit the discrete-time model, the data must first be restructured to obtain a set of Bernoulli responses $\{y_{ti}\}$, $t = 1, \ldots, t_i$, for each individual i. The Bernoulli responses for a given individual will be a series of zeros for each time point at which the event has not yet occurred. The final response will be coded as a one if the event occurs and a zero if the individual is censored. For example, if an individual experiences the event of interest in the sixth month of observation, he or she will contribute six records: five zeros, followed by a one. The total number of records in the restructured data set will be equal to the total exposure time for all individuals in the sample. The covariates for each of these new responses will be assigned whatever value they take at that time point, enabling time-varying covariates to be accommodated easily.

Clearly this approach can generate a huge data set, particularly when the time units are small relative to the period of observation. However, in the case of a constant or piecewise-constant hazards model, it is possible to reduce the size of the data set by grouping the Bernoulli responses over time intervals for which the hazard is assumed to be constant to a single binomial response. For the binomial response, the number of 'successes' is the number of events that occur over the interval (0 or 1) and the number of trials is the number of discrete time points in the interval. Alternatively, if all covariates are categorical it is more efficient to analyse the data at a group level, rather than at the individual level, by forming a contingency table. Standard log-linear techniques can then be applied. If the number of categories of the time variable and other covariates is large, however, the contingency table approach quickly becomes impractical. Another way in which the data set may be reduced is to consider making the discrete time units larger. For example, rather than having a Bernoulli response for every month, it may be possible to have a response for every three-month period or even broader time periods. Diamond et al. (1986) found that little precision was lost by grouping durations into reasonably broad groups, and that the computational time was reduced considerably. However, parameter estimates and standard errors for the time aggregated models should be compared with those obtained for the model using the original time units to ensure that the loss of information does not affect the estimates too greatly.

4.5 Repeated Events

Many events studied by social scientists can occur more than once to an individual over their lifetime: repeated birth intervals, contraceptive use intervals and employment spells are just a few examples. Recurrent events present a methodological problem since the fundamental assumption of most statistical models is that all observations are independently distributed. However, one would usually expect repeated durations on an individual to be correlated. For example, in an analysis of birth interval durations, a woman who has a high level of fecundity is likely to become pregnant very quickly and may, therefore, contribute a series of short birth intervals. In the context of contraceptive discontinuation, one may expect a woman who experiences difficulties with use to discontinue early, contributing a sequence of short spells, while others who have fewer problems may manage to continue for longer periods and perhaps contribute only one long spell.

The simplest approach, and the one most commonly used, is to consider each occurrence of the event separately. This approach is statistically inefficient, however, because in many cases the process determining event occurrence will vary very little for successive events for the same individual. Alternatively, the observations can be pooled across all individuals and covariates capturing the characteristics of an individual's prior event history can be incorporated into the analysis to take into account possible dependencies between observations on the same individual. In a study of contraceptive discontinuation, for example, these might include the number of previous uses of contraception, the durations of previous use intervals or the number of previous failures or other method-related problems. However, there are still likely to be some unobserved factors for which we cannot control and, therefore, some correlation between durations for a given individual is likely to persist.

Another approach would be to include a covariate representing the order of the event. This can then be interacted with other covariates in the model, in particular time, to test whether the process varies across repeated occurrences of the same event. The order of event is only meaningful, however, if each individual's complete event history is known. Contraceptive histories, for example, are usually collected for a period of only several years before the survey. In such situations, it is not possible to determine the true order of a contraceptive use spell: the first interval in the observation window may be the first time a woman has used contraception, but in most cases there would have been some prior use on which we have no information.

4.6 Multiple Event Types

In many cases, there will be more than one way of leaving a particular state. For example, a patient may die from one of a number of 'competing' causes, or a woman may discontinue use of a contraceptive method for any one of a number of reasons such as failure or side-effects. Rather than treat all causes of death or reasons for contraceptive discontinuation as alike, it is often of interest to distinguish between different kinds of events. To accommodate multiple event types, biostatisticians have developed techniques for 'competing risks' and demographers have extended the traditional single-decrement life table to multiple decrements.

Multiple event types can be classified into two broad groups. The first class consists of situations where one causal process determines the occurrence of an event of any kind, and another process determines the type of event. In this situation, a two-step nested analysis strategy can be used. In the first stage, an event history model of the type described in Section 4.4.2 is fitted to model the occurrence of an event, making no distinction between the different types. Next, a multinomial logit model is fitted to determine the relative risk of each type of event among those individuals who have experienced an event. This approach is appropriate if it is reasonable to assume that the same causal process underlies the occurrence of each type of event, that is, the covariates and their effects are the same across event types. Tutz (1995) uses a model of this type to analyse unemployment durations. First, a discrete-time event history model is employed to model the time to the termination of unemployment. In the second stage, a binary logit model is fitted to determine the probability that the job is full-time versus part-time, conditional on the fact that an individual has left unemployment.

The second class of multiple event types consists of situations where the causal process determining the occurrence of an event varies according to the type of event. In such cases, the factors affecting event occurrence may be different for each event or the same set of factors may be operating in different ways. This class of multiple event types contains an important subclass for which the occurrence of one event type removes the individual from risk of any of the other event types. This situation is known as 'competing risks' in the survival analysis literature and is the most frequently studied. Cause of death is an example, where an individual who dies from one cause is no longer at risk of dying from any other cause. A demographic example is contraceptive discontinuation where a woman who stops use because of side-effects is removed from the risk of contraceptive failure or any other type of discontinuation.

4.6.1 Life Table Methodology for Multiple Event Types

Two life table approaches are commonly used in situations where there is more than one type of event of interest. The first is an extension of the single-decrement life table to allow for more than one type of exit—the multiple-decrement life table. This method yields a set of dependent or net rates where the probability of one type of event occurring depends on the probabilities for the other types of event. An alternative approach is to calculate a series of single-decrement life tables where all events other than the one of interest are treated as censored. This yields independent or gross rates which are the hypothetical rates that would occur if all other event types were removed.

The Multiple-Decrement Life Table

Suppose that an individual can occupy one of s states at any given point in time, where the first $s-1$ are types of event or decrement and the final state s represents right-censored cases for whom the event has not yet occurred. Let

$$d_{rt} = \text{no. events of type } r \text{ in } [t, t+1),\ r = 1, \dots, s-1.$$

A hazard rate can be defined for each event type r

$$h_{rt} = \Pr(\text{event of type } r \text{ in } [t, t+1) | T \geq t),\ r = 1, \dots, s-1,$$

which can be estimated as

$$\hat{h}_{rt} = \frac{d_{rt}}{n_t - 0.5c_t}.$$

The overall hazard rate for an event of any type is

$$h_t = \sum_{r=1}^{s-1} h_{rt},$$

and the probability that no event of any type occurs is

$$h_{st} = 1 - h_t.$$

We can also define the unconditional probability of event occurrence for each type of event as

$$f_{rt} = \Pr(\text{event of type } r \text{ in } [t, t+1)).$$

The overall probability of survival to time t is

$$\begin{aligned} S_t &= \Pr(\text{no event of any type before } t) \\ &= \prod_{j=1}^{t-1} (1 - h_t). \end{aligned}$$

The hazard of an event of type r can be expressed in terms of f_{rt} and S_t as

$$h_{rt} = \frac{f_{rt}}{S_t},$$

and, using this relationship, the cumulative probability of an event of type r before time t can be written

$$\begin{aligned} F_{rt} &= \sum_{j=1}^{t-1} f_{rj} \\ &= \sum_{j=1}^{t-1} h_{rj} S_j. \end{aligned}$$

As for the single-decrement case, the estimate of h_{rt} can be used to derive an estimate of S_t and, therefore, $\hat{F}_{rt}$.

The multiple-decrement life table yields the hazard of an event of type r occurring in the presence of all other competing risks. These rates reflect the actual experience of the sample since, in most situations, individuals are simultaneously at risk of experiencing a number of different types of event.

Associated Single-Decrement Life Tables

In some situations, comparisons of rates obtained from different multiple-decrement life tables may be misleading. This is because the risk of one type of event occurring is dependent on the frequency with which the other events occur. As an example, suppose we wish to compare failure rates of two methods of contraception, say the pill and withdrawal. In the multiple-decrement life tables for the two methods, the risk of failure will depend upon the number of discontinuations for reasons other than failure. Therefore, failure rates can be very different for the pill and withdrawal, not necessarily because one method is more effective than the other, but because one method has a higher rate of discontinuation due to other reasons than for failure. For instance, the pill may have a lower failure rate because the rate of discontinuation due to side-effects is high for the pill, but negligible for withdrawal. An alternative is to use independent rates which represent the underlying risk of a particular event occurring in the absence of all other competing risks. These rates are unaffected by the degree of risk associated with other event types. They are calculated by constructing a single-decrement table for each type of event, treating the occurrence of all events other than the one of interest as censored observations.

A major drawback of the associated single-decrement table approach, however, is the need to assume that the competing risks are independent. This strong assumption may be untenable in many applications, but cannot be tested (Kalbfleisch and Prentice 1980). Kost (1993) gives the following example to illustrate how the independence assumption may be unrealistic in the study of contraceptive discontinuation. In order to calculate the risk of contraceptive failure, one needs to assume that women who stop to become pregnant or for other reasons would have faced the same risk of failure as those who continue use. However, it is possible that women who ultimately stop to get pregnant would have had a higher chance of failure than continuing users because they are less concerned about the consequences of failure.

Gross rates are purely theoretical since in most real populations individuals are simultaneously at risk of experiencing any of the competing event types. Therefore, if the primary interest is to describe the observed risks associated with each event type in a population or in comparing patterns across subgroups of the population, multiple-decrement rates are more appropriate. However, gross rates may be useful to compare the risks of a particular event type across different populations or in the same population over time.

4.6.2 The Discrete-Time Competing Risks Model

If we wish to incorporate covariates into the analysis, the discrete-time model described in Section 4.4.2 can be extended to handle multiple event types.

Suppose there are $s-1$ decrements or causes of 'death', then we define the discrete-time cause-specific hazard by

$$h_{rti} = \Pr(T_i = t, R = r | T_i \geq t, \mathbf{x}_{rti}), \; r = 1, \ldots, s-1,$$

where h_{rti} is the probability that individual i experiences an event of type r at time t, given survival to time t and a set of covariates x which may be cause-specific.

The overall hazard of an event of any type is

$$h_{ti} = \sum_{r=1}^{s-1} h_{rti},$$

and the hazard that no event of any type, i.e. an event of type s, occurs is

$$h_{sti} = 1 - \sum_{r=1}^{s-1} h_{rti}.$$

As for the single event case, the relationship between h_{rti}, time and covariates can be expressed in a general form

$$h_{rti} = g^{-1}(\alpha_{rt} + \mathbf{x}'_{rti}\beta_r), \; r = 1, \ldots, s-1,$$

where the dependency of the hazard on time and the covariates, represented by α_{rt} and β_{rt} respectively, can vary according to the type of event.

For any link function $g(.)$, the likelihood can be shown to be

$$L = \prod_{i=1}^{n} \left[\frac{h_{rti}}{1-h_{ti}}\right]^{c_i} \prod_{j=1}^{t} (1 - h_{ji}),$$

where c_i is the censoring indicator as before.

The likelihood for a continuous-time competing risk model can be factorised into separate components for each kind of event (see e.g. Kalbfleish and Prentice 1980). This result is very convenient as it enables the continuous-time model to be estimated via a series of models for single events, one for each type of event where other types of events are treated as censored cases. Therefore no new methodology needs to be developed for modelling multiple kinds of events in continuous time. Unfortunately, the likelihood for the discrete-time model does not factorise in this way and maximum likelihood estimation must be carried out simultaneously for all types of event.

Although other link functions $g(.)$ can be applied, a generalisation of the logistic function is most commonly used as it is the most tractable. This yields the following multinomial logit model

$$\log\left(\frac{h_{rti}}{h_{sti}}\right) = \alpha_{rt} + \mathbf{x}'_{rti}\beta_r, \; r = 1, \ldots, s-1. \tag{4.4}$$

The cause-specific hazard for event type r is

$$h_{rti} = \frac{\exp(\alpha_{rt} + \mathbf{x}'_{rti}\beta_r)}{1 + \sum_{k=1}^{s-1} \exp(\alpha_{kt} + \mathbf{x}'_{kti}\beta_k)}, \; r = 1, \ldots, s-1, \tag{4.5}$$

and the overall hazard is

$$h_{ti} = \frac{\sum_{k=1}^{s-1} \exp(\alpha_{kt} + \mathbf{x}'_{kti}\beta_k)}{1 + \sum_{k=1}^{s-1} \exp(\alpha_{kt} + \mathbf{x}'_{kti}\beta_k)}. \tag{4.6}$$

Substituting (4.5) and (4.6) into the discrete-time competing-risks likelihood will yield the likelihood for multinomial response data. Therefore, the model can be fitted using any software which can perform multinomial logit regression. A series of $s-1$ contrasts are estimated simultaneously, each one comparing the hazard of one type of event with the baseline category s which represents 'no event'. Similar to the binary logit model for single events, the data must be restructured to obtain a multinomial response with s categories for each time point t. For any individual, their response in the discrete time points before an event occurs will be coded s. Their final record will be coded r if an event of type r occurs (r=1,...,s-1) or s if they are censored.

As for the case where there is only one type of event of interest (Section 4.4.2), the cause-specific hazards can be estimated directly from the fitted model. For an individual i with covariate vector $\mathbf{x}_{rti}$, the estimated hazard of an event of type r is given by

$$\hat{h}_{rti} = \frac{\exp(\hat{\alpha}_{rt} + \mathbf{x}'_{rti}\hat{\beta}_r)}{1 + \sum_{k=1}^{s-1} \exp(\hat{\alpha}_{kt} + \mathbf{x}'_{kti}\hat{\beta}_k)}, \quad r = 1, \ldots, s-1,$$

and the estimated hazard that no event occurs is

$$\hat{h}_{sti} = 1 - \sum_{r=1}^{s-1} \hat{h}_{rti}.$$

The estimates $\hat{h}_{rti}$ can then be used to derive estimates of S_{ti}, the probability of survival, and F_{rti}, the cumulative probability of an event of type r occurring, using the formulae presented in Section 4.6.1, i.e.,

$$\hat{S}_{ti} = \prod_{j=1}^{t-1} \hat{h}_{sji},$$

and

$$\hat{F}_{rti} = \sum_{j=1}^{t-1} \hat{h}_{rji}\hat{S}_{ji}.$$

The multinomial model in (4.4) is a multivariate analogue of the multiple-decrement life table, and thus yields dependent or net rates. In order to obtain independent or gross rates, one can fit a separate binary logit model (Section 4.4.2) for each event type, treating all other events as censored.

4.7 Multilevel Discrete-Time Event History Models

4.7.1 Unobserved Heterogeneity

A major problem with the discrete-time event history models described in Sections 4.4.2 and 4.6.2, and in analogous continuous-time models, is the assumption that the covariates represented by $\mathbf{x}$ explain all the variation in the hazard rate. In most situations, this is clearly untenable: some covariates may have been omitted because they were unavailable, unmeasurable or their importance was unsuspected. As a result, there is likely to be some heterogeneity in the population which has not been observed. The presence of 'unobserved heterogeneity' or 'frailty', as it is often termed in the medical literature, has important consequences. In general, if individuals differ by some unobserved factors, the observed dynamics at the population level will tend to be different from those observed at the individual level (Vaupel et al. 1979, Blossfeld and Hamerle 1989). Even if the hazards of individuals in a population are constant over time, the aggregate population hazard tends to vary across time, typically showing a decreasing hazard with time. This can be explained by selection effects: if a population is heterogeneous in its susceptibility to experiencing an event, then those individuals with high hazard rates tend to experience the event first, leaving the individuals with the lowest hazards in the at risk population. Therefore at each time point, the population is depleted of those individuals most likely to experience the event and thus the overall population hazard rate declines with time.

4.7.2 A Multilevel Approach to Incorporate Unobserved Heterogeneity

The usual approach to incorporate unobserved heterogeneity in the model is to add an error term or random effect whose variance represents the unexplained heterogeneity between individuals. Allison (1987) discusses unobserved heterogeneity models with logit or probit links for clustered binary response data. Such models are suitable for the discrete-time event history model described in Section 4.4.2 since the data have a two-level clustered structure: binary responses for each discrete time unit (level 1) nested within individuals (level 2). The random effect can be introduced either inside or outside the linear predictor yielding an 'internal' or 'external' model. For the logit link, for example, the internal model for the hazard function is

$$h_{ti} = \text{logit}^{-1}(\alpha_t + \mathbf{x}'_{ti}\beta + u_i),$$

and the external model is

$$h_{ti} = \text{logit}^{-1}(\alpha_t + \mathbf{x}'_{ti}\beta) + u_i,$$

where u_i is the random effect for individual i.

An example of an external model is the beta-binomial model (see e.g. Egger 1992) where the probability of event occurrence is assumed to follow a beta distribution. However, internal models which assume that the unobserved covariates are operating on the same scale as the observed covariates are the most commonly used. The random effect is usually assumed to follow a normal distribution with zero mean and variance σ_u^2. An advantage of the internal logistic model is that it can be estimated using any software which performs multilevel logistic regression, e.g. MLn, EGRET, VARCL, SABRE and BUGS. The estimation procedures these packages use are outlined in Section 2.7.

The majority of applications of the internal model appear in the econometrics literature, and in particular in studies of employment and unemployment durations. For example, Davies et al. (1992) incorporate a normally distributed individual-specific error term in a logit model to allow for residual heterogeneity in the risk of becoming unemployed. Elias (1994) uses a similar approach to control for person-specific heterogeneity in the duration of employment spells. Narendranathan and Elias (1993) also analyse employment durations, but use two different formulations for the random effect distribution. The first uses the same approach as Davies et al. (1992) and Elias (1994), assuming a normal distribution for the heterogeneity. The second uses the non-parametric approach developed by Heckman and Singer (1984). Narendranathan and Stewart (1993) also considered parametric and non-parametric characterisations of the heterogeneity distribution in an analysis of the length of unemployment spells, using probit and logit event history models. Several studies also distinguish between 'movers', individuals who experience the event of interest regularly, and 'stayers', individuals who have a zero or low probability of ever having an event. For example, Elias (1994) uses a 'mover-stayer' model which incorporates end-points—one to represent individuals who persistently change jobs and another to represent those who never change their employer. The discrete-time mover-stayer model, with unobserved heterogeneity, can be estimated using SABRE. As well as providing estimates of the fixed parameter estimates and the random effect variance, SABRE yields estimates of the probabilities of being a 'mover' or a 'stayer'.

A multilevel formulation of the model enables a number of possible extensions to the basic model, each of which can be estimated straightforwardly in available software. For example, the unobserved heterogeneity may be a function of time or any of the other covariates in the model, that is the effect of the unobserved factors may vary across time or by individual characteristics. Such a model is an example of a random coefficients model. Another extension would be to incorporate more levels into the model, for example to explore areal differences in hazard rates.

4.7.3 Multilevel Models for Repeated Events

In situations where it is possible for an individual to experience an event more than once over the observation period, the two-level unobserved heterogeneity model can be extended to three levels: discrete time units (level 1) nested within events (level 2) within individuals (level 3). This would allow one to test whether an individual's random effect value is constant across successive events in their event history, that is, the unobserved individual characteristics are the same for all observed durations for that individual. Alternatively the extravariation in an individual's risk changes from event to event, in which case there is unobserved heterogeneity between the observations within each subject. A different approach to model recurrent events could be to fit a two-level model, with individuals at level 2, and include a covariate denoting the order of event. A random coefficient model, in which the coefficient of the order of the event is allowed to vary randomly across individuals, could then be used to test whether the amount of unobserved heterogeneity varies across successive events.

4.7.4 Multilevel Models for Multiple Event Types

If there are multiple types of event, the single-level competing-risks model presented in Section 4.6.2 can be extended to incorporate cause-specific random effects to control for unobserved heterogeneity. This yields the following multilevel multinomial logit model for the cause-specific hazard function

$$\log\left(\frac{h_{rti}}{h_{sti}}\right) = \alpha_{rt} + \mathbf{x}'_{rti}\beta_r + u_{ri}, \; r = 1, \ldots, s-1,$$

where u_{ri} is the random effect associated with event type r for individual i, $u_{ri} \sim N(0, \sigma_r^2)$ and $\text{cov}(u_{ri}, u_{r'i}) = \sigma_{rr'}, r \neq r'$.

This model can be estimated as a multilevel multinomial model in MLn (see Section 2.8). The above model is a two-level random intercepts model, but can be extended to further levels or to allow for random coefficients, where the effect of time or other covariates varies randomly across individuals.

As for the single-level discrete-time competing risks model, the data must first be restructured to obtain a series of multinomial responses for each discrete time unit during which an individual was exposed to risk. In order to fit the multinomial model, the data need to be expanded further: if the response has s categories, then $s-1$ binary indicator variables for the $s-1$ event types need to be generated for each multinomial response. This procedure is described in more detail in Section 2.8.2. The restructured data set can potentially be very large, especially if there is a large number of competing alternatives. In addition, a separate set of parameters needs to be estimated for each contrast of the hazard of an event type with the hazard of no event occurring, which often results in a large number of parameters. This coupled with the size of the data set means that fitting the multilevel multinomial model tends to be highly computer intensive. As a result, convergence can be very slow. It is possible to reduce the number of iterations required for convergence, however, by specifying suitable starting values. For the fixed part parameters, these may be estimates from the corresponding single-level model. Alternatively one could fit a series of pairwise binary logit models as an approximation to the multinomial model to obtain initial values for both the fixed and random part parameters. This process involves fitting separate models that compare each response category with the baseline, and ignoring the other categories (Begg and Gray 1984). Possible strategies for fitting multilevel models to discrete response data are discussed in more detail in Section 2.9.

Other approaches to incorporate unobserved heterogeneity into a discrete-time competing risks model include a study by Enberg et al. (1990) on transitions between work and welfare. They also employ a multinomial logit model with cause-specific random effects, but the random effects are assumed to follow independent discrete distributions. Thus they assume that the unobservables governing the transition rates out of the different states are uncorrelated. In many situations this assumption may be unrealistic, since it is likely that some of the unobserved risk factors will be common to more than one of the competing alternatives. For example, in an analysis of contraceptive discontinuation, a woman's biological characteristics may potentially affect both her risk of a contraceptive failure and other method-related discontinuations. The restriction that the random effects for different types of event are independent is equivalent to what is called the Independence of Irrelevant Alternatives (IIA) assumption (Ben-Akiva and Lerman 1985). This is a problem if, for some individuals, one of the event types cannot occur, perhaps because of unobserved factors, as this will in general affect the odds of experiencing the other types of event. However, with the IIA assumption, the odds are assumed to be unaffected by the propensity of alternative types of event. One way to relax the IIA assumption is to allow the random effects to be correlated across competing alternatives as in the multilevel model described above. Hill et al. (1993) adopt another approach and allow for the possibility of shared unobserved risk factors explicitly by separating the random effect for each competing alternative into two components: one which is the same for each alternative and another which varies across alternatives.

4.7.5 Misspecification of the Unobserved Heterogeneity Distribution

One important issue, raised by Heckman and Singer (1984), is the sensitivity of parameter estimates to the choice of the unobserved heterogeneity distribution. In their study, a series of continuous-time Weibull hazards models were fitted, differing only in the distributional assumptions for the random effect. The results showed an alarming difference in the estimates of the covariate coefficients depending on whether the unobserved heterogeneity was assumed to follow a normal, lognormal or gamma distribution. Heckman and Singer (1984) proposed an alternative non-parametric approach and simulation results showed that the estimates for this model were fairly robust to a range of 'true' heterogeneity distributions. However, Elias (1994) suggests that for discrete-time binary response event history models, parametric approaches yield robust estimates of the regression coefficients. Narendranathan and Elias (1993) and Narendranathan and Stewart (1993) use both a parametric approach, assuming normally distributed heterogeneity, and the non-parametric approach of Heckman and Singer (1984), and find that the two sets of estimates are very similar. Their findings are consistent with the results of a simulation study by Neuhaus et al. (1992) which examines the impact of misspecification of the random effects distribution when fitting random effects logistic models. Binary data with variable cluster size and substantial intraclass correlation were generated using five different random effect distributions: two gamma and t distributions (one symmetric and the other highly skewed) and a normal distribution. Logistic-binomial models were then fitted, using a binomial approximation to the normal distribution (see Section 2.7.4). The simulation results showed that the estimates of both the regression parameters and standard errors obtained from the logistic-binomial model were robust to misspecification of the random effects distribution.

4.8 Summary

In this chapter, a number of approaches to the analysis of event history data are described. Life table techniques for both single types of event and multiple event types are presented. Discrete-time event history models, multivariate analogues to both the single-decrement and multiple-decrement life table, are then described. These models can be estimated using standard regression techniques for binary and polychotomous response data. The problem with standard discrete-time models, however, is that they assume that all variation in individual hazard rates can be explained by the covariates included in the model. Therefore, these models have been extended to incorporate random effects which account for unobserved heterogeneity using a multilevel framework. This enables discrete-time event history models with unobserved heterogeneity to be estimated using the multilevel procedures for discrete response data described in Chapter 2. Models for single or multiple event types can thus be fitted using available software. Another advantage of the multilevel approach is that recurrent events can be analysed together. A three-level model can be used where discrete time units (level 1) are nested within repeated events (level 2) within individuals (level 3). Further hierarchical levels can be added to explore extravariation between higher units of aggregation, for example geographical areas.

Chapter 5

Child Immunisation Uptake in Rural Bangladesh

5.1 Introduction

In this chapter, a multilevel discrete-time event history model is used to analyse the factors affecting the level of child immunisation uptake in several rural communities in Bangladesh. Using a unique, longitudinal data set collected by Save the Children USA, the relative influence of maternal, household and community factors in the determination of immunisation status of children are examined. The data refer to a five year period, 1988 to 1993, when several new programmes of service delivery were introduced with considerable reported success. These programmes were in the areas of immunisation against six major childhood diseases, family planning services delivered to the doorsteps of rural women, and the provision of micro credit to rural poor women through an innovative institutional mechanism.

A total of five levels of nesting can be identified: child, mother, household, village and intervention area. A number of studies have found evidence of death clustering for children with the same mother, owing to shared maternal characteristics (see for example Curtis et al. 1993; Curtis and Steele 1996; Madise and Diamond 1995) and one might expect familial clustering to extend to other health outcomes such as immunisation uptake. The household is also an important unit of analysis as in Bangladesh it is at the household level which most decision-making takes place. Household factors have been found to influence patterns of fertility, mortality, migration and marriage (Foster 1993). The strength of household-level effects after controlling for community-level variation are also examined. The data were collected in four intervention areas with considerably different levels of infrastructure which leads to a large amount of areal variation in immunisation uptake, and there is also variation among villages within the four regions.

5.2 The Expanded Programme on Immunisation and its Implementation in Bangladesh

The Expanded Programme on Immunisation (EPI) was launched by the World Health Organisation in 1974. The programme focused on tackling six major childhood diseases: measles, tuberculosis, pertussis (whooping cough), diphtheria, tetanus and poliomyelitis with the aim of universal child immunisation by 1990. Efforts were intensified after 1982 and by 1990 the global coverage rate was estimated at 80%. Its main goals for the 1990s include the elimination of poliomyelitis and neonatal tetanus by the year 2000. Under the EPI, children receive one dose of BCG for protection against tuberculosis, three doses of the triple vaccine DPT (diphtheria, pertussis and tetanus), three doses of either IPV (injectable) or the more commonly used OPV (oral) for poliomyelitis protection and one dose of the measles vaccination by their first birthday.

EPI guidelines recommend that the BCG vaccination be given as soon as possible after birth. It is usually administered at approximately six weeks along with the first doses of DPT and OPV, although these first shots are frequently given at birth which is often the only contact the mother has with the health services. The second and third doses of DPT and OPV are usually given at four week intervals. Sometimes a fourth dose of DPT is given at 18 months followed by a booster later on in childhood. The recommended optimal age for immunisation against measles is nine to twelve months. The timing of the measles vaccination is particularly important as it is ineffective if administered too early due to the presence of persisting maternal antibodies which interfere with the immune response. In addition it is recommended that all women of reproductive age, with particular attention to pregnant women, be given a tetanus toxoid (TT) immunisation which offers protection for newborns against neonatal tetanus.

The EPI had a late start in Bangladesh compared to other countries. In 1985 coverage was estimated as low as 2%. It was not until 1986 that the first phase of the programme was ready to be implemented, initially in eight of the total of 464 *thanas* (sub-district administrative units). The programme was extended in subsequent years to other *thanas* until in 1989 the EPI had been introduced to the whole country. However, immunisation coverage rates were still low with only 38% of children immunised by their first birthday (Huq 1991). Therefore the campaign was stepped up in 1989 and the Ministry of Health and Family Planning joined forces with other relevant government bodies and non-governmental organisations (NGO) to improve service delivery and communication. As a result, coverage increased rapidly so that by 1990 it was estimated at 69% for the full three doses of DPT and OPV, 65% for measles and 86% for BCG (Huq 1991).

There are a number of practical considerations in the implementation of a large-scale health initiative such as the EPI. On the supply side, efficient distribution and administrative systems are essential as are adequate vaccine storage facilities and sufficient numbers of well-trained health workers. Important considerations on the demand side include communication with mothers and households to encourage them to bring their children to be immunised.

In the initial stages of the EPI, vaccination sessions were organised by male health assistants and female family welfare assistants from the Health and Family Welfare services. In rural areas each ward was divided into a number of outreach sites where satellite immunisation camps were held on a monthly basis. In 1989 other ministries including social welfare, women's affairs, information and communication became involved. At the local level, the help of community volunteers, school teachers and local organisations was also enlisted. In particular, there has been collaboration between government workers and NGOs. For example, in Save the Children USA intervention areas, regular meetings are held between government workers and Save the Children workers to discuss how best they can combine their limited resources to provide an efficient service.

Although the EPI camps are government-run, the Save the Children workers play a key role in motivating mothers to attend the clinics on their monthly house-to-house visits.

Another major practical consideration is vaccine storage. Vaccines must be stored and transported to the immunisation camps at a temperature not exceeding eight degrees centigrade as high temperatures can affect the potency of a vaccine. Low temperatures are also harmful for DPT and TT vaccines. If a refrigerator breaks down or if there is an electricity failure, there may be a break in the cold chain which reduces a vaccine's efficacy. Vaccine storage can pose problems in many developing countries, particularly in remote rural areas where, if there is any electricity, the supply is likely to be unreliable. Recent technological developments which have helped to minimise these problems include special refrigerators which can operate on less than eight hours of electricity a day, freeze-dried vaccines which remain stable for up to a week at temperatures as high as 37 degrees centigrade and more heat-stable vaccines. Ice boxes and ice packs are used in the transportation of vaccines to ensure they remain cold. In Bangladesh, vaccine storage facilities have been installed in each *thana* since 1989.

An efficient administrative system is vital. All new births need to be registered in order to identify which children are eligible for immunisation. Children under two years and women aged 15-45 years are registered by government health workers on their household visits. However, their lists are updated only once a year. In the Save the Children intervention areas, lists of target women and children are drawn up from their household registration system. Once a child has received its first set of vaccinations, records need to be kept to ensure that they complete their immunisation schedule. In most countries, vaccination cards are given to women and children after their first vaccinations with the date of their next appointment and these have proved successful in reducing the problem of drop-out.

Even if there is a demand for health care, the costs to the mother, in terms of time and travelling expenses, are likely to be prohibitive unless health services are easily accessible. Immunisation services are brought close to the community via the use of mobile clinics. These are held every month at a designated household, the location of which is communicated to all mothers of children on the target lists before the day of the clinic. The issue of accessibility is especially important in a society such as Bangladesh where the system of purdah, which restricts the mobility of women, might prevent them from attending a clinic outside their neighbourhood. However in many rural areas, accessibility is problematic as roads are often inadequate. One of the Save the Children intervention areas, Nasirnagar, is very remote and inaccessible. During the monsoon, much of the area is inundated and boats provide the only means of transport between villages. The situation is even worse in the dry season as the poor condition of the roads makes it impossible to travel by rickshaw and often walking is the only option. This makes it very difficult for government and Save the Children workers to reach some of the more remote villages.

An effective system of communication to inform mothers of the nature and benefits of immunisation and to motivate mothers to bring their children to immunisation clinics is crucial. Mothers need to be given reliable information about immunisation if fears about modern medicine are to be overcome. The ideas of preventative health care are unfamiliar to many who only seek medical help when they are ill. Although immunisation does not depend on the mother having to learn, remember and repeat a new procedure, as is the case in oral rehydration therapy or contraceptive use, the mother has to be sufficiently well informed about immunisation if she is to be encouraged to attend a clinic and to return for additional doses.

In 1989 when efforts to expand immunisation coverage were intensified, a campaign of 'social mobilisation' was launched to raise motivational levels among parents. A variety of media and publicity approaches such as a series of national immunisation days were employed in attempts to increase awareness about the importance of immunisation. However, there has been much debate about whether such methods are effective in the long term. The grassroots health and family planning workers who make regular house-to-house visits play a vital role in raising public awareness and it is generally felt that personal contact is a more sustainable means of motivating parents to accept immunisation. These workers are also responsible for informing mothers when and where immunisation is being provided before the day of the clinic and reminding them when to return for follow-up doses.

The immunisation programme is monitored through monthly vaccination performance reports and field visits are made each month by medical officers and supervisory staff. Feedback is given on each district's performance and reward systems are in operation to generate local-level competition to improve coverage rates.

5.3 The Save the Children USA Surveillance System

The data for the study come from a prospective surveillance system operated by Save the Children USA in four areas in rural Bangladesh. Save the Children first came to Bangladesh in 1972 with relief operations after severe famine and floods. Their first development programme began in Rangunia in Chittagong district which was followed in 1975 by the initiation of programmes in three other areas: Nasirnagar in Brahmanbaria district, Ghior in Manikgonj and Mirzapur in Tangail. In 1988 their coverage was further extended to include a number of new villages in Nasirnagar. Starting with a broad-based community development approach in the 1970s which included programmes to improve the infrastructure of villages, Save the Children has moved towards more focused activities. These have included various health initiatives, education for children and income generation schemes, in particular women's savings groups to improve the economic participation of women.

Since 1986, Save the Children has used a computerised programme management information system to maintain demographic, socioeconomic, health, nutrition, sanitation, family planning and project participation information from each of the four intervention areas, with an approximate total population of 75,000. The system was initiated in July 1986 through universal registration of all households in the intervention areas. At the time of registration, a 'family registration form' was completed for each household. Data on all men and women in the household were collected, including details such as age, marital status, relationship to the head of the household, education and occupation. For women, information on each pregnancy experienced in the observation period was collected, such as pregnancy outcome, ante-natal care, place of delivery and type of birth attendant. Some details of methods of contraception used were also recorded. In addition, information on the socioeconomic situation of the household was collected.

This included the source of drinking water, sanitation, type of building material used to construct the house, income, food sufficiency, whether the household had savings or debts and ownership of various household items such as a television, radio, bedding and lighting (electricity or hurricane lamp). Save the Children have used these household-level data to construct a four-category classification of socioeconomic status to identify the poorest households who they then target in their development efforts.

After registration, the demographic core of the database was updated through monthly vital event reporting of births, deaths and migrations. Data on service delivery, including immunisation and contraceptive use, were collected on a quarterly basis using computer-generated rosters which were circulated to field staff and, after completion, returned for entry into the computerised system in the head office in Dhaka. Save the Children used these data primarily to obtain demographic profiles of the population in each intervention area and to generate various service statistics which were used for short-term operation planning, to monitor the progress of their interventions and for long-term strategic planning.

To assess the quality of the data, random samples of households were selected and their responses to a series of questions were compared to those recorded in the manual rosters. Ten samples of 20 randomly selected households, each containing at least one child under the age of five years, were assessed in two of the intervention areas, Rangunia and Nasirnagar. First a check was made that the selected household was actually currently registered in Save the Children's family enrolment system. The household was then interviewed on issues such as immunisation status and presence of a vaccination card. The information obtained in the interview and details from the vaccination card were then cross-checked with the corresponding information recorded in the rosters. A sample was rejected if less than 85% of the information from the interviews and cards matched that in the rosters. In general, the recorded information in the registration system was found to be complete and matched with that collected in the samples. All sampled households were found to be registered and only one lot was rejected for reasons of data quality.

To evaluate the accuracy of the reporting of immunisation dates, additional checks were made to ensure that the date of an immunisation post-dated the child's date of birth and that vaccinations were administered at roughly the recommended optimal ages and in the correct sequence. For example, if the first two doses of DPT had been received, a check was made that the second dose followed the first. Although several inconsistencies arose from these checks, the immunisation data were found on the whole to be of a high quality.

For this study, all children born between 1 January 1988 and 30 June 1993 were considered. This start date was used because Save the Children did not begin their development work in the new area of Nasirnagar until the end of 1987. Births after June 1993 were excluded because data collection ended in Rangunia, Mirzapur and Ghior around this time when Save the Children moved out of these areas to concentrate their efforts in Nasirnagar. In the preliminary descriptive analysis (Section 5.5), children who were born less than 12 months before the end of the observation period were excluded in order to calculate coverage rates for children who had sufficient time to complete their immunisation schedule. This resulted in a sample of 10,084 children. In the multivariate analysis using multilevel event history techniques (Section 5.6), children aged at least six months were considered. Thirty children were reported as being fully immunised before the age of six months. These children were excluded from the analysis as were children who died before the age of six months or who were born less than six months before the end of the study period to give a sample size of 11,033.

5.4 Methodology

A multilevel discrete-time event history model (Section 4.7.4) is used to analyse the factors that influence immunisation acceptance. A multilevel approach is employed in order to determine the extent to which characteristics of the individual child, mother, household, village and *thana* affect immunisation uptake. Clearly the social context in which a child lives is likely to have a major impact on their likelihood of being immunised. Factors operating at any of these levels may affect a child's chance of receiving immunisation, some of which are likely to have been unobserved.

A number of studies have found evidence of clustering of health outcomes for children in the same family, associated with maternal characteristics (Das Gupta 1990; Curtis et al. 1993; Guo 1993; Curtis and Steele 1996; Madise and Diamond 1995; Pebley et al. 1996). There may also be clustering at the household level. Characteristics of the household which may potentially influence immunisation uptake include socioeconomic status and the knowledge of and attitudes towards health care of the decision-makers within the household. In the present study, the household is defined as those individuals who eat together on a regular basis.

Factors operating at the community level may also affect immunisation acceptance. These might include the attitude of a community leader or of other families in the neighbourhood towards immunisation. Finally, there are likely to be factors at the *thana* level influencing health-related behaviour. The four *thanas* in which Save the Children worked are highly varied in terms of infrastructure and level of development. Mirzapur and Ghior are within easy motorable distance from the capital Dhaka. Rangunia is also close to a large town, while Nasirnagar is very remote and not easily accessible.

An event history model is used for two main reasons. First, although under the EPI guidelines children are supposed to be fully immunised by the age of 12 months, many children in the sample, particularly in the Nasirnagar area, do not receive their final immunisation until as late as two years or older. Therefore, rather than choosing a cut-off age of 12 months, say, at which children are classified as fully immunised or not fully immunised, an event history model is used to allow for these later immunisations. Second, the use of an event history model allows censored cases to be included in the analysis. In this case, censoring occurs if the child has died during the study period before being immunised or the study period has ended before the child has been immunised. In particular, an event history model approach allows the inclusion of children born 6-11 months before the end of the study period, most of whom are censored observations because they have not had sufficient time to complete their immunisation schedule. If a cut-off point of 12 months were used, all children aged less than 12 months would have to be excluded from the analysis due to inadequate exposure to the chance of being fully immunised. For the technical reasons mentioned in Section 4.4.1, a discrete-time formulation of the model is employed rather than a continuous-time version. The recommended age for the measles vaccination, the last in the immunisation schedule, is nine to 12 months. The majority of children receive their measles shot between these ages which results in a large number of ties. A continuous-time model would assume a zero probability of two children being fully immunised at the same time which is clearly unrealistic in this case.

Initially, models with random effects corresponding to the mother, household and village levels were fitted. Since there are too few intervention areas to allow the estimation of the distribution of a random effect at this level, the intervention area is represented by a fixed effect. This involves including a categorical covariate for intervention area in the fixed part of the model. Models which incorporate an additional child level were also considered, but there was no evidence to suggest the presence of unobserved heterogeneity at this level. This is not surprising as it seems likely that immunisation uptake is influenced more by the characteristics of the person making the decisions about the child's health care than those of the child itself. There was also no significant extra-variation at the mother level which implies that there is no difference in the chance of child immunisation for mothers living in the same household. However, since only 10% of households contain more than one mother who contributes children to the analysis sample, it is possible that there are too few multiple-family households to enable the mother and household effects to be disentangled. Therefore, the model was simplified to include random effects for the household and village levels only.

It was found that once a child begins the immunisation schedule the chance of dropping out is very low. For this reason, it was decided to represent immunisation status by a simple dichotomy: fully immunised versus not fully immunised (See Section 5.5 for a fuller discussion of drop-out rates). In a population where drop-out is high, an ordered multinomial response could be used to represent the number and type of immunisations received.

Let y_{tijk} be the binary response at age interval t for child i in household j in village k, where $y_{tijk} = 1$ if the child is immunised and $y_{tijk} = 0$ otherwise. Let $h_{tijk} = \Pr(y_{tijk} = 1)$, the hazard of being immunised in age group t. Then the general three-level discrete-time event history model can be written as

$$y_{tijk} = h_{tijk} + e_{tijk},$$

where

$$\log\left(\frac{h_{tijk}}{1-h_{tijk}}\right) = \alpha_t + \mathbf{x}'_{ijk}\beta + \mathbf{z}'_{uijk}\mathbf{u}_{jk} + \mathbf{z}'_{vijk}\mathbf{v}_k.$$

The vector $\mathbf{x}_{ijk}$ contains covariates which may be defined at the child, mother or household level, and β contains the associated parameters. Although none of the background characteristics considered in this analysis are time-dependent, it would be straightforward to incorporate time-varying covariates in the model. α_t is some function of time which in the present case is measured by the child's age. In this general multilevel model, the coefficients of the covariates in $\mathbf{z}_u$ and $\mathbf{z}_v$ are permitted to vary randomly across households and villages respectively. Usually $\mathbf{z}_u$ and $\mathbf{z}_v$ will be subsets of $\mathbf{x}$. e_{tijk} is the error term defined at the lowest level in the model, the age interval level, which is assumed to follow a Bernoulli distribution. $\mathbf{u}_{jk}(\sim N(0, \Omega_u))$ and $\mathbf{v}_k(\sim N(0, \Omega_v))$ are vectors of random effects at the household and village levels respectively. The above model is an example of a multilevel event history model where there is only one type of event of interest: the completion of the immunisation schedule (see Section 4.4.2).

As discussed in Section 4.4.2, the data must be restructured to obtain a binary response indicating immunisation status at each age interval t and for each child in the sample before a discrete-time event history model can be fitted. In this analysis, the children's ages are defined in intervals of three months, the first interval being 6-8 months. Observation of a child ends at the age when the last immunisation was received for those who are fully immunised, otherwise if the child has not been fully immunised (censored cases) the last recorded age corresponds to their age at the end of the study period or their age at death. This data reconstruction results in a very large data set: the 11 033 children in the analysis sample contribute a total of 49 156 records. For computational reasons, it was decided to take a random 50% sample of households in each village to obtain a final sample of 5,435 children, contributing 23,700 records. After the data have been restructured in this way, intervals for ages of more than 18 months were grouped together. Therefore, five distinct age intervals are considered, assuming a constant hazard of becoming fully immunised in each: 6-8, 9-11, 12-14, 15-17 and 18+ months. The age effect, represented by α_t, is incorporated into the analysis by including a categorical variable, with a separate category for each of the five age intervals, as a covariate in the model. A categorical formulation of age was found to give the best fit. Linear and quadratic age effects failed to reflect the pattern of the observed hazard which is low at 6-8 months, peaks at 9-11 months and then starts to decline.

5.5 Results from the Preliminary Analysis

Under EPI, each child is supposed to have received three doses of DPT, three of OPV, one BCG and one measles by their first birthday. Table 5.1 shows the percentage of children aged 12 months or more who have actually received the full dosage of each type of immunisation, and the percentage who have completed their immunisation schedule, for each intervention area. In the more developed areas, Rangunia, Ghior and Mirzapur, coverage is very high at around 90% for each type of immunisation. The proportion receiving the full complement of eight vaccinations is lower in the old intervention area in Nasirnagar at around 76%. However in the more recently intervened area only 40% of children have been fully immunised.

The average drop-out rates across all areas are low. Of those children who have received the first dose of DPT and OPV, 97% return for the second dose and 96% of those return for the final dose. On average, only 12% of children who received their first immunisation fail to complete the full schedule. The majority of those who have dropped out have missed only the measles vaccination, though a small group have received only the first doses of DPT and OPV and BCG, and a few have missed the third doses of DPT and OPV in addition to measles. Since drop out is so low, it was decided to represent immunisation status in the multilevel event history analysis by a dichotomous variable: either fully immunised or not fully immunised.

Table 5.1: Percentage coverage rates for children aged at least 12 months by type of immunisation and intervention area

Type of immunisation	Intervention area					All areas
	Rangunia	Ghior	Mirzapur	Old Nasirnagar	New Nasirnagar	
DPT (All 3 shots)	93.5	90.5	94.8	80.5	45.4	74.0
OPV (All 3 shots)	93.6	89.7	94.1	80.5	42.4	73.0
BCG	95.2	92.3	96.2	85.9	53.5	79.0
Measles	90.3	88.6	92.7	76.6	46.6	72.1
Full Immunisation	90.0	87.8	91.7	76.1	39.9	69.6
n	1582	1037	659	3775	3011	10064

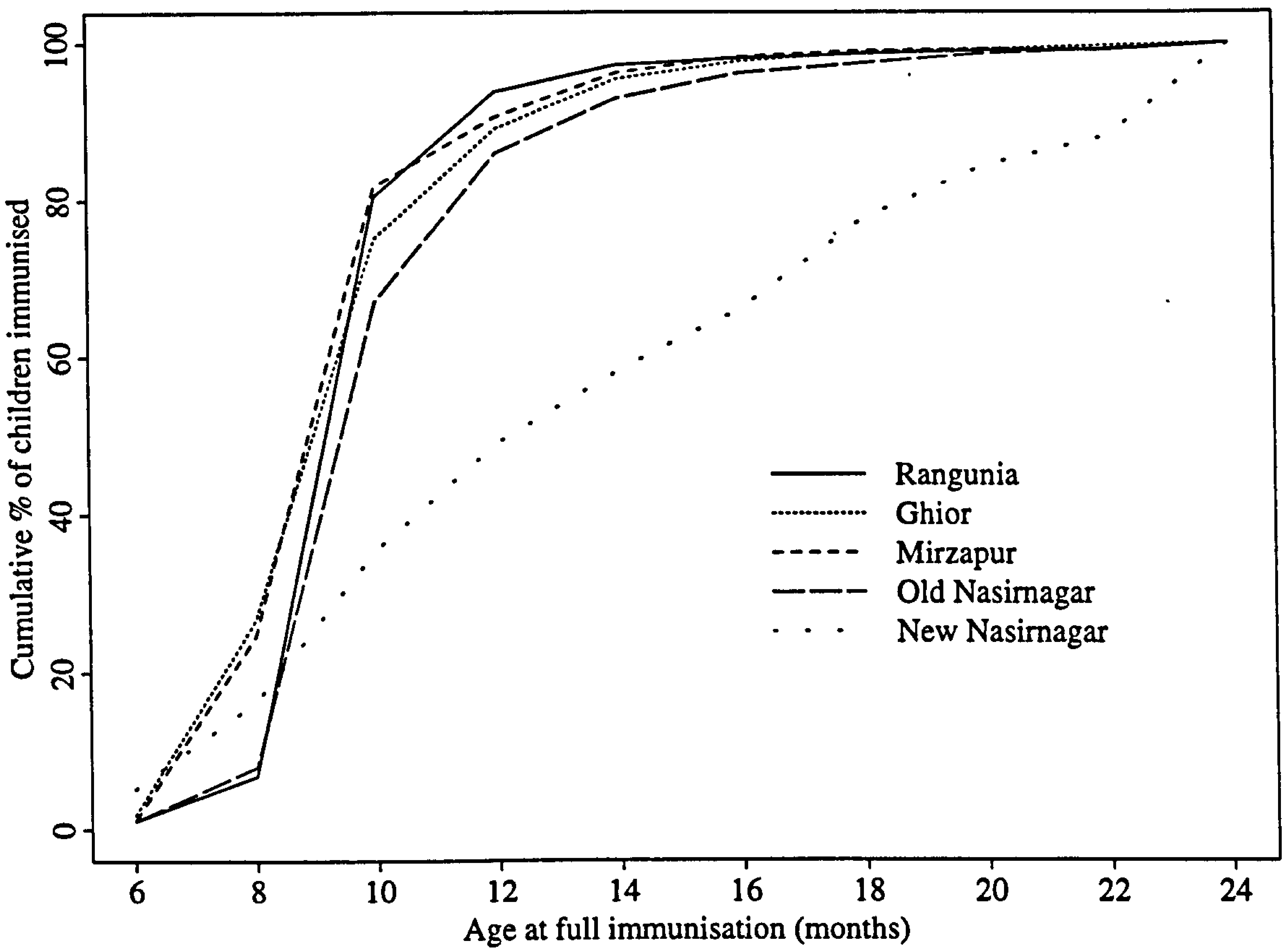
Table 5.2 shows the percentage of children aged at least 12 months who have completed their immunisation schedule in each village. In Rangunia and Mirzapur, there is little evidence of variation between villages. There is slightly more variation in Ghior with a 9% difference between the highest and lowest rates. Most of the variation between villages is in Nasirnagar, particularly in the new area where the immunisation rate ranges from 0% in Andrabaha to 69% in Beruin. Field workers in Nasirnagar report that the zero rate in Andrabaha may be due to vaccine shortages and storage problems caused by power failures.

Although children are supposed to have received all eight vaccinations by their first birthday, there is evidence that children in the new area of Nasirnagar are not being immunised until considerably later. In Rangunia, 95% of children who have been fully immunised have received their last immunisation by the end of their first year (see Figure 5.1). In the older area of Nasirnagar, this proportion is slightly lower at 85%, but in the new area only half have been immunised at the recommended optimal age. In this area, a large percentage of children are being immunised as late as 18 to 24 months.

Table 5.2: Percentage rates of full immunisation for children aged at least 12 months by village

Village	% children fully immunised	n
Rangunia		
Minagazirtila	88.7	575
Ichamoti	89.8	127
Sayed Bari	89.9	417
Kulkurmai	91.8	463
Ghior		
Tarail Kakjore	84.2	152
Jabra	85.2	385
Goaldungi	88.6	237
Baniajuri	92.8	263
Mirzapur		
Ranashal	90.9	77
Rashid deo hata	91.5	236
Dherua	91.7	84
Baimail	92.0	262
Old Nasirnagar		
Nurpur	70.4	1251
Gokarna	76.0	973
Choirkuri	78.7	164
Kunda	79.0	1025
Moslendapur	86.2	362
New Nasirnagar		
Andrabaha	00.0	81
Jethagram	17.9	571
Dighar	33.2	359
Brahmanshasan	33.9	316
Suchiura	35.4	147
Chotipara	42.7	157
Mohishber	48.9	515
Kahetura	51.9	484
Pathanisha	59.0	39
Bitui	64.7	207
Beruin	68.9	135
Overall average	69.8	10064

Figure 5.1: Distribution of age at immunisation for children who are fully immunised, by intervention area



To determine whether there is any evidence of clustering of immunisation uptake within households, the distribution of the number of children who are fully immunised per household is examined (Table 5.3). Around 34% of households contribute two or more children, accounting for 53% of children in the sample. Table 5.3 shows some evidence of clustering at the household level. In the majority of cases, households with two children have had both or neither immunised, and if there are three or more children the household tends to immunise at least two or none at all.

Table 5.3: Distribution of the number of children fully immunised per household

No. of children per household	No. of households	%	No. of children immunised	%
1	4776	66.5	0	27.8
			1	72.2
2	2022	28.2	0	15.2
			1	32.7
			2	52.0
3	293	4.1	0	13.7
			1	16.4
			2	33.8
			3	36.2
4+	88	1.2	0	14.8
			1	11.4
			2	17.0
			3	18.2
			4	33.0
			5	5.7
Total	7179	100.0		

A large number of demographic, socioeconomic and health-related covariates were considered in the multilevel analysis. Tables 5.4, 5.5 and 5.6 display the percentages of children aged at least 12 months who have received all eight immunisations by selected child-, mother- and household-level characteristics respectively. Covariates defined at the child level include the sex of the child, the year of birth of the child, maternal age, the number of doses of tetanus toxoid (TT) received by the mother during pregnancy, the occurrence of a child death in the household prior to the index child's birth and the women's savings group (WSG) membership status of women in the household in the period preceding the child's first birthday. Characteristics of the mother which were considered include parental education, both of the mother and the father, the father's occupation and the mother's relationship to the head of the household. Variables relating to the household include socioeconomic category, source of drinking water, sanitation, the building material of the house, food sufficiency, the existence of a separate kitchen and ownership of a radio or television.

A number of studies in South Asian countries have found evidence of a sex bias in the allocation of food and health care within a household which leads to excess female mortality (see, for example, Das Gupta 1987). In poor households, the low participation of female children in agriculture and income-generating activities in general means that scarce resources are often allocated to the males in the family. However in the areas considered here there is no evidence of any sex preference towards boys, which is possibly an effect of Save the Children's interventions. This is consistent with the findings of Amin (1993) in an analysis of immunisation uptake in Bangladesh. In this study, sex differentials were observed in *thanas* where the EPI had not yet been introduced, but were no longer evident in EPI covered areas. Similarly, Pebley and Amin (1991) found that health care interventions reduced excess female mortality in Punjab in India relative to pre-intervention levels.

Maternal age is often found to have an important impact on health-related behaviour, with older women being more slow to adopt modern medicine and preventative health care. In this case there is little evidence of such an effect, though there is a slight decrease in immunisation rates for mothers aged 40 years or more.

Table 5.4: Percentage rates of full immunisation for children aged at least 12 months by selected child-level variables

Variable	% immunised	n
Sex		
Male	69.9	5053
Female	69.4	5011
Year of birth		
1988	78.2	2595
1989	79.6	2328
1990	71.9	2171
1991-92	52.7	2970
Maternal age (years)		
< 20	68.6	1979
20-24	70.8	3398
25-29	70.2	2640
30-39	68.5	1859
40+	63.8	188
No. of TT shots received by mother during pregnancy		
0	59.7	5080
1	83.4	2841
2 +	33.8	2143
Previous child death in household		
Yes	72.9	824
No	69.4	9240
WSG membership in HH prior to child's 1st birthday		
None	64.9	5621
Mother of child only	72.3	2200
Other female only	73.2	989
Mother and other female	83.5	1254
Overall average	69.6	10064

The number of TT shots received by the mother during pregnancy is included as one may expect mothers who are immunised themselves and have, therefore, already accepted preventative health care to be more likely to have their children immunised. Also those women who have been immunised against TT have had some contact with health workers which may encourage them to seek health care for their children. There is strong evidence of such an effect as women who have received one dose of TT have a 24% higher immunisation rate than those women who have had no doses. However, the immunisation uptake rate is lower for mothers who have had two or more shots during pregnancy than for those who have had only one. This is as expected because women are given two doses if they are being immunised with TT for the first time, whereas those who receive just one shot have been immunised previously and require only a booster. Therefore, there is evidence of a cumulative effect of TT immunisation: a woman who has accepted TT for herself during a previous pregnancy and has had contact with health services on at least two occasions appears to be more receptive to the idea of child immunisation.

The occurrence of a child death in the household prior to the index child's birth was considered to explore whether a previous death leads to a change in health behaviours which would motivate a mother to adopt preventative measures to reduce the risk for subsequent children. However, there was no evidence here to support this.

Since the late 1970s, Save the Children has organised WSGs, each containing around 15-20 members, among poor rural women. Each WSG member deposits a small amount every month and then the collective group funds are invested and loans are made available to members for various income generating projects. Members of the groups also participate in discussions on health, family planning and social issues for the purpose of education and exchange. In recent years WSGs have been a key component in Save the Children's development programmes as a means of encouraging women to become economically active and giving them credit with the aim of enhancing their status within their household. In this study, the impact of WSG membership of women in the household during the period prior to the child's first birthday on immunisation uptake is explored. Four categories of household membership are considered: no member, the mother only, another female in the household only, and the mother and at least one other female in the household. The bivariate relationship shows that immunisation uptake is higher in households with at least one member, and highest for those where the mother and another woman are both members.

Levels of parental education have been found to have a strong impact on the propensity to utilise health services (Streatfield et al. 1990; Amin 1993) which leads to differentials in morbidity and mortality levels. In this case, education of both the father and the mother appears to have an impact on immunisation, though the relationship is stronger for maternal education. The immunisation rate is 17% higher among mothers who have been educated to at least secondary school level than among those who have received no education. Father's occupation is included as a proxy for the socioeconomic status of the child's family. The highest immunisation rates are observed among children whose fathers work in service with lower rates among farmers or day labourers.

The mother's relationship to the head of the household may be expected to influence immunisation uptake as it is an indicator of her status within the household. In her study of maternal health in Nasirnagar, Blanchet (1991) reports that if a woman is the daughter-in-law of the head of the household she often has a low status and most of the household decisions are made by the men and her mother-in-law. This is likely to extend to decisions regarding the health care of her children. Such decisions would probably be made by her mother-in-law who may have more traditional view towards modern preventative health care. However, in this case the bivariate analysis shows no evidence of such an effect.

Table 5.5: Percentage rates of full immunisation for children aged at least 12 months by selected mother-level variables

Variable	% immunised	n
Mother's education		
None	66.5	7632
Primary	77.2	1516
Secondary +	83.1	916
Father's education		
None	65.1	6235
Primary	75.8	1960
Secondary +	78.5	1869
Father's occupation		
Farmer	65.8	3224
Service	82.2	1125
Business/trade	73.8	1717
Day labourer	67.8	3646
Other	64.2	352
Mother's relationship to head of household		
Wife	70.1	6780
Daughter-in-law	69.0	2666
Overall average	69.6	10064

A number of socioeconomic indicators relating to the household have been considered. Save the Children uses a four category classification of socioeconomic status which is derived from 17 indicators, including those listed in Table 5.6. Class A corresponds to households with highest socioeconomic status, while class D corresponds to those with the lowest. Many of Save the Children's development programmes target only the poorest classes, C and D. There is little difference between the four categories in immunisation coverage rates. In fact one of the poorest groups, class C, has the highest rate, though the rate is slightly lower for class D. This suggests that the Save the Children interventions may have been successful in lessening socioeconomic differentials by increasing the access of the poorest women to health care services.

Table 5.6: Percentage rates of full immunisation for children aged at least 12 months by selected household-level variables

Variable	% immunised	n
Socioeconomic status		
A (High)	68.5	1072
B	70.1	1917
C	72.8	4501
D (Low)	64.2	2574
Sanitation		
Sanitary/water sealed	81.0	716
Other	68.7	9348
Source of drinking water		
No tubewell	54.6	280
Tubewell for part of year	69.2	1490
Tubewell all year round	70.2	8294
Household building material		
Mud/bamboo/straw	66.6	5633
Bamboo/corrugated iron	71.8	2233
Mud/wood/corrugated iron	75.3	2021
Brick/corrugated iron	75.7	177
Sufficiency of own food production		
Sufficient for < 3 mths	71.4	4925
Sufficient for 3-6 mths	71.0	1806
Sufficient for 6-9 mths	67.3	1216
Sufficient for 12 mths	66.9	989
Surplus	64.8	1128
Ownership of radio or TV		
No	68.3	8632
Yes	77.9	1432
Household has separate kitchen		
No	69.4	5650
Yes	70.0	4414
Overall average	69.6	10064

Households which have access to tubewell water for all or part of the year experience a 15% higher immunisation rate than those who do not have a tubewell. Also immunisation coverage is higher among households which have a sanitary or water-sealed latrine as opposed to an open latrine or none at all. These variables are included as proxies for economic status and also as indicators of the general health behaviour of the household. Lower immunisation rates are observed among the poorest households whose homes are constructed using mud, bamboo and straw than among wealthier households who use corrugated iron sheet. The other indicators of economic status, whether the household has a separate kitchen and food sufficiency, do not appear to have an important impact.

Ownership of a radio or television has been considered for two reasons; first, as another proxy for economic status and second in an attempt to assess whether the household would have been exposed to the media campaign on immunisation. The EPI campaign has made extensive use of television and radio to increase awareness about immunisation and to motivate parents to vaccinate their children, and here we find that immunisation coverage rates are higher among households which own a television or radio.

5.6 Results from the Multilevel Discrete-time Event History Model for Immunisation

In the next stage of the analysis, multilevel discrete-time event history models are used to evaluate the impact of the covariates considered in Section 5.5 net of all other factors. The primary aim of this analysis is to determine the extent of extravariation at the household and village level after controlling for the influences of important demographic, socioeconomic and health-related variables. In the selected model, the age of the child, intervention area, year of birth, maternal age, number of tetanus toxoid vaccinations received by the mother during pregnancy, membership of women's savings groups, the mother's relationship to the head of the household and several indicators of the household's socioeconomic status were found to have a significant effect on child immunisation uptake.

After selecting a main effects model, all pairwise interactions were tested for significance. Of particular interest are interactions between intervention area and other covariates. For example, one might expect household socioeconomic factors to play a less important role in the more developed and accessible areas with a high standard of health care provision. Also of interest are interactions with the child's age to test whether the timing of immunisation depends on background characteristics. To examine whether the effect of any covariate varies over time, interactions with the child's year of birth were also considered. In this analysis, two significant interactions were found between intervention area and the child's year of birth, and between area and the child's age.

Even after controlling for a wide range of background characteristics, a substantial amount of variation persists within intervention between villages areas, and within villages between households. In the general three-level model described in Section 5.4, the coefficients of any covariate (not necessarily restricted to those in the fixed part of the model) can vary randomly across households or villages. For instance, WSGs in some villages may be more successful in promoting good health behaviour among their members than those in other villages. Similarly, the socioeconomic situation of a household may not have as great an influence if the household has a high degree of health care knowledge. To test for these effects, a random coefficients model may be fitted where the coefficients of the WSG variable are allowed to vary randomly across villages, and the coefficient of a socioeconomic indicator varies across households. In the present analysis, however, no covariate was found to vary randomly at either the household level or the village level. Therefore, the results from the simplified model, where only the intercept term varies across level 2 and 3 units, are presented (Table 5.7).

Although the bivariate results in Table 5.5 show a strong relationship between maternal education and child immunisation status, this effect becomes nonsignificant after paternal education was added to the model. This is because they are highly correlated, since in rural Bangladesh an uneducated man would be unlikely to marry an educated woman, and the father's education has the stronger influence. The importance of the father's educational level is expected since the husband would hold more sway in decisions regarding household matters and on issues relating to their children's welfare. However, the effect of father's education is no longer significant once intervention area is added, indicating that the lower immunisation rate among fathers with no education can be explained by areal differentials in educational levels. It is also possible that education

Table 5.7: Parameter estimates (and standard errors) from the multilevel discrete-time event history model for immunisation

Variable	Estimate	(SE)
Constant	1.22	(0.36)
Age (months) (base=9-11)		
6-8	-3.61***	(0.08)
12-14	-1.00***	(0.08)
15-17	-1.89***	(0.13)
18+	-3.95***	(0.14)
Intervention area (base=Rangunia)		
Ghior	-0.56	(0.51)
Mirzapur	-0.89*	(0.53)
Old Nasirnagar	-0.73	(0.47)
New Nasirnagar	-3.59***	(0.43)
Year of birth (base=1988)		
1989	-0.16	(0.21)
1990	-0.47**	(0.20)
1991	-0.49**	(0.21)
1992	-1.91***	(0.21)
Maternal age (years) (base=20-24)		
< 20	-0.27***	(0.08)
25+	0.02	(0.07)
No. TT shots received by mother (base=0)		
1	0.58***	(0.08)
2+	0.33***	(0.08)
Household WSG membership (base=not mother)		
Mother only	0.27***	(0.08)
Mother and other female	0.46***	(0.10)
Mother's relationship to head of HH (base=wife)		
Daughter-in-law	-0.23***	(0.08)
Other	-0.38	(0.13)
Household sanitation (base=non-sanitary)		
Sanitary/water-sealed	0.39***	(0.13)
Household drinking water supply (base=tubewell)		
Other	-0.15**	(0.08)
Separate kitchen (base=no)		
Yes	0.24***	(0.07)
Area by year of birth interaction		
Ghior/1989	0.69**	(0.32)
Ghior/1990	0.74**	(0.33)
Ghior/1991	1.38***	(0.35)
Ghior/1992	0.38	(0.39)
Mirzapur/1989	1.18***	(0.37)
Mirzapur/1990	1.40***	(0.38)
Mirzapur/1991	1.52***	(0.38)
Mirzapur/1992	2.36***	(0.40)
Old Nasirnagar/1989	0.22	(0.24)
Old Nasirnagar/1990	0.25	(0.23)
Old Nasirnagar/1991	-0.28	(0.24)
Old Nasirnagar/1992	-0.41	(0.28)
New Nasirnagar/1989	0.87***	(0.24)
New Nasirnagar/1990	0.18	(0.25)
New Nasirnagar/1991	-0.82***	(0.28)
Area by age (months)		
New Nasirnagar/6-8	2.62***	(0.18)
New Nasirnagar/12-14	0.80***	(0.18)
New Nasirnagar/15-17	1.42***	(0.22)
New Nasirnagar/18+	3.00***	(0.20)
Household-level variance σ_u^2	0.85***	(0.07)
Village-level variance σ_v^2	0.42***	(0.13)

* p=0.10; ** p=0.05; *** p=0.01

is a proxy for factors operating at the community level and, therefore, becomes unimportant when these are accounted for. To test whether mass education as opposed to individual education plays a more important role in immunisation uptake, the average number of years of mother's and father's education in each village were considered. The level of education in the village was found to have a highly significant effect on immunisation acceptance, but this too disappeared after controlling for area.

Similarly, the effect of father's occupation is no longer important after controlling for area. This is likely to be a reflection of the differences in the type of employment opportunities open to men in Rangunia, Mirzapur, Ghior and Nasirnagar. Mirzapur and Ghior are a short distance from Dhaka and many men commute on a daily or weekly basis to work in the city. Rangunia is close to a large town where many of Bangladesh's largest and oldest industries are based, providing men with a range of employment options. In contrast, Nasirnagar is a very remote area where for most men there is no alternative to agricultural work. As Nasirnagar has the lowest immunisation coverage of the four areas, this would explain the low rates observed among farmers and day labourers.

5.6.1 Interpretation of the Fixed Covariate Effects

To aid interpretation of the covariate effects, 12-month cumulative probabilities of full immunisation are calculated from the estimated model. For illustration, probabilities for one of the study areas, Old Nasirnagar, for children born in 1989 are displayed in Table 5.8. The estimated probabilities are calculated for each covariate in turn while all other covariates are fixed at average values, which correspond to proportions in this case since all the covariates are categorical. At this stage, the household and village random effects are held at their mean value of zero.

Teenage mothers are less likely to immunise their children than mothers aged 20 years or more at their child's birth. Also, mothers who are married to the head of the household have an increased chance of immunising their children compared with daughters-in-law. Immunisation uptake is even lower among other relations, including sisters and sisters-in-law, who would have little influence over household decision-making. It is possible that these two effects are linked since a very young mother is unlikely to be the wife of the head of the household.

Mothers who have received at least one dose of the TT immunisation are more likely to immunise their children than those who have not been vaccinated against TT. Mothers who are being immunised against TT for the first time, and, therefore, receive two doses, have a lower odds of having their child immunised than women who have had previous contact with health workers for their own immunisation on more than one occasion and require only a booster. It is also possible, however, that mothers' immunisation is endogenous since it is another example of health seeking behaviour. If this is the case, the effect of TT immunisation vaccination may be overstated.

Immunisation rates are higher among mothers who have been WSG members prior to their child's first birthday than among non-members. If another woman in the household had also been a member, the chance of being immunised is further increased. However, there is no significant difference in immunisation uptake between households where no woman had been a WSG member and households with a woman other than the mother who had been a member. The WSGs organised by Save the Children have strong educational components on health and family planning and WSG meetings are forums for discussions on a variety of issues. WSG members would, therefore, be provided with information on immunisation and encouraged to have their children immunised. It is also possible, however, that there is a selection effect operating since WSG membership is voluntary and women who join WSGs could be more empowered; WSG members may have a greater say in household decision-making and a higher degree of social independence which would allow them to visit satellite clinics unaccompanied.

The effects of the household socioeconomic indicators are in the expected directions: households with sanitary latrines, tubewells for drinking water and separate kitchens are all more likely to seek child immunisation. After controlling for these variables, socioeconomic status as defined by Save the Children is no longer an important predictor of uptake.

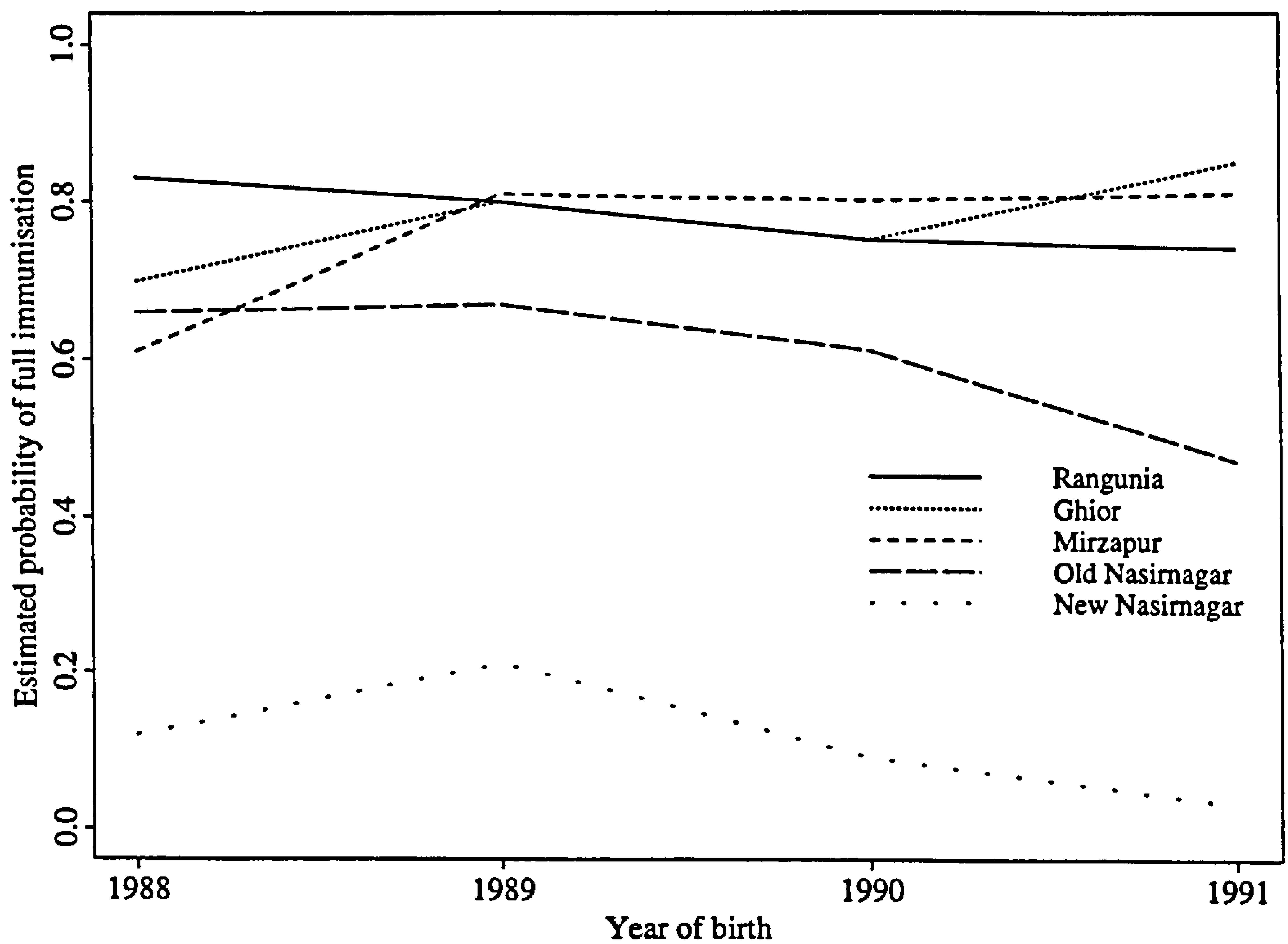
Table 5.8: Estimated 12-month cumulative probabilities of immunisation by selected characteristics: Old Nasirnagar 1989

Variable[a]	12-month probability of immunisation
Maternal age (years)	
< 20	0.64
20-24	0.70
25+	0.70
No. TT shots	
0	0.64
1	0.76
2+	0.71
Household WSG membership	
Not child's mother	0.66
Mother only	0.72
Mother + other woman	0.76
Mother's relationship to head of household	
Wife	0.71
Daughter-in-law	0.66
Other	0.62
Sanitation	
Non-sanitary	0.68
Sanitary	0.77
Drinking water supply	
Tubewell	0.70
Other	0.66
Separate kitchen	
No	0.67
Yes	0.72

[a]Probabilities calculated for each covariate in turn while all other covariates are fixed at average values. The household and village random effects, u_{jk} and v_k, are fixed at zero.

To examine the interaction between intervention area and the child's year of birth, estimated 12-month cumulative probabilities of immunisation have been calculated for each area by year (Figure 5.2). All other covariates and the household and village random effects are fixed at average values. Two of the areas, Mirzapur and particularly the more recently intervened area of Nasirnagar, show a large increase in immunisation uptake for children born in 1989. This is likely to reflect the high-profile government campaign to promote immunisation from 1989 to 1990. However, the most striking feature of Figure 5.2 is the steep decline in immunisation rates among children born after 1989 in both areas of Nasirnagar. There is also evidence of a slight decrease in Rangunia, but the decline is most marked in the more remote intervention areas. Save the Children workers in Nasirnagar offered two possible explanations. First, there were no government health promoters in the area until 1993. As outreach workers play an important role in providing information on immunisation and motivating mothers to attend immunisation clinics on their house-to-house visits, this is likely to have had a large impact on immunisation uptake. Second, field workers reported a shortage in vaccine supplies, mostly in Nasirnagar, between 1990 and 1991.

Figure 5.2: Estimated 12-month cumulative probabilities of immunisation, by year of birth and intervention area



Note: All other covariates are fixed at average values. The household and village random effects, u_{jk} and v_k, are fixed at zero.

As shown in Figure 5.1, children who were immunised in New Nasirnagar tended to complete their immunisation schedule at a considerably later age than those in other intervention areas. This effect was incorporated in the model by adding an interaction between area and child's age. Since no significant differences in the timing of immunisation were found between Rangunia, Ghior, Mirzapur and Old Nasirnagar, these four areas are combined in the interaction term. To illustrate the difference in the timing of immunisation between New Nasirnagar and the other areas, 12- and 24-month cumulative probabilities of immunisation are calculated from the estimated model for children born in 1989 (Table 5.9). As before, all other covariates and the household and village random effects are fixed at average values. In Rangunia, Ghior, Mirzapur and Old Nasirnagar, there is a small increase in the probability of immunisation between the ages of 12 and 24 months. This is expected since although most children in these areas are immunised at the optimal ages of 9-11 months, some of them receive their final measles immunisation slightly later than this. In the more recently intervened area of Nasirnagar, however, a much higher proportion of children are being immunised after the recommended age of 12 months which is reflected in the large increase in the cumulative probability of immunisation from 12 to 24 months.

Table 5.9: Estimated 12- and 24-month cumulative probabilities of immunisation by intervention area: 1989

Area[a]	12-month probability	24-month probability
Rangunia	0.80	0.88
Ghior	0.80	0.88
Mirzapur	0.81	0.89
Old Nasirnagar	0.67	0.76
New Nasirnagar	0.21	0.44

[a]Probabilities calculated for each covariate in turn while all other covariates are fixed at average values. The household and village random effects, u_{jk} and v_k, are fixed at zero.

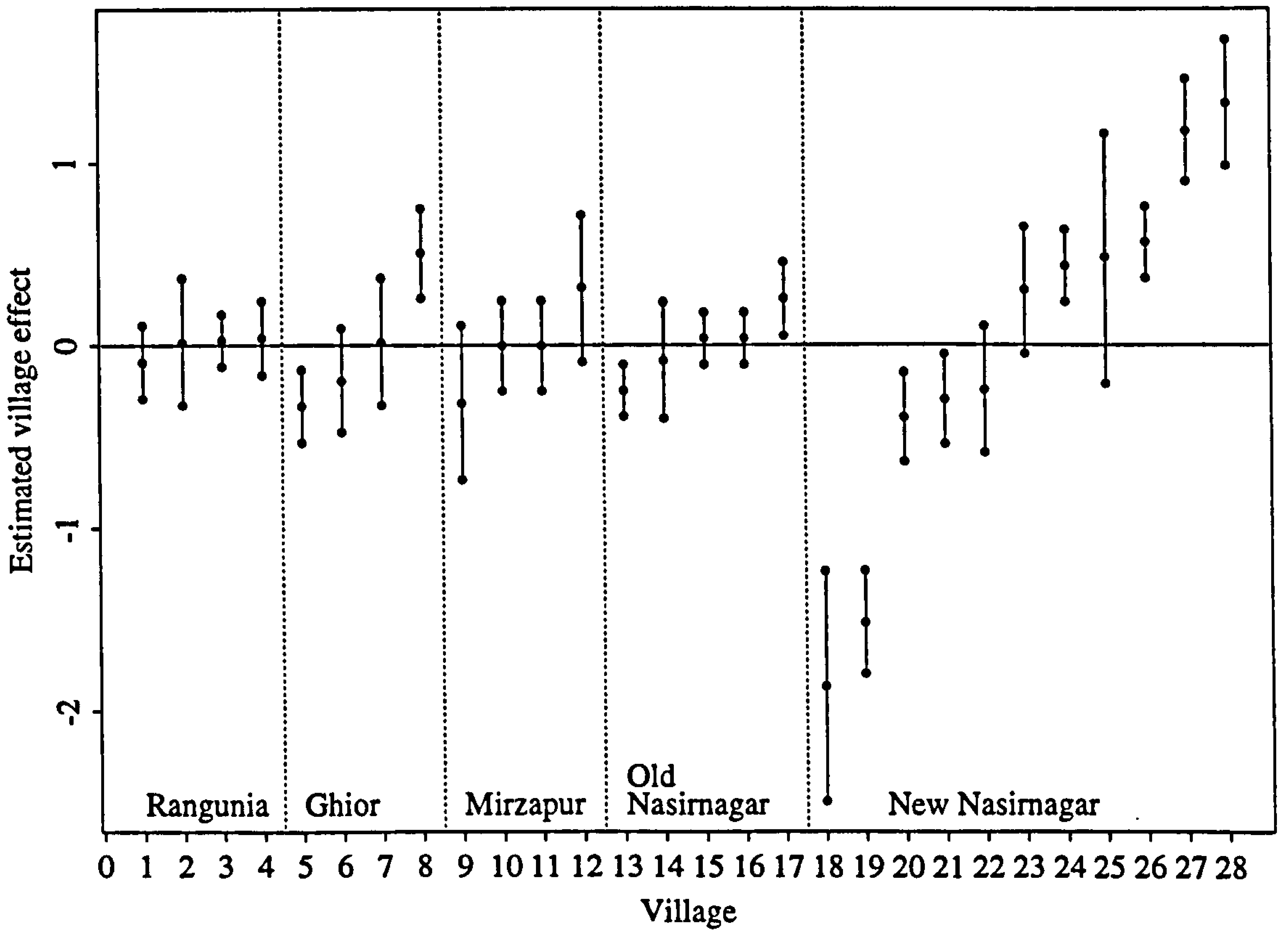
5.6.2 The Household and Village Effects

The results in Table 5.7 show that there is a large amount of extravariation both within intervention areas between villages, and within villages between households. To examine the the extent of village-level variation, estimates of the level 3 residuals v_k have been calculated for each village with simultaneous 95% confidence intervals (Figure 5.3). These represent 'village effects', that is, the contextual effect of the village in which a child lives on their chance of being fully immunised, after controlling for important observed background characteristics and the unobserved household effects. The village-level residuals can be used to obtain adjusted rankings of villages according to their average immunisation coverage rate. Confidence intervals have been constructed for each village-level residual to test for differences in immunisation uptake between villages. The procedure proposed by Goldstein and Healy (1995) is used to construct simultaneous confidence intervals to allow multiple comparisons between any pair of the 28 villages (a more detailed description of this procedure can be found in Section 2.5.2). A pair of villages have significantly different average immunisation rates at the 5% level if their associated confidence intervals do not overlap.

Since the confidence intervals for the four villages in Rangunia all overlap, there is no evidence at the 5% level of differences in immunisation uptake after controlling for the observed characteristics of their inhabitants and unobserved household-level effects. Similarly, there are no significant differences between villages in Mirzapur. In Ghior, Baniajuri has a significantly higher immunisation rate than two of the other villages. Most of the village-level variation is in Nasirnagar, particularly in the new area where there are two villages, Andrabaha and Jethagram, with much lower immunisation coverage than the others.

One possible explanation for this village-level variation is differences in their accessibility which would affect both mothers' ability to attend an immunisation clinic and the regularity of house-to-house visits made by the outreach workers. During the dry season, inadequate roads mean that some areas in Nasirnagar can only be reached on foot which makes it very difficult for health workers to make frequent visits to each household. In contrast, the villages in Rangunia, Ghior and Mirzapur are either within a short walking distance from each other or connected by good roads. Accessibility problems are likely to explain explain the zero immunisation rate in the village of Andrabaha in New Nasirnagar. This village is very isolated and is completely cut-off from other villages in the area during the monsoon.

Figure 5.3: Estimated village effects v_k with approximate simultaneous 95% confidence intervals



Note: All other covariates are fixed at average values. The household random effect, u_{jk}, is fixed at zero.

Village key:

Rangunia: 1.Minagazirtila 2.Ichamoti 3.Sayed Bari 4.Kulkurmai

Ghior: 5.Tarail Kakjore 6.Jabra 7.Goaldungi 8.Baniajuri

Mirzapur: 9.Ranashal 10.Rashid deo hata 11.Dherua 12.Baimail

Old Nasirnagar: 13.Nurpur 14.Choirkuri 15.Gokarna 16.Kunda 17.Moslendapur

New Nasirnagar: 18.Andrabaha 19.Jethagram 20.Dighar 21.Brahmanshasan 22.Suchiura 23.Chotipara 24.Mohishber 25.Pathanisha 26.Kahetura 27.Bitui 28.Beruin

Another reason for the unobserved variation between villages may be the influence of attitudes of community leaders and other families in the neighbourhood towards health care. This might explain why Gokarna in the old Nasirnagar area has the second lowest immunisation rate since in this village there is a religious leader who strongly opposes immunisation and discourages parents from having their children immunised. In New Nasirnagar, Beruin has the highest immunisation rate which is likely to be due to its proximity to the *thana* headquarters, but also to the co-operation of a local school teacher who motivates parents to attend immunisation clinics.

A further possible explanation for the village-level variation could be differences in motivational levels and the effectiveness of health workers, and whether any Save the Children workers are living in the area. In New Nasirnagar, Bitui has the second highest immunisation rate possibly because Save the Children workers live there, whereas Andrabaha with no Save the Children workers has the lowest rate.

There is also substantial extravariation within villages between households. This household effect represents factors operating at the household level which have not been observed. A number of socioeconomic household variables have been considered, which would suggest that this extravariation is unlikely to be due to economic reasons. Possible contributing factors are differences in knowledge and attitudes towards health care, particularly of the persons responsible for decision-making in the household. Another potential explanation is differences in the frequency of contact between households and health services.

In order to evaluate the relative impact of household- and village-level effects on immunisation, 12 month cumulative probabilities of immunisation have been calculated for a range of values of the household and village random effects, u_{jk} and v_k. For each intervention area, probabilities have been estimated for 1) the lowest estimates of u_{jk} and v_k in the area (the 'worst' household and village—representing the least favourable combination of the unobserved household and village effects from the estimated model); 2) the lowest estimate of u_{jk} and v_k=0; 3) u_{jk}=0 and v_k=0 (the 'average' household and village); 4) the highest estimate of u_{jk} and v_k=0; and 5) the highest estimates of u_{jk} and v_k (the 'best' household and village in the area—representing the most favourable combination of household and village effects). The estimated probabilities for each of these five combinations of observed household and village effects are shown in Table 5.10. These have been calculated for children born in 1989, while all other observed covariates are fixed at average values.

Table 5.10: The impact of household- and village-level extravariation on estimated 12 month cumulative probabilities of immunisation: 1989

Area[a]	Rangunia	Ghior	Mirzapur	Old Nasirnagar	New Nasirnagar
'worst' HH/ 'worst' village[b]	0.44	0.34	0.42	0.23	0.02
'worst' HH/ 'average' village[c]	0.47	0.43	0.51	0.27	0.10
'average' HH/ 'average' village[d]	0.82	0.81	0.83	0.69	0.28
'best' HH/ 'average' village[e]	0.95	0.94	0.94	0.91	0.69
'best' HH/ 'best' village[f]	0.95	0.97	0.96	0.93	0.93

[a]All other covariates are fixed at average values.

[b]lowest $\hat{u}_{jk}$ and $\hat{v}_k$ in area.

[c]lowest $\hat{u}_{jk}$; v_k=0.

[d]u_{jk}=0; v_k=0.

[e]highest $\hat{u}_{jk}$; v_k=0.

[f]highest $\hat{u}_{jk}$ and $\hat{v}_k$ in area.

In all areas, the probability of immunisation increases considerably when the estimate of u_{jk} corresponding to the 'worst' household is replaced by u_{jk}=0 (the 'average' household), even when the village effect v_k remains unchanged (rows 2 and 3 in Table 5.10). For example, in an 'average' village in Rangunia the probability of immunisation ranges from 0.47 for the 'worst' household to 0.82 for an 'average' household, while changing the value of the village-level effect has little impact (rows 1 and 2). This reflects the greater variability between households than between villages in each area, a result also found by Pebley et al. (1996) in Guatemala. Their study reveals substantial correlation of children's immunisation probabilities within families, but rather less intra-community correlation for immunisation uptake than for other health outcomes. They attribute this to the nationwide coverage of the EPI campaign.

In New Nasirnagar, however, where the village-level extravariation is highest, the impact of changing the village effect v_{jk} from the lowest estimate in the area to zero (rows 1 and 2), or from zero to the highest estimate (rows 4 and 5), is much more marked than in the other intervention areas.

5.7 Summary

In this chapter, a multilevel event history model is used to analyse the factors affecting child immunisation uptake in Save the Children USA's intervention areas in rural Bangladesh. Of primary interest is the extent to which unobserved factors operating at the child, household and village level play a role in determining a child's chance completing their immunisation schedule. This study has identified several demographic and socioeconomic characteristics which influence the uptake of immunisation, including the child's year of birth, maternal age, the number of TT shots received by the mother during pregnancy, the mother's relationship to the head of the household and household WSG membership. Even after controlling for these observable factors, however, there remains a large amount of unexplained variation in the acceptance of immunisation between intervention areas, within intervention areas between villages, and within villages between households.

The extravariation between households could be attributed to differences in attitudes and beliefs about preventative medicine or in the frequency of contact with the health services. Another possibility is the power relationships within the household. There is evidence of lower immunisation rates among daughters-in-law and teenage mothers. In these households, it has been shown that daughters-in-law are less likely to have their babies immunised. This is because it is probable that it is the mother-in-law who makes most of the household's decisions and she may have a more traditional attitude towards health care. These results suggest that Save the Children may improve their programmes through targeting them at the household level.

There is also some variation within intervention areas between villages, especially in the Nasirnagar area. Possible explanations for this community level variation are differences in accessibility, attitudes of local influential leaders and the effectiveness of health care workers in the area. The influence of community and religious leaders has been noted elsewhere (e.g. Samosir 1994). This study cannot identify which village-level characteristics are important in each village but, through the development of village rankings, has been able to identify those villages which do particularly well or poorly and hence can target research which identifies the key influences.

Chapter 6

Contraceptive Discontinuation in China

6.1 Introduction

In Chapter 3, the application of multilevel techniques to the study of contraceptive choice was illustrated using current status data from Bangladesh. However, the process of contraceptive choice or acceptance is just one aspect of contraceptive behaviour. Contraceptive use is a dynamic process and as the level of acceptance increases, a number of other important questions need to be addressed in order to assess the impact of a family planning programme on reducing fertility. Such questions include: What happens *after* acceptance? How soon and why is contraceptive use terminated? In other words, what is the risk of method discontinuation? The most commonly studied type of discontinuation is contraceptive failure which refers to pregnancies that have occurred while a method of contraception was being used. Clearly, if failure rates are high and these unintended pregnancies lead to unwanted births, the impact of failure on fertility levels will be considerable. Other reasons for discontinuation include side-effects or health concerns, problems with availability or accessibility of methods or family planning services, or to become pregnant. If a method is discontinued, it is then important to know whether contraceptive use is resumed later and, if so, which method is adopted. The process of 'switching' to a new method or continuing non-use following a discontinuation is another important aspect of contraceptive use dynamics. An analysis of the determinants of switching behaviour, using data from Bangladesh, is presented in Chapter 7.

In this chapter, we focus on contraceptive discontinuation and, in particular, the influences of social, economic and demographic factors. Event history models are used to study the timing of discontinuation, and a competing risks framework is used in order to distinguish between the different reasons for stopping use. In addition, woman-specific random effects are incorporated to account for unobserved heterogeneity. If there are women who have discontinued on more than one occasion during the observation period, these random effects also allow for the possibility of correlations between the durations of their repeated use intervals. This methodology is used to study contraceptive discontinuation in China, using retrospective data from the 'Two-per-Thousand' fertility survey of 1988.

6.2 The Study of Contraceptive Discontinuation

The continuation rate is defined as the proportion of acceptors who are still using a method after a given period of exposure to the risk of discontinuing, usually calculated for a period of 12 months. Its complement is the discontinuation rate which measures the proportion who are no longer using the method at the end of the period (Jejeebhoy 1990). The reason for discontinuation which tends to be of most interest to family planning policy makers is failure, that is enforced discontinuations after the occurrence of an unintended pregnancy. Three definitions of contraceptive failure are commonly used (Bongaarts and Potter 1983; Trussell and Kost 1987). The narrowest definition is theoretical failure which refers to failures that occur under ideal laboratory conditions and perfect use. This is a measure purely of method-failure. A more useful measure is use-failure which considers unintentional pregnancies that occur under average use conditions, allowing for inconsistent or incorrect use. An even broader definition is extended use-failure which includes all accidental pregnancies, even if the method was switched or use was interrupted or discontinued at the time of conception. In this study, the definition of use-failure is used.

The main objectives of the study of discontinuation and other contraceptive use dynamics are to measure these processes and to attempt to understand some of the underlying socioeconomic and programmatic factors which influence behaviour. Although measures of contraceptive acceptance and choice of method are necessary starting points in assessing the potential demographic impact of a family planning programme, the level of prevalence alone is not a sufficient measure of a programme's success. High levels of use do not always lead to lower fertility as in some countries high prevalence is offset by high rates of discontinuation and use failure among method acceptors (Bongaarts and Rodriguez 1990; Jejeebhoy 1990). Failure becomes a more prominent fertility determinant as fertility preferences decline and use of contraception increases. Bongaarts and Rodriguez (1990) show that as fertility targets fall, couples are practising contraception for longer periods and have an increased risk of experiencing an unintentional pregnancy. If contraceptive effectiveness is low, a large proportion of couples will experience a contraceptive failure which, if converted to unwanted births, will contribute substantially to a country's level of fertility.

While failure is the type of discontinuation which is likely to have the most impact on fertility levels, it is also important to study other reasons for discontinuation. Method-related reasons tend to be most frequently cited as reasons for stopping use. Discontinuations for these reasons imply dissatisfaction with specific method characteristics such as use-effectiveness, side-effects or inconvenience with use. The rate of discontinuation due to method-related reasons is an indicator of method acceptability. Further, a discontinuation followed by a switch to another method can lead to an increased risk of unintentional pregnancy because of the learning curve associated with contraceptive use. Since failure rates tend to be highest during the first six months of use when the user is still unfamiliar with the method, couples who discontinue and have to learn how to use a new method will pass through that high risk period again. In many cases contraceptive use is abandoned altogether, showing a dissatisfaction not only with the current method but with all other available methods. Clearly discontinuations of this type will have a more direct impact on fertility levels. Therefore, to assess the merits of any method of contraception, both its failure rate and other method-related discontinuation rate need to be considered. The ideal method would have a low failure rate and high level of acceptability, measured by low rates of discontinuation due to side-effects, health concerns and inconvenience.

One of the main reasons for studying contraceptive use dynamics is to provide accurate information on contraceptive effectiveness for family planning policy makers. The processes of method choice, continuation and switching are influenced not only by individual characteristics, but by the level of family planning service provision. If accessibility or availability of a range of methods is low, this is likely to have an adverse effect on prevalence and continuation. Information on use dynamics allows the adequacy of the family planning programme to be evaluated. In addition, a set of reliable failure rates is useful for directing future policy as it allows the calculation of the level of prevalence required to achieve a target level of fertility. Also of interest are the effects of individual characteristics on use patterns as this enables those subgroups of the population with particularly low levels of use and continuation to be identified. Some women may be more likely to use contraception, to do so more effectively, and to continue use than others. For the success of a future family planning programme, one needs to be able to identify those women who are most likely to experience problems with using different methods of contraception so as to be able to offer couples a suitable mix of methods with which to control their fertility. Information on the effectiveness and health risks of methods is also necessary for couples if they are to make an informed choice on which method to adopt. The reasons for discontinuation will vary according to the method used. For example, side-effects are common among pill users, but are unlikely to be experienced by users of traditional methods. Therefore, studying method-specific discontinuation rates can give valuable insights into the advantages and disadvantages of each method which is important information for both policy makers and prospective users.

The study of contraceptive discontinuation is particularly timely in China which has undergone a rapid fertility transition after the introduction of the one-child policy in 1979. By 1988, around 70% of currently married women were using contraception, with the most common methods being IUDs and female sterilisation (Wang and Diamond 1995). This has resulted in a considerable drop in the TFR to just over two children per woman. However, although contraceptive prevalence is high, there is evidence that first year failure rates in China are among the highest in the world. Failure rates for all methods are above the levels of many countries in Asia, Latin America and Africa (Moreno and Goldman 1991; Wang and Diamond 1995). In particular the failure rate for the IUD, the most commonly used reversible method, is very high at about 10% (Wang and Diamond 1995). These levels are likely to have a marked impact on the level of fertility in the future.

6.3 Population Growth and the Family Planning Programme in China

The Chinese government has been concerned about the country's rate of population growth since the foundation of the People's Republic of China in 1949. From 1949 to 1976, the population of China grew from 542 million to 933 million, representing an annual growth rate of 2% (United Nations 1989). This rapid population increase was the result of a rise in the birth rate coupled with a sharp decline in the death rate, especially for infant mortality. Before 1949, the high death rate had counteracted the high birth rate, leading to a low rate of natural increase. After Liberation, however, large-scale improvements in public health led to a reduction in the incidence of infectious disease which lowered the death rate. During the 1950s, the government became increasingly concerned about population growth and in 1956 the first family planning programme was launched. This programme was disrupted in 1958, however, by the launch of the Great Leap Forward and Communisation movements to collectivise agriculture and decentralise industry. During this period of economic hardship, the total fertility rate (TFR) plummetted from around six in the early to mid-1950s to 3.3 in 1961 (Poston 1986). Between 1961 and 1963, the economy began to recover and the government adopted a pronatalist policy, arguing that a large population was necessary to achieve the rapid expansion of China's economy. As a result, the TFR rose to a peak of 7.5 in 1963 and the rate of natural increase surged once again (Poston 1986).

In the late 1960s, it became apparent that a pronatalist policy would actually hinder China's future economic development. High rates of population growth in an already densely populated country such as China would lead to a number of problems in the future, including mass unemployment, a drop in living standards, and pressures on housing authorities, medical facilities and other services. The government realised that rather than encouraging couples to have more children, they would have to adopt drastic measures to control population growth. In 1971, a major family planning campaign was launched and in 1973 family planning was delivered to people's homes for the first time. Family planning became a national priority as the government advocated '*late* marriage, *longer* birth intervals, and *fewer* children' (Tien et al. 1992). This led to a rapid decrease in the TFR from 5.4 in 1971 to 2.7 in 1979 (Poston 1986). In the late 1970s, however, there were fresh concerns about population growth as the cohorts of children born during the baby booms of the 1950s and early 1960s were approaching childbearing age. Therefore, in 1979 the government announced its most ambitious family planning programme yet—the 'one-child' policy. To achieve their long-term economic goals, the government set a population target of no more than 1.2 billion for the year 2000 (Peng 1991). A set of incentives to persuade couples to have only one child was introduced. Couples who had just one child received a number of benefits which included priority for housing, monthly subsidies, higher pensions on retirement and free education for the child. A couple who had a second child after accepting a one-child certificate was subject to financial penalties. It should be emphasised, however, that the policy has never been uniformly 'one-child' across the whole country. In particular, rural areas and ethnic minorities have always enjoyed a more lenient policy. These groups of the population are permitted to have a second child without penalty if the first child is female. Second births tend to be more strictly discouraged in urban areas and among the Han majority. There is also some areal variation in policy as some provinces allow couples to have a second child provided there is a long interval between the first and second birth.

In the early 1980s, the campaign proved highly successful. The TFR continued to fall and the 1982 Census reported that almost half of married women aged 20-29 had only one child (Poston and Yu 1985). However, some of the hard-line tactics used in the campaign, such as mandatory IUD insertions for women with one child, abortions for unauthorised pregnancies, and sterilisation for couples with two or more children, were extremely unpopular and met with strong resistance, particularly in rural areas (Greenhaulgh 1986; Peng 1991). In 1984, a more realistic and flexible policy was introduced, the aim of which was to increase voluntary participation by offering a policy which was more acceptable to the people (Greenhaulgh 1986; Hardee-Cleaveland and Bannister 1988). Since 1984, a wider range of contraceptive methods has been made available as alternatives to the IUD and sterilisation. Family planning workers make regular visits to deliver free contraceptive supplies and to provide education on contraceptive use. In addition, financial assistance is made available to women travelling to clinics for IUD insertions and sterilisations. The conditions under which couples could have two children were also expanded. Couples with two children were given the option of signing a contract with the government agreeing not to have a third child. As a consequence of this relaxation in policy, however, the TFR began to rise again and in 1986 the net increase in the population was 1.68 million more than the target figure for that year (Hardee-Cleaveland and Bannister 1988). In 1990, the target population size for the year 2000 was modified from 1.2 billion to 1.29 billion (Poston 1992), and it now seems likely that the projected population will be nearer 1.3 billion (Tien et al. 1992).

There are a number of features of the Chinese family planning programme which have contributed to its success. The strong commitment of the national leadership to population policy since 1971 is a major factor. Also important is China's highly organised social structure which enables strong political control of policy at both the top and local levels. Central government sets birth quotas for each province and pays for contraceptive supplies, while implementation of the programme is organised by local units. Another important element of the programme is a strong information, education and motivation component which uses a variety of approaches to encourage couples to conform. Finally, a range of free contraceptive methods have been made widely available and accessible.

Since the introduction of the one-child policy in 1979, contraceptive prevalence in China has remained very high. In 1988, the prevalence rate was estimated at 71.1%, and 98.6% of users were using effective modern methods (Wang 1994). Sterilisation was the most common method, accounting for 48.6% of all contraceptive use, followed by the IUD used by 41.6% of users. Levels of use and the methods used vary according to the type of region of residence. Prevalence tends to be lower in rural areas and the use of modern methods other than the IUD is more common in urban areas. There are also large variations in use patterns across geographical regions and ethnic groups. Other important influences on a woman's likelihood of using contraception in China include her level of education attainment, occupation, age and characteristics of her fertility and contraceptive use history (Wang 1994).

As mentioned earlier, failure rates in China are among the highest in the world. Due to the high risk of contraceptive failure, abortion is used as a backup method by many women. As a result, the number of abortions nearly tripled between 1971 and 1988 (Tien et al. 1992) and it is estimated that around 70% of abortions are due to contraceptive failure (Delfs 1990). Although abortion rates are high, around 37% of IUD failures lead to live births (Wang and Diamond 1995). Wang and Diamond (1995) estimate that 7% of the general fertility rate of currently married women aged 15-49 for a 12-month period was attributable to contraceptive failure, mainly due to the high failure rate associated with IUD use. Therefore, contraceptive failure has serious implications for both women's health and national fertility levels. Reasons for discontinuation other than failure, especially those which indicate some dissatisfaction with a method, are also of interest. In particular, IUD expulsions are common in China due to poor insertion techniques and low quality IUDs (Wang 1994). Side-effects such as pain or excessive bleeding are also frequently reported problems among IUD users.

6.4 The Chinese National Survey of Fertility and Contraceptive Prevalence 1988

The data for this study are from the National Survey of Fertility and Contraceptive Prevalence, conducted by China's State Family Planning Commission from 1 July to 15 July 1988. This survey represented a sample of two per 1000 persons in mainland China, targeting ever-married women aged 15-57. All 31 administrative areas—three municipalities directly under central government, five autonomous regions and 23 provinces—were surveyed, including for the first time Tibet. Within provinces, sampling fractions were used in order to obtain data which was representative for each province. A total of around half a million women were interviewed.

The information collected in the survey included complete pregnancy and contraceptive use histories from each woman. For each continuous period of contraceptive use, the month and year of starting use, the method used, the reason for using the method, the date of stopping use and the reason for discontinuation were all recorded. Demographic and socioeconomic data were also collected on the woman, her husband and other members of the household.

Due to the reconstruction of the data required to fit a discrete-time event history model and the further expansion of the data set needed to extend to a multilevel competing risks model (Sections 2.8.2 and 4.6.2), it is computationally infeasible to use the complete sample of 500,000 women. Therefore it was decided to take a 10% subsample and to consider only four provinces. Beijing, Shandong, Guizhou and Gansu were chosen in an attempt to reflect some of the geographical and socioeconomic differentials in fertility and contraceptive use across China. The strength of the family planning programme varies between developed and less developed provinces, between urban and rural areas, and between the majority Han nationality and ethnic minorities. Beijing, which includes the capital, is well developed and has a strong family planning policy. Shandong is an agricultural province, but is more highly developed than Guizhou and Gansu, the other rural provinces. Gansu is one of the poorest provinces and is located in the remote northwest; Guizhou is characterised by its high concentration of ethnic minorities, who enjoy a more lenient family planning policy.

All women who have some experience of using reversible methods of contraception are considered; sterilisations are excluded. A woman who has been sterilised, however, is not excluded completely from the analysis: intervals of reversible method use before the sterilisation are retained in the analysis sample. The analysis is further restricted to intervals starting after 1979 to control for period changes in contraceptive behaviour following the introduction of the one-child policy in 1979. The final sample consists of 2,375 women who contribute 3,361 spells of use.

6.5 Methodology

There are a number of features of contraceptive discontinuation data which need to be considered in the analysis. First, one must distinguish between different reasons for stopping because the factors affecting the risk of one type of discontinuation may not be the same as those affecting the risk of another type, or because the same factors may affect the risks in different ways. Second, the duration of use before a discontinuation also needs to be considered as the risk of stopping typically decreases as duration increases; method-related problems such as side-effects or failure tend to be more likely to occur in the first few months of use. Finally, as with all event history data, there will be right-censored observations as many women will initiate use of a method during the observation period and continue use of the same method beyond the time of the survey. These features suggest the use of multiple-decrement life table techniques (Section 4.6.1) or competing risks event history methodology (Section 4.6.2).

There are relatively few studies on contraceptive discontinuation and switching. Most research on use dynamics focuses on contraceptive prevalence rather than on behaviour after method acceptance. Those studies that have examined discontinuation tend to use life table methodology. Grady et al. (1983) use a multiple increment-decrement life table to calculate 12-month cumulative net rates of failure and continuation. A multiple-increment table is used because a large proportion of use intervals began before the start of the observation period. For example, a woman who has already been using a method for six months is included in the life table from six months on. The associated single-decrement life tables are then calculated to obtain gross failure and continuation rates, that is, the probabilities of failure and continuation with all other competing risks eliminated. Multiple-decrement life tables have also been used by Choe and Zablan (1991) and Kost (1993) to study contraceptive discontinuation by reason. Recently, Ali and Cleland (1996) used life table techniques in a comparative analysis of discontinuation in six developing countries. Associated single-decrement life tables were calculated to yield hypothetical 12- and 24-month probabilities of cause-specific discontinuation in the absence of competing causes.

The disadvantage of the life table approach is that the number of control variables that can be considered is usually limited by sample size. Although all of the above studies have looked at bivariate relationships between the probability of discontinuation and a set of background characteristics, one would usually wish to look at the effects of a larger number of variables simultaneously. If life table methodology is used, a separate life table must be calculated for every subgroup defined by the cross-classification of all covariates. This limits the number of variables that can be introduced as the size of some subgroups will often be too small to provide reliable estimates of discontinuation rates. To overcome this problem, Schirm et al. (1982) used a log-linear hazards model to study the determinants of contraceptive failure. This is essentially a discrete-time event history method which can be used if all covariates and the duration variable are categorical. The data are aggregated to form a contingency table, where each cell represents a combination of the covariate values for a particular duration. The dependent variable is the logarithm of the failure rate which is assumed to follow a Poisson distribution. The same approach was adopted by Hammerslough (1984) to study discontinuation, where all types of discontinuation are grouped together. Grady et al. (1988) used a discrete-time piecewise-constant hazards model to study discontinuation due to two reasons: changing to another method or abandoning use of contraception altogether. Separate models were fitted to yield gross rates of each type of discontinuation which were then converted to net rates. Bracher and Santow (1992) used a continuous-time competing risks model to examine the correlates of discontinuations due to failure, side-effects and general dissatisfaction with a method. Since the likelihood for a continuous-time competing risks model can be factorised into separate components corresponding to each type of discontinuation, standard survival analysis techniques could be applied (Section 4.6.2). This involved fitting a separate model for each type of discontinuation of interest, treating all other reasons for discontinuation as censored.

In the present analysis, multiple-decrement life tables are calculated to explore the relationships between discontinuation rates and selected background characteristics. Discrete-time competing risks event models are then used to assess the simultaneous impact of these variables on different types of discontinuation.

A total of 10 possible reasons for discontinuation were listed in the questionnaire: 1) side-effects, 2) illness, 3) failure, 4) to change to another method, 5) expulsion (IUD users only), 6) to get pregnant with a license, 7) to get pregnant without a license, 8) menopause, 9) widowed, marital separation or divorced, or 10) another unspecified reason. In this analysis, these reasons are grouped into three categories: 1) failure, 2) method-related reasons (side-effects, illness, to change to another method and expulsion), and 3) nonmethod-related reasons (to get pregnant and other unspecified reasons). Intervals which end because the woman has reached menopause or because she is widowed or separated from her husband are treated as right-censored. It seems reasonable to assume that these women would have continued use if they had any need for contraception.

As discussed in Section 4.6.2, the discrete-time competing risks model can be estimated as a multinomial logit model. Starting with the month when a new method was adopted, a multinomial response is recorded for each month until either the method is discontinued or the time of the survey (censored cases), whichever event occurs first. Observation on a woman then ends unless she starts to use another method, in which case a new interval begins. The observation period is from January 1980 to the end of June 1988. Thus the maximum exposure time for a woman who starts use in January 1980 is 102 months. The multinomial response indicates the woman's contraceptive use status at each month of observation. A woman can be in one of four mutually exclusive and exhaustive states at any time after initiating use: she may have 1) experienced a contraceptive failure, 2) discontinued for a method-related reason, 3) discontinued for a nonmethod-related reason, or 4) continued use of the same method. As an example, suppose a woman has been using a method successfully for five months before experiencing a failure in the sixth month of use. She will contribute six records to the data set with the response pattern 4 4 4 4 4 1.

The single-level competing risks model can be specified as the following multinomial logit model

$$\log\left(\frac{h_{rti}}{h_{4ti}}\right) = \alpha_{rt} + \mathbf{x}'_{rti}\beta_r, \quad r = 1, 2 \text{ or } 3, \tag{6.1}$$

where h_{rti} is the hazard of a discontinuation of type r at time t for use interval i and h_{4ti} is the hazard of continuing use of the same method (the reference category). To examine the effect of duration on the risk of discontinuation, a covariate for duration is included in the model. The duration effects are represented by α_{rt}. $\mathbf{x}_{rti}$ contains a set of covariates which may be woman-specific (e.g. education) or use interval-specific (e.g. method). The covariates may even be time interval-specific, although in this case no time-dependent information is available. The vector β_r contains the regression parameters associated with $\mathbf{x}_{rti}$. A separate set of parameters is estimated for each contrast of a type of discontinuation with the 'still using' reference category.

The model in (6.1), however, assumes that all variation in the hazard rate can be explained by the covariates in $\mathbf{x}$. In most situations, such an assumption is untenable as there will be risk factors that have not been observed, perhaps because they are difficult to measure, and as such cannot be included in the model. Further, a woman can experience more than one discontinuation over the study period. In the presence of unobserved heterogeneity at the woman level, the durations of repeated use intervals for the same woman may be correlated. To take into account this extravariation between women, (6.1) can be extended to a two-level model

$$\log\left(\frac{h_{rtij}}{h_{4tij}}\right) = \alpha_{rt} + \mathbf{x}'_{rtij}\beta_r + \mathbf{z}'_{rtij}\mathbf{u}_{rj}, \quad r = 1, 2 \text{ or } 3, \tag{6.2}$$

where h_{rtij} is now the hazard of a discontinuation of type r at time t for use interval i of woman j. The vector $\mathbf{u}_{rj}$ contains woman-specific random effects, representing unobserved factors operating at the woman level. $\mathbf{z}_{rtij}$ is a set of covariates, the coefficients of which may vary across women. For instance, the effect of unobserved woman-level characteristics may vary according to the type of method used; unobserved biological factors would be likely to play a role among pill or IUD users, but not among users of traditional methods. Such an effect can be tested by the inclusion of method in the random part of the model. The random effects $\mathbf{u}_{rj}$ are assumed to follow multivariate normal distributions with mean 0 and variance-covariance matrix $\mathbf{\Omega}_r$, r=1,2 or 3. A separate set of random effects is estimated for each contrast of the risk of a type of discontinuation r with the risk of continuation, and these may be correlated to allow for shared unobserved risk factors across the competing alternatives.

If there are women who experience repeated discontinuations of the same type over the observation period, one can test whether the woman-level random effect is constant for recurrent events. In other words, it is possible to determine whether there is unobserved heterogeneity between the durations of use intervals within each individual. Such unobserved heterogeneity between the durations for each individual can be accounted for by incorporating an additional use interval-specific random effect in (6.2) to obtain a three-level model. If there is no significant extravariation at the use interval level, it can be assumed that the unobserved woman-level characteristics that influence the risk of a discontinuation of type r, represented by $\mathbf{u}_{rj}$, affect all observed use intervals for that woman.

As described in Section 4.7.4, both the two-level model in (6.2) and the extended three-level model can be fitted using estimation procedures for multilevel multinomial models.

6.6 Preliminary Analysis

6.6.1 Contraceptive Prevalence in the Four Study Provinces

Table 6.1 shows method-specific contraceptive prevalence rates for currently married women aged 15-49, for each of the four provinces considered. The overall prevalence rate is highest in Beijing province and lowest in one of the agricultural provinces, Guizhou. The low rate of use in Guizhou can be partly explained by its high ethnic minority population for whom the family planning policy is more relaxed. There are large provincial differences in the method-specific use rates. Beijing, in particular, has considerably lower proportions using sterilisation or IUDs than the other provinces. In Beijing, 30% are using the pill or condom, compared to less than 5% elsewhere. There is also evidence of variation in method mix between the agricultural provinces, Shandong, Guizhou and Gansu. Sterilisation is the most commonly used method in these provinces, with a particularly high rate in Gansu. There are large differences in male sterilisation rates across the four provinces. Although women are more likely to be sterilised than their husbands in each province, the proportion of male sterilisation acceptors is much higher in Shandong and Guizhou than in Beijing or Gansu.

Table 6.1: Percentage of currently married women aged 15-49 using contraception by province and method

Province	Sterilisation		IUD	Pill	Condom	Others	Total	n
	Male	Female						
Beijing	0.2	7.9	36.9	13.3	16.7	4.4	79.4	1305
Shandong	15.0	23.2	30.7	1.7	3.0	1.0	74.6	1759
Guizhou	12.7	19.7	27.2	1.3	0.5	0.6	62.0	1272
Gansu	0.1	45.7	19.4	1.2	0.7	1.4	68.5	1384

6.6.2 Multiple-Decrement Life Table Discontinuation Rates

Multiple-decrement life tables were used to calculate net rates of discontinuation. Table 6.2 presents 24-month cumulative probabilities of stopping use according to the reason stated for discontinuation by selected background characteristics. A range of demographic and socioeconomic covariates were considered, including the woman's age at the start of use, the number and sex of the woman's living children, her level of education, her ethnicity, the type of region of residence (rural or urban), and province of residence. Also considered were two factors relating to past experience with contraception: the number of use intervals preceding the index interval and the number of previous failures. These were calculated using all intervals in a woman's contraceptive history rather than only those which began after 1980.

Overall, failure is the major reason for discontinuation. Nearly 20% of users have experienced a contraceptive failure within two years of initiating use. Approximately 13% have discontinued for some other method-related reason. The woman's age at the start of use has a strong impact on the risk of failure; failure rates decrease as age increases. The motivation to space and limit births and thus to avoid an unwanted pregnancy is likely to increase with age. The low rate among women aged 35 or more may also be a reflection of declining fecundity with age. There is little effect of age on other method-related reasons. The risk of a nonmethod-related discontinuation, however, is also lower for older women. This is as expected because this category of discontinuers consists mainly of women who have stopped to become pregnant. It is likely that most women aged 30 or more will already have at least one child and would not plan to have another.

To examine the impact of the number and sex of a woman's children on her likelihood of discontinuing use, a variable representing the sex composition of her living children was constructed. If a preference toward male children exists, women with no sons might be expected to have higher discontinuation rates than women with at least one son as they are less motivated to avoid pregnancy. The probabilities in Table 6.2 show a marked decline in the failure rate as the number of children increases, but no evidence of a sex differential. The rate of other method-related discontinuation increases with the number of children, though there is evidence of a decrease for women with two girls. As expected, the risk of stopping for nonmethod-related reasons is highest for couples with no children and decreases sharply for those with one or more child. There is also a suggestion of son preference. For women with one girl, the rate of a method-related discontinuation is almost double that for women with one boy. This is likely to be a reflection of the family planning policy which allows some couples to have a second child if the first is a girl. Among couples with two children, the chance of a nonmethod-related discontinuation is almost three times higher if they are both girls than if they are both boys.

Another factor which might be expected to affect contraceptive behaviour is the change in family planning policy in 1984. To examine the effect of this on discontinuation, the year during which a new method was adopted was considered. Two time periods were used: 1980 to the end of 1983, when the family planning campaign was at its peak, and 1984 to 1988, when the policy was more relaxed. Although there is little difference in failure rates between the two periods, the rate of other method-related discontinuation is lower in the post-1983 period. The likelihood of stopping use for a nonmethod-related reason is slightly lower for intervals started after 1983.

Table 6.2: 24-month cumulative probabilities of discontinuation by selected background characteristics

Variable	Reason for discontinuation Failure	Other Method-related[a]	Nonmethod-related[b]	Continue	n
Age at start (years)					
15-24	0.240	0.126	0.061	0.573	975
25-29	0.217	0.135	0.053	0.596	1500
30-34	0.150	0.131	0.032	0.687	625
35+	0.066	0.094	0.024	0.816	261
No. and sex of children					
None	0.449	0.055	0.328	0.168	166
1 boy	0.200	0.106	0.016	0.677	1057
1 girl	0.217	0.096	0.029	0.658	855
2 boys	0.180	0.181	0.034	0.604	207
2 girls	0.178	0.123	0.100	0.599	186
1 boy, 1 girl	0.140	0.184	0.073	0.603	418
3+	0.143	0.187	0.026	0.644	472
Year of starting use					
$<$ 1984	0.209	0.150	0.042	0.598	1393
$\geq$ 1984	0.185	0.109	0.056	0.650	1968
Education					
None	0.140	0.183	0.039	0.639	822
Primary	0.192	0.107	0.040	0.660	555
Secondary	0.210	0.129	0.041	0.620	1147
Senior+	0.242	0.088	0.077	0.594	837
Ethnicity					
Han	0.201	0.130	0.049	0.620	3134
Minority	0.150	0.106	0.053	0.690	227
Province					
Beijing	0.224	0.111	0.062	0.603	1515
Shandong	0.198	0.146	0.036	0.620	962
Guizhou	0.146	0.123	0.035	0.696	475
Gansu	0.147	0.163	0.042	0.647	409
Region					
Urban	0.213	0.100	0.059	0.627	1303
Rural	0.187	0.147	0.042	0.623	2058
Method					
IUD	0.147	0.123	0.026	0.704	2098
Pill/condom	0.272	0.143	0.093	0.493	1063
Other	0.317	0.118	0.061	0.504	200
No. of previous uses					
0	0.201	0.133	0.060	0.607	2046
1	0.209	0.131	0.031	0.628	814
2+	0.163	0.109	0.034	0.694	501
No. of previous failures					
0	0.191	0.140	0.057	0.611	2493
1+	0.215	0.096	0.026	0.663	868
Overall	0.197	0.129	0.049	0.625	3361

[a]Discontinuations due to IUD expulsion, switching to another method, side-effects

[b]Discontinuations to become pregnant or for other unspecified reasons (menopause treated as censored)

Surprisingly, the risk of failure increases with the woman's level of education. A possible explanation is the higher level of pill and condom use among educated women, as these methods have higher failure rates than the IUD. The relationship between the risk of other method-related discontinuations and education is in the expected direction: educated women are less likely to stop use because of expulsion, side-effects or to switch to another method. There is little evidence of an education effect on nonmethod-related discontinuations among those with up to secondary education, but women who have been educated to senior level or above are almost twice as likely to stop for these reasons than less educated women.

Since the one-child policy is more lenient toward ethnic minority couples, one would expect discontinuation rates for method-related reasons to be higher for minority women as they would have less motivation for avoiding an unintentional pregnancy. However, the risks of failure and other method-related reasons are in fact lower among ethnic minority women than among those of the majority Han nationality. This again may be a reflection of differences in method mix between Han and minority couples: minorities are unlikely to use any modern method other than the IUD which has the lowest failure rate among modern reversible methods.

The risk of failure is higher in Beijing and Shandong than in the rural provinces, Guizhou and Gansu. This could be due to the higher level of pill and condom use in Beijing and the more developed rural province of Shandong (see Table 6.1). Beijing, a mainly urban province, has the lowest rate of other method-related discontinuation and the highest rate of nonmethod-related discontinuation. Similar patterns are observed among urban women in all four provinces.

The risk of each type of discontinuation varies according to the method used. The chance of stopping for any reason is considerably lower among IUD users than among users of any other method. In particular, IUD users experience almost half the failure rate of users of the pill or condom or other, mainly traditional, methods. There are few differences between methods for other method-related reasons, although users of the pill or condom have a slightly higher risk of stopping for these reasons, mainly due to the side-effects commonly associated with the pill. The likelihood of stopping for a nonmethod-related reason is much lower for IUD users than for users of other methods. This suggests that couples who wish to stop childbearing and who are not sterilised are more likely to use the IUD rather than some other reversible method.

To test whether current contraceptive behaviour is influenced by past experience, the number of previous uses and failures were considered. If a woman has used a method before, she may be more experienced in using contraception and less likely to discontinue for method-related reasons. It is also possible, however, that multiple prior use is an indication of repeated method-related problems which have caused her to discontinue on more than one occasion. In this case, there is evidence that women with more experience of contraceptive use, who have had two or more previous uses, have lower method-related discontinuation rates than those who have had no or only one previous use. The effect of a previous contraceptive failure could potentially affect the risk of an interval ending in failure in one of two ways. A previous failure may indicate some difficulties with contraceptive use which may occur repeatedly, in which case the risk of failure is increased. Alternatively, a woman who has already experienced an unintentional pregnancy may be extra-careful in avoiding a repeat failure, which would lead to a decrease in the current risk of failure. In the present case, no effect is found on the failure rate, but the risk of another type of method-related discontinuation is lower among women who have experienced a failure. The negative effect of prior failure on the risk of a nonmethod-related discontinuation is expected. If a woman has experienced a failure and did not have an abortion, she will have at least one child and is, therefore, less likely to stop use to have another.

6.6.3 Repeated Discontinuation

Table 6.3 shows the distribution of the number of use intervals contributed to the analysis sample by each woman. Twenty-eight percent of women have contributed more than one spell, accounting for nearly half of all intervals in the sample. This suggests that the durations of intervals will not be independent if a correlation exists between the durations of spells for the same woman; in such a case, a single-level competing risks model would not be appropriate. For example, one would not expect a woman who reports a history of difficulty with use and who contributed a series of short intervals to have the same risk of discontinuation as a woman who has not previously experienced problems, even if their demographic and socioeconomic characteristics are similar. Even if variables relating to a woman's previous contraceptive history, such as the number of prior uses and failures, are included, a single-level analysis may still not capture the full extent of this correlation between durations of successive spells for a woman.

Table 6.3: Distribution of the number of use intervals contributed to the analysis sample by each woman

No. of intervals	No. of women	% women
1	1696	71.4
2	463	19.5
3	155	6.5
4	42	1.8
5+	19	0.8
Total	2375	100.0

To determine whether some women experience the same type of discontinuation more than once, the distributions of the number of use intervals ending because of failure, other method-related reasons and nonmethod-related reasons are considered (Table 6.4). Although only 5.1% of women have experienced failure more than once in the observation period, multiple failures account for 36.4% of all failures in the sample. Also, the 2.1% of women who have contributed two or more other method-related stops contribute 20.7% of these types of discontinuation in the sample. Since method-related problems tend to arise during the initial period of use, the durations of these repeated intervals may be correlated.

There is little clustering of nonmethod-related discontinuations for any women. Very few women in the sample have stopped use more than once for a nonmethod-related reason. This finding is expected because the principal nonmethod-related reasons is stopping to become pregnant. As a direct result of the one-child policy, this is likely to occur only once for urban women and not more than twice for rural women.

Table 6.4: Distribution of the number of use intervals contributed to the analysis sample by each woman, by reason for discontinuation

	% women by reason for discontinuation		
No. of intervals	Failure	Other method-related reason	Nonmethod-related reason
0	74.6	80.5	89.3
1	20.3	17.4	10.4
2	4.0	2.0	0.2
3+	1.1	0.1	0.0
Total	100.0	100.0	100.0
No. of women	2375	2375	2375

6.7 Results of the Multilevel Discrete-time Competing Risks Event History Model for Discontinuation

6.7.1 Specification of Duration Dependency and Model Selection

The first step of the competing risks modeling was to specify the form of the relationship between the hazard of discontinuation and the duration of use. A number of alternative formulations for duration dependency, represented by α_{rt} in (6.1) and (6.2), were tested. These were 1) $\alpha_{rt} = \alpha_{r0} + \alpha_{r1}t$ (linear); 2) $\alpha_{rt} = \alpha_{r0} + \alpha_{r1}t + \alpha_{r2}t^2$ (quadratic); 3) $\alpha_{rt} = \alpha_{r0} + \alpha_{r1}\log t$ (linear in log time); 4) a piecewise-constant model, with durations grouped into 11 categories; and 5) a piecewise-constant model, with durations grouped into six categories. A single-level discrete-time competing risks model was fitted for each of these five specifications. At this stage, no other covariates were included in the models. Also, the interval between discrete time points was one month, that is each woman's contraceptive use status was recorded at monthly intervals. Since only models 1) and 2) are nested, standard tests for comparing models such as the likelihood ratio test or the Pearson chi-squared test are inappropriate. Therefore, the Akaike Information Criterion (AIC, Maddala 1988) for comparing nonnested models was used. For each model, the following statistic was calculated

$$\text{AIC} = -2(\text{maximised log-likelihood of the fitted model}) + 2p,$$

where p is the number of parameters in the fitted model. The model which minimises the AIC statistic among those considered is selected. While a low value for -2 log-likelihood is desirable, one would also like to limit the number of parameters in the model. The AIC takes into account both the precision and parsimony of the model; a model with a low -2 log-likelihood value is penalised if it has a large number of parameters.

Table 6.5 displays the values of -2 log-likelihood, p and the AIC statistic for each of the five duration formulations. Two of the models have very similar AIC values: the quadratic model and the piecewise-constant model where the duration is grouped into 11 categories (six month intervals for the first four years and one category for durations of four years or more). The quadratic model is selected since it has the lower AIC value and considerably fewer parameters.

Table 6.5: Modelling the dependency of discontinuation risks on duration of use

α_{rt}	-2 log-likelihood[a]	No. of parameters	AIC[b]
$\alpha_{r0} + \alpha_{r1}t$	18265.8	6	18277.8
$\alpha_{r0} + \alpha_{r1}t + \alpha_{r2}t^2$	18249.2	9	18267.2
$\alpha_{r0} + \alpha_{r1}\log t$	18377.2	6	18389.2
Piecewise-constant (11 categories[c])	18204.7	33	18270.7
Piecewise-constant (6 categories[d])	18243.2	18	18279.2

[a]Calculated for models including only α_{rt}, with data in 1 month intervals

[b]The Akaike Information Criterion

[c]0-2, 3-5, 6-8, 9-11, 12-17, 18-23, 24-29, 30-35, 36-41, 42-47, 48+ months

[d]0-5, 6-11, 12-23, 24-35, 36-47, 48+ months

After specifying the duration dependence, forward selection was used to determine which of the covariates in Table 6.2 had a significant net effect on the risk of at least one type of discontinuation. Single-level models were used in this initial screening process. Duration of use, method, the woman's age at the start of use, family composition, the period of starting use, education, the type of region of residence and her previous experience with contraceptive use were all found to have a significant effect on the risk of one or more types of discontinuation. All two-way interactions between these variables were also tested. In particular, interactions with method were considered to test whether the effects of covariates varied across method. All interactions with duration were considered to test for non-proportionality of hazards. However, no two-way interactions were found to be significant.

Before extending the selected single-level model to the multilevel model, attempts were made to reduce the size of the data set by increasing the interval between discrete time points. The current model, using one month intervals, requires the data set to be expanded to give a multinomial response for each month of contraceptive use. This procedure generates a huge data set with over 90,000 records. Fitting a multilevel multinomial model to such a large data set would be extremely computationally intensive. The data would need to be further extended to obtain a set of three binary responses for each month (see Section 2.8.2) which would generate over 270,000 records. In addition, the multinomial model has a large number of parameters since a separate set of parameters is estimated for each contrast of a type of discontinuation with the reference 'still using' response category. Therefore, the selected model was refitted using broader discrete-time intervals of three and six months. It was found that increasing the interval from one month to six months actually had little effect on the parameter estimates and standard errors. Most of the parameter estimates only differed in the second decimal place. Since the increase to six-month intervals reduced the data set from over 90,000 records to a more manageable 16 705 and the computational time was also reduced considerably, the model using six-month intervals was selected. As a final check, the five alternative specifications for the duration dependency (Table 6.5) were retested. Once again, the quadratic model was found to provide the best fit to the data.

In the final stage of the analysis, the selected competing risks model was extended to a two-level model which allows for unobserved heterogeneity in discontinuation risks between women. A further level corresponding to use interval was added, but no evidence of extravariation at this level was found. This implies that the unobserved woman-level factors affecting the risk of discontinuation are constant across repeated intervals. However, it is also possible that too few women experienced more than one discontinuation of a particular type during the observation period to enable use interval and woman effects to be distinguished. Therefore, the two-level model was selected. A further extension of the basic random intercepts model was considered in which the coefficients of covariates were allowed to vary randomly across women. However, no evidence of variability in the slope coefficients was found.

Table 6.6 displays the parameter estimates and standard errors for the selected two-level random intercepts model. The effects and significance of the covariates selected using the single-level model do not change markedly after the addition of the woman-specific random effects. However, the results show that a significant amount of the variation in discontinuation risks between women remains unexplained by the covariates included in the model: there is significant extravariation in the risk of failure and of other method-related discontinuations, but not in the risk of nonmethod-related discontinuations.

6.7.2 Interpretation of the Fixed Effects

To aid interpretation, the parameter estimates presented in Table 6.6 have been converted to estimated 24-month cumulative probabilities of discontinuation (Table 6.7). These probabilities have been estimated for each covariate using formulae presented in Section 4.6.2, while all other covariates in the model are held at average values. At this stage, the random effects are also fixed at their average values of zero.

The duration of use has a significant effect on the risk of failure and nonmethod-related discontinuation. As expected, the risk of failure increases initially and then decreases as the duration of the use interval increases. This would suggest that the longer a woman manages to use a method successfully, the more experienced she becomes and the less likely she is to fail. There is a downward, though not statistically significant, trend in

Table 6.6: Estimates (and standard errors) from the multilevel discrete-time competing risks model for discontinuation

Variable	Type of discontinuation					
	Failure/ Continue		Other method-related/ Continue		Nonmethod-related/ Continue	
	Est.	(S.E.)	Est.	(S.E.)	Est.	(S.E.)
Constant	-3.54	(0.23)	-2.65	(0.25)	-6.74	(0.42)
Duration	0.10**	(0.05)	-0.08	(0.05)	0.22***	(0.06)
Duration^2	-0.02***	(0.004)	-0.004	(0.004)	-0.009**	(0.004)
Age at start of use (Base=15-24)						
25-29	-0.16*	(0.10)	-0.18	(0.12)	-0.21	(0.16)
30-34	-0.61***	(0.14)	-0.59***	(0.16)	-1.18***	(0.25)
35+	-1.67***	(0.24)	-1.37***	(0.22)	-0.88***	(0.26)
No. and sex of children (Base=1 boy)						
0	1.34***	(0.17)	-0.06	(0.33)	3.22***	(0.27)
1 girl	0.09	(0.11)	-0.09	(0.14)	0.73***	(0.25)
2 girls	0.26	(0.20)	0.38	(0.23)	2.52***	(0.27)
2 (at least 1 boy) or 3+	0.27**	(0.12)	0.88***	(0.13)	1.78***	(0.24)
Year of starting use (Base=<1984)						
$\geq$ 1984	-0.41***	(0.08)	-0.45***	(0.10)	-0.04	(0.15)
Education (Base=None)						
Primary+	0.08	(0.12)	-0.46***	(0.12)	-0.08	(0.18)
Region of residence (Base=rural)						
Urban	-0.40***	(0.10)	-0.13	(0.12)	-0.47***	(0.17)
Method (Base=IUD)						
Pill/condom	0.97***	(0.09)	0.99***	(0.12)	1.42***	(0.17)
Other	1.08***	(0.16)	0.94***	(0.21)	1.18***	(0.28)
Contraceptive experience (Base=prior use, but no failure)						
No prior use	0.29**	(0.13)	-0.22	(0.13)	0.48**	(0.22)
Prior use and failure	0.22	(0.14)	-0.50***	(0.15)	-0.15	(0.25)
Random effect variance σ^2_{ur}	0.24***	(0.09)	0.25**	(0.12)	0.19	(0.18)

* p=0.10; ** p=0.05; *** p=0.01

Table 6.7: Estimated 24-month cumulative probabilities of discontinuation by selected background characteristics

Variable	Reason for discontinuation Failure	Other Method-related[a]	Nonmethod-related[b]	Continue
Age at start (years)				
15-24	0.160	0.137	0.029	0.673
25-29	0.142	0.118	0.025	0.715
30-34	0.097	0.084	0.010	0.808
35+	0.036	0.041	0.015	0.907
No. and sex of children				
None	0.308	0.061	0.140	0.492
1 boy	0.106	0.084	0.008	0.802
1 girl	0.116	0.076	0.016	0.793
2 girls	0.125	0.112	0.085	0.678
2 ($\geq$ 1 boy), 3+	0.123	0.180	0.039	0.657
Year of starting use				
$<$ 1984	0.124	0.107	0.021	0.748
$\geq$ 1984	0.087	0.072	0.022	0.819
Education				
None	0.113	0.148	0.022	0.716
Primary+	0.127	0.097	0.021	0.755
Region				
Urban	0.100	0.102	0.017	0.782
Rural	0.142	0.111	0.025	0.722
Method				
IUD	0.090	0.078	0.013	0.819
Pill/condom	0.194	0.174	0.045	0.588
Other	0.216	0.165	0.035	0.584
Contraceptive experience				
No prior use	0.129	0.111	0.026	0.733
Prior use and failure	0.125	0.087	0.015	0.774
Prior use, but no failure	0.098	0.140	0.017	0.746
Overall average	0.124	0.107	0.021	0.748

[a]Stopping due to IUD expulsion, switching to another method, side-effects

[b]Stopping to become pregnant or for other unspecified reasons (menopause treated as censored)

the risk of a method-related discontinuation. The risk of stopping for a nonmethod-related reason increases with duration of use, but levels out at longer durations. Again, this is as expected because a pregnancy would usually be planned in advance. Thus a couple is unlikely to change to a new method, only to stop within a few months to have a child.

The woman's age at initiation of use remains an important factor in determining the risk of discontinuation after controlling for other characteristics such as family composition. For the method-related reasons—failure, expulsion, switching methods and side-effects—older women's lower risk of stopping could be due to their greater experience. Also, women aged 30 or more are considerably less likely to discontinue to have a child or for other nonmethod-related reasons.

The number and sex composition of a woman's children has a significant effect on discontinuation risks. Although the original variable tested had seven categories (see Table 6.2), initial models showed no significant difference in the risk of any type of discontinuation between women with two children (including at least one boy) and those with three or more children. No evidence of a sex bias in the chance of failure was found: all women who have at least one child experience similar risks. However, the risk of failure and other method-related discontinuation is significantly higher for families with two or more children. The higher failure rate among those with more than one child may reflect a tendency to report an intended pregancy as a failure to avoid being penalised for having another child. If a woman has one child, she is more likely to stop to become pregnant if that child is a girl than if it is a boy. Women who already have a child tend to be more strongly motivated to avoid a further pregnancy, and so less likely to fail, than women without children.

The time period during which use of a method was initiated has a strong impact on the risk of method-related discontinuation. Women starting use during the pre-1984 period, when the family planning policy was strictly enforced, are more likely to discontinue for method-related reasons than those who started after 1984. One might speculate that those women who did not wish to use contraception, but who were under great pressure from the family planning programme before 1984, would have cited method-related reasons to explain why they discontinued. In contrast, after 1984, when the programme became more flexible, women would have been less likely to feel the need to cite method-related reasons for discontinuation.

After controlling for method and other characteristics, the negative effect of education on failure risks observed in the life table analysis (Table 6.2) loses its statistical significance. However, there is evidence that better-educated women are less likely than the uneducated to stop because of expulsion, method switches, or side-effects. Also, women living in rural areas are more likely to stop to have a child, probably because rural couples are permitted to have two children without penalty. Also, since urban areas in China are characterised by high population density and high cost of living, it is probable that couples living in these conditions are more easily persuaded of the advantages of having one child than rural women. When the type of region of residence and the level of education are controlled, differences between Han and ethnic minorities are no longer significant. Similarly, the province of residence becomes unimportant, although there was some evidence of higher discontinuation rates in Guizhou and Gansu, the poor rural provinces, before the inclusion of type of residence.

Users of IUDs have the lowest rate of any type of discontinuation. Failures, method switches, and side-effects are more common among users of the pill or condom, or among those using other methods, mainly inefficient methods such as rhythm or withdrawal. IUD users are the least likely to discontinue use in order to become pregnant. This suggests that the IUD is regarded as a long-term method, an alternative to sterilisation, whereas other methods are more likely to be used as temporary measures among couples who plan to have another child.

To explore the influence of previous contraceptive behaviour on the chance of discontinuation of the current method, the number of previous uses and failures were considered (see Table 6.2). However, since the majority of women have had no prior use and thus cannot have experienced a failure, these two variables were amalgamated. The constructed variable has three categories: 1) no prior use, 2) at least one prior use and failure, and 3) at least one prior use, but no failure. An interesting result is that the increased risk of failure observed in the corresponding single-level model among women who have already experienced failure is not significant when the model is extended to the multilevel version. This suggests that the occurrence of a previous failure may be a proxy for unobserved woman-level characteristics which increase her susceptibility to repeated failure. The multilevel model includes a woman-specific random effect to account for unobserved factors such as these. In this model, it is women who have never used contraception before who have the highest risk of failure. However, a previous failure still has a negative effect on the risk of other method-related discontinuations. Women who have had no previous use are more likely to stop for nonmethod-related reasons than women who have had some prior experience with contraception.

6.7.3 Interpretation of the Random Effects

Even after controlling for a range of demographic and socioeconomic variables and characteristics of previous contraceptive use experience, there is a significant amount of variation in method-related discontinuation risks between women that remains unexplained. However, there is no significant extravariation in the risk of nonmethod-related discontinuations. The significance of the random effect variances for failure and other method-related discontinuation suggest that there remain some unobserved variables that increase or decrease a woman's susceptibility to the risk of stopping use for method-related reasons. These unobservables may include biological factors. For example, some women may be more likely to experience side-effects or IUD expulsion due to biological reasons. Fecundity is another example of an unobserved biological factor; some women have an inherently higher chance of conceiving and, therefore, have a higher risk of failure, regardless of which method they are using. Another explanation for the unobserved heterogeneity in the risk of stopping for method-related reasons could be characteristics of prior use. Although the number of prior uses and failures have been considered, it is unlikely that these fully capture a woman's history of contraceptive use and the problems that she has encountered which could affect her current risk of discontinuation. Further possibilities are factors relating to the availability and quality of family planning services. For example, Wang and Diamond (1995) suggest that much of the IUD failure rate can be attributed to the low quality of IUDS used in China. It is possible that IUD quality and poor insertion techniques leads to an increase in failure, side-effects and expulsion risks.

While biological and family planning supply factors are possible explanations for the extravariability in method-related discontinuation risks, however, it is unlikely that they will affect the risk of nonmethod-related discontinuation. The decision to stop use in order to become pregnant is more likely to be a function of factors such as family composition and regional and ethnic differentials in family planning policy, all of which have been considered in the model. This could explain the lack of significance of the random effect variance for nonmethod-related reasons for discontinuation.

To illustrate the impact of this extravariation on method-related discontinuation rates, 24-month cumulative probabilities of stopping due to failure and other method-related reasons have been calculated for each method type (Tables 6.8 and 6.9). These have been calculated with the observed covariates held at their sample means. Following the approach of Curtis et al. (1993), the random effect corresponding to failure is varied between two standard deviations below the mean value of zero and two standard deviations above the mean, representing an approximate 95% confidence interval for women effects (Table 6.8). The random effects for other method-related and nonmethod-related discontinuations are fixed at zero. Similarly, the random effects corresponding to failure and nonmethod-related stops are held at mean values of zero in order to examine the impact of the other method-related random effect on other method-related discontinuations (Table 6.9). In both cases, the probability of method-related discontinuation varies considerably according to the value of the random effect. This extravariability across women leads to overlaps in the approximate 95% confidence intervals for the three method types. For instance, if all else is equal, the risk of failure for the IUD is less than half that associated with the pill, condom and other methods. It is possible, however, that a woman who is fitted with an IUD, but who is in a higher risk group in terms of her unobserved characteristics (with a random effect value greater than $\hat{\sigma}_{u1}$) has a greater chance of failure than a similar woman who uses the pill or condom, but who is in a lower risk group (with a random effect value below -$\hat{\sigma}_{u1}$).

Table 6.8: Estimated 24-month cumulative probabilities of failure by method, for a range of values of the woman-level random effect u_{1j}

	Random effect u_{1j}				
Method	$-2\hat{\sigma}_{u1}$	$-\hat{\sigma}_{u1}$	0	$\hat{\sigma}_{u1}$	$2\hat{\sigma}_{u1}$
IUD	0.040	0.062	0.093	0.140	0.206
Pill/condom	0.098	0.147	0.215	0.309	0.427
Other	0.109	0.163	0.238	0.338	0.462

Table 6.9: Estimated 24-month cumulative probabilities of another method-related discontinuation by method, for a range of values of the woman-level random effect u_{2j}

	Random effect u_{2j}				
Method	$-2\hat{\sigma}_{u2}$	$-\hat{\sigma}_{u2}$	0	$\hat{\sigma}_{u2}$	$2\hat{\sigma}_{u2}$
IUD	0.035	0.053	0.081	0.122	0.180
Pill/condom	0.086	0.130	0.191	0.277	0.389
Other	0.082	0.123	0.183	0.265	0.372

6.8 Summary

In this chapter, a multilevel discrete-time competing risks model was used to study the factors affecting contraceptive discontinuation in four provinces of China. A competing risks framework allows one to distinguish between the risk factors associated with discontinuations due to failure, other method-related reasons and nonmethod-related reasons, while a multilevel model allows for omitted covariates at the woman level. The model incorporates women-specific random effects corresponding to each type of discontinuation. These allow for unobserved heterogeneity between women, and correlations between durations of repeated use intervals for the same woman.

The substantive results are much as expected: discontinuation rates are lowest among IUD users; older women are less likely to stop use; women with no children have an increased chance of discontinuing because of failure or because they wish to become pregnant; method-related reasons become less prevalent after the relaxation in policy in 1984; and rural dwellers are more likely than their urban counterparts to discontinue. Of primary interest are the random effects. These can be regarded as representing the woman's unobserved characteristics. Some of these may be observable—for example, the type of IUD used—but others may be unobservable such as fecundity. In any case, there is a substantial amount of extravariation between women in discontinuation rates for failure and for other method-related reasons. The estimated probabilities calculated for a range of values of each random effect show that these unobserved variables have a greater impact on the risk of method-related discontinuation than the observed covariates.

Chapter 7

Contraceptive Switching in Bangladesh

7.1 Introduction

In Chapter 6, multilevel event history analysis techniques were used to study contraceptive discontinuation. In this chapter, we look at another important aspect of use dynamics—contraceptive switching. Most work on use dynamics has focused on the process of discontinuation and the reasons for stopping use rather than the method chosen after discontinuation. The study of switching is concerned with contraceptive behaviour following discontinuation. Questions of interest include: Does the woman switch to an alternative method or does she abandon use altogether? If she switches to another method, which method does she choose? Contraceptive discontinuation and switching are closely related since the impact of discontinuation on fertility levels depends on switching behaviour. Of particular concern are switches from modern methods to less effective methods or non-use because of method-related reasons such as side-effects, health concerns or supply problems. A high degree of switching from modern methods to inefficient traditional methods or non-use would clearly have a large impact on fertility rates due to an increase in the number of unwanted pregnancies. Switches of this type also suggest that the family planning services have not been successful in meeting these women's needs: it could be the case that they are not offering a suitable range of method options or that they are failing to provide adequate information on method use. If a woman immediately switches to a new equally effective method, the implications for fertility levels are obviously less serious. Such switches may in fact be the result of an increase in the range of contraceptive options made available by family planning services which enable women to change to the method which best suits their individual needs. However, any switching between methods can potentially lead to an increased risk of unwanted pregnancy since failure rates are typically higher during the first few months of use when a couple may not be fully familiar with a new method.

The analysis presented in this chapter employs multilevel discrete-time competing risks models to study the determinants of contraceptive switching over a six-year period, using calendar data from the 1993-94 Bangladesh Demographic and Health Survey. A competing risks framework is used to distinguish between the various types of switches that may occur. For example, a pill user may abandon use, switch to another modern method, switch to a traditional method or continue using the pill. Of major interest is the extent of unobserved variation in switching rates at the woman, sampling cluster and district level. Therefore, random effects corresponding to each level of the hierarchy and for each type of switch are incorporated in the model.

7.2 Approaches Used to Study Contraceptive Switching

Very few studies have looked at switches between contraceptive methods, and those that have each use a different analysis approach. The approach chosen depends, to a large extent, on the data available: for example, the length of the period for which contraceptive use histories were collected and the type of information recorded. Some surveys collect retrospective data on all contraceptive use since marriage, while others only collect data for a three- to six-year window before the survey. In some cases, information on the exact duration of method use is unavailable, and only data on the type of methods used between pregnancies are collected. This automatically precludes a hazards modelling approach. DaVanzo et al. (1989) utilise data of this type by looking at method changes between the pair of intervals before and after a pregnancy.

Their analysis is descriptive with transition matrices showing the probabilities of moving between methods, including non-use, and continuing use of the same method after pregnancy. They also look at bivariate relationships between switching and background characteristics such as the woman's education level, age and pregnancy order.

Curtis and Hammerslough (1995) propose an approach which uses multiple-decrement life tables to calculate cumulative 12-month discontinuation rates by use status in the month immediately after discontinuation. Separate life tables are calculated for each type of origin method and five types of decrement are considered: 1) non-use because the woman has no need for contraception (discontinuations because the woman is pregnant, wants to have a child, or as a result of marital separation or menopause), 2) non-use while at risk of an unintended pregnancy, 3) a switch to a modern method, 4) a switch to a traditional method, and 5) continuation with the same method. This approach was adopted by Mitra and Sabir (1996) to study contraceptive switching in Bangladesh, using calendar data from the Bangladesh Demographic and Health Survey.

Several studies have employed regression techniques to analyse contraceptive switching. Samosir (1994) extends the bivariate approach of DaVanzo et al. (1989) by using a multinomial logit regression model to look at switches between successive pairs of inter-pregnancy intervals. The response variable is the destination method, the method used after pregnancy. To allow the estimation of method-specific transition probabilities, the origin method, the method used before pregnancy, is included as a covariate. Hamill et al. (1990) extend their descriptive analysis of switching in rural Sri Lanka (Tsui et al. 1989) to a multivariate analysis, again using multinomial logit regression. They only have information on contraceptive use for three years before the survey and the first interval for each woman is always left-censored. Since eliminating these intervals would mean the exclusion of a large number of cases, they reject a hazards approach. Instead of modelling the duration of use until the occurrence of a switch explicitly, they fix two points in time, two years apart, and analyse the probability of switching from one method to another between these two time points. Very few women experience more than one switch during the two-year period. If at either time point the woman is pregnant or amenorrheic, her status is classified as the method used most recently when she was at risk of conception. Separate models are fitted for each origin method, the method used at the start of the two-year period, where the destination method used at the end of the period is the multinomial response variable.

Grady et al. (1989) use a competing risks hazards framework in their analysis of switching in the United States. They model the time since method adoption to the first switch, either to non-use or to an alternative method. If contraceptive use is interrupted by a period of no intercourse, the use interval is treated as continuous, but the period of non-use does not contribute to the duration of the interval. Intervals of non-use are only considered if the woman is at risk of an unintended pregnancy, that is, if she is non-sterile, having intercourse and not trying to become pregnant. A series of models are fitted, one for each type of destination method, and in each model continuations with the same method or switches to a method other than the type of interest are treated as censored. The origin method is included as a covariate and the fitted model is used to estimate cumulative two-year transition probabilities.

Islam (1994) also uses a hazards modelling approach to analyse changes in contraceptive behaviour in Matlab, Bangladesh. He describes a multistate model where states are either reversible or absorbing. Intervals of non-use or temporary method use are classified as reversible, while use of an irreversible method is an absorbing state. For reversible states, repeated transitions are possible. For example, if an individual moves from non-use to use of a temporary method, discontinues use and then adopts another temporary method, this is counted as two transitions of the type non-use to temporary use. The use of an irreversible method is absorbing because no further transitions are possible after entry into this state. Islam fits a series of models for each type of transition, with separate models for repeated transitions of the same type. An approach for testing the equality of regression parameters for two models is also developed. In particular, this allows one to test whether the underlying factors affecting a transition act in the same way for repeated transitions of the same type. However, due to a lack of data on pregnancy outcomes and reasons for discontinuation, Islam's analysis was unable to distinguish between non-use intervals when a woman is not at risk of an unintended pregnancy (for example, pregnant or trying to become pregnant) and other types of non-use.

Samosir (1994) uses a multilevel approach to allow for potential correlation between outcomes for the same woman. Correlations are likely to arise if a woman makes repeated transitions over the observation period. For example, a woman who experiences problems with contraception may have repeated discontinuations and therefore would contribute a series of switches from use to non-use. In the two-level model, use intervals (level 1) are nested within women (level 2) and a woman-level random effect is included which controls for the possibility of clustering of outcomes for intervals contributed by the same woman.

7.3 Methodology

The present analysis employs multilevel discrete-time competing risks models to examine switching behaviour among contraceptive users in Bangladesh. Three separate models are used to study switches from use of the pill, other modern methods and traditional methods. The pill is considered separately from other modern methods since it is by far the most commonly used method in Bangladesh, accounting for 64% of modern reversible use in 1993-94 (Mitra et al. 1994). Following the life table approach of Curtis and Hammerslough (1995) and Mitra and Sabir (1996), the type of switch which occurs is determined by the woman's use status in the month immediately after method discontinuation. A woman may be in one of four states at any time after initiation of a method: she may have 1) switched to non-use, 2) switched to a modern method, 3) switched to a traditional method, or 4) continued to use the original method. In addition, data on the reason for discontinuation are utilised to distinguish between two broad categories of non-use: non-use because the woman has no need for contraception at that time, and non-use where the woman does have a need for contraceptive protection and is at risk of an unintended pregnancy. The first group of non-users consists of women who have discontinued because they want to get pregnant, or are pregnant as a result of a contraceptive failure, those who are having no or infrequent intercourse because they have experienced a marital separation or their husband is away, and women who have reached menopause or who consider themselves infertile. If a woman stops use because of menopause, infrequent sex or a marital separation, the interval is treated as censored, that is, we assume that had she a need for contraception then she would have continued use of the same method. Intervals which end in a failure or discontinuation in order to become pregnant are considered as a separate response category.

Of most interest are transitions to non-use when the woman is at risk of an unintended pregnancy, but does not immediately switch to another method after stopping use. Such transitions are the result of discontinuations for method-related reasons and indicate dissatisfaction not only with the current method but with the other available methods. This type of switch follows discontinuations because of side-effects or health concerns, inconvenience with use, husband's disapproval, the need for a more effective method and supply or access problems with the current method. In the present analysis, these are examined in detail by including a separate response category for transitions to non-use for method-related reasons. This category also contains a group of women who became pregnant immediately after discontinuation, but not as a consequence of a method failure or because they wanted to become pregnant. These women are classified as non-users since before becoming pregnant they must have first gone through a short period of non-use.

Each transition from the pill or another modern method is classified as one of five types, with a separate response category for each: 1) non-use where the woman has no need for contraception (failure or wants to become pregnant), 2) non-use and at risk of an unintended pregnancy, 3) an alternative modern method, 4) traditional method, or 5) continue use of same modern method. Traditional method users can also make transitions of type 1), 2) and 3), but since they cannot switch to the same method, the final category is now 4) continue traditional use. Although contraceptive failures are a special and important case, they are not considered separately here since the focus of this study is the behaviour of users who discontinue use while still in need of contraceptive protection. A comprehensive analysis of contraceptive use dynamics in Bangladesh would also examine the factors affecting the risk of failure and other types of discontinuation. An analytic approach similar to that used in Chapter 6 could be employed.

The discrete-time competing risks model described in Section 4.6.2 can be used to analyse these data. Suppose there are s competing alternatives for a method user (including continuing use), then the model can be written

$$\log\left(\frac{h_{rti}}{h_{sti}}\right) = \alpha_{rt} + \mathbf{x}'_{rti}\beta_r, \ r = 1, \ldots, s-1 \tag{7.1}$$

where h_{rti} is the hazard of a transition of type r at time t for use interval i; h_{sti} is the hazard of a transition of type s (the reference category) which in this case is the hazard of continuing use of the same method. A use interval is observed from the time the method was adopted until a transition occurs or until the time of the survey (right-censored cases). Duration effects are incorporated in the analysis by inclusion of α_{rt} which is some function of the time since starting use of the origin method. $\mathbf{x}_{rti}$ is a vector of possibly time-dependent covariates and β_r is the vector of associated regression parameters.

This approach is similar to that of Grady et al. (1989). Rather than fitting a separate model for each destination state, however, the multinomial formulation of the model allows transition probabilities to be estimated simultaneously. In the model for transitions from modern methods other than the pill, the origin method is included as a covariate to enable the calculation of method-specific transition rates.

The model in (7.1) assumes that all variation in women's risks of switching can be explained by the covariates $\mathbf{x}$. In most situations, however, this assumption is invalid since there will probably be important risk factors which have been omitted, some of which may be unobservable. To allow for unobserved heterogeneity, random effects defined at the woman level can be included in the model. It is also possible that there will be unobserved factors operating at two additional levels: the primary sampling unit (PSU) and the district. The multilevel analysis of contraceptive choice in Bangladesh in Chapter 3 revealed a large amount of extravariation both between districts, and within districts between PSUs. An aim of the present analysis is to examine whether the patterns of clustering in contraceptive choice observed in Chapter 3 extend to contraceptive behaviour over time. For example, it was suggested that the variation at the PSU level may be attributed to characteristics of the family planning services which are implemented at that level. One would expect such factors as quality of services and the range of methods available in the locality to continue to play a role among method acceptors. To test for the presence of unobserved covariates acting at the woman, PSU and district level, the following four-level model is used

$$\log\left(\frac{h_{rtijkl}}{h_{stijkl}}\right) = \alpha_{rt} + \mathbf{x}'_{rtijkl}\beta_r + \mathbf{z}'_{urtijkl}u_{rjkl} + \mathbf{z}'_{vrtijkl}v_{rkl} + \mathbf{z}'_{wrtijkl}w_{rl}, \; r = 1, \ldots, s-1 \tag{7.2}$$

This is a multilevel extension of the discrete-time competing risks hazards model given in (7.1), where h_{rtijkl} is the hazard of a transition of type r at time t for use interval i of woman j in PSU k in district l. The covariates in $\mathbf{x}$ can be defined at any level in the hierarchy, including the time interval which is the lowest level in the model. In the multilevel model, a set of random effects for each transition type r is included: $\mathbf{u}_{rjkl}$ are the woman effects, $\mathbf{v}_{rkl}$ the PSU effects and $\mathbf{w}_{rl}$ the district effects. As usual, these are assumed to follow normal distributions: $\mathbf{u}_{rjkl} \sim N(0, \Omega_{ur})$, $\mathbf{v}_{rkl} \sim N(0, \Omega_{vr})$ and $\mathbf{w}_{rl} \sim N(0, \Omega_{wr})$. In the general random coefficients model presented in (7.2), each of the random effects vectors $\mathbf{u}_r$, $\mathbf{v}_r$ and $\mathbf{w}_r$ has an associated vector of covariates, $\mathbf{z}_{ur}$, $\mathbf{z}_{vr}$ and $\mathbf{z}_{wr}$ respectively. These are usually subsets of $\mathbf{x}_r$, the effects of which vary across women, PSUs or districts.

If there are women who experience the same type of switch more than once over the six-year observation period, the model in (7.2) can be extended to a five-level model which incorporates an additional set of random effects corresponding to the use interval. This would enable one to test whether the unobserved woman-level factors represented by $\mathbf{u}_r$ are constant for each occurrence of a switch of type r.

7.4 The Bangladesh Demographic and Health Survey 1993-94

The data used for this analysis come from the Bangladesh Demographic and Health Survey (BDHS) conducted between November 1993 and March 1994 under the authority of the National Institute of Population Research and Training (NIPORT) of the Ministry of Health and Family Welfare. A two-stage sample design was used. In the first stage, 304 primary sampling units (PSU) were selected with probability proportional to size. PSUs correspond roughly to villages in rural areas and neighbourhoods in urban areas.

Since fieldwork was not possible in three of the areas, a total of 301 PSUs were covered - 231 rural and 70 urban. In the second stage, a list of households was drawn up for each PSU and used to obtain a systematic sample of households. The average number of households selected in each PSU was 25 in urban areas and 37 in rural areas. In the selected households, all ever-married women aged 10-49 were interviewed giving a total sample size of 9640. The BDHS used a series of four questionnaires to collect data on households, women, husbands and service availability. The service availability module collected information on family planning, health and other services from family planning workers and key informants in the community.

In the BDHS, retrospective data on contraceptive use were collected in the form of a calendar. Information on births, pregnancies, contraceptive use, reasons for discontinuation of use and marital status was recorded for each month for a period of six years, from April 1988 to March 1994 or the month of interview. Pregnancies and births were entered first in the calendar, starting with those most recent and then working backwards towards the start of the six-year window. The interviewer then asked the respondent which methods she used between pregnancies, if necessary using children's names and dates of birth as reference points to aid recall.

The calendar approach has proved a very useful and reliable means of collecting retrospective data. One important advantage of the calendar is that it makes it easier for the interviewer to detect inconsistencies in an individual's responses. For example, a woman who reports that she was pregnant in a particular month may also say she was using contraception. In questionnaires traditionally used in fertility surveys, pregnancy and contraceptive use histories are collected in different sections, making it difficult for reporting errors of this type to be detected during the course of the interview. The calendar technique has also been found to be more effective in helping a respondent to recall events as the interviewer is able to probe using important events such as pregnancies and births as reference points. Although the calendar approach has been shown to be superior to other techniques for the collection of retrospective longitudinal data (Goldman et al. 1989; Westoff et al. 1990 and Becker and Sosa 1992), it is not immune to data quality problems. In particular, all retrospective data are subject to recall error which can lead to omission or misreporting of the timing of events. One simple way of testing for recall error is to compare calendar data with current status data from an external source. Mitra and Sabir (1996) used this approach to check the quality of the BDHS calendar data. They compared method-specific contraceptive prevalence estimates for 1989 and 1991 calculated using the BDHS calendar with those obtained using current status data from two previous fertility surveys. They found that the rates from the two different sources were extremely close, indicating that the BDHS calendar data are of a high quality.

A total of 620 women or their husbands had been sterilised before the start of the observation period. These women were excluded from the analysis since for them no further changes in contraceptive use are possible. In addition, any use interval which ended in sterilisation was excluded. There were actually very few such transitions observed. The rate of sterilisation acceptance was considerably lower in the 1990s than in earlier years and of the women surveyed in 1993-94, only 248 were sterilised during the six-year observation period. This is because although the Bangladeshi government has never formally dropped the compensation scheme for sterilisation patients, funding from international aid agencies ended in the late 1980s, and, therefore, the programme of financial incentives was carried out with much less zest than in earlier years. Further, only 42% of these women had ever used any other method of contraception prior to sterilisation, and among those who had used another method previously, sterilisation was usually performed after a birth. A total of only 36 transitions from a reversible method to sterilisation were excluded. Any intervals contributed before the sterilisation, however, were retained in the analysis.

A large proportion of married women (35%) used no form of contraception during the entire six-year window and, therefore, were not considered in this analysis. Left-censored use intervals which began before April 1988 and continued into the observation period were also excluded since women were not asked the starting date of these spells and, therefore, their duration is not known. This led to the exclusion of 4.5% of users who continued to use the same reversible method throughout the observation period. The final sample consists of 3,826 intervals of pill use, 2,345 of other modern methods and 1,490 traditional use intervals.

7.5 Preliminary Analysis

7.5.1 Multiple-Decrement Life Table Switching Rates by Method

In the first stage of the analysis, cumulative probabilities of switching by each month in the six-year study period are estimated for each method of contraception. Multiple-decrement life tables (see Section 4.6.1) are used to calculate the hazard of each type of switch at every month in the observation period. These are then used to estimate cumulative probabilities of switching within two years of initiating method use.

Figures 7.1 to 7.5 show life table estimates of the cumulative probabilities of switching for each method. To summarise these results, 24-month cumulative probabilities of switching are also presented for each method (Table 7.1). The main focus of interest is switches among women who still have a need for contraception, that is, switches to alternative methods or to non-use when the woman is at risk of an unwanted pregnancy. Pill users who discontinue use are more likely to abandon use completely rather than switch to another modern method, and are very unlikely to move to a traditional method. Given the popularity of the pill in Bangladesh, the proportion discontinuing use and not adopting another method is of particular concern and is likely to have a large impact on fertility levels. It is important to determine the reasons for discontinuation among these women and whether their decision to abandon use of contraception is as a result of their experience with the pill or because of a lack of a suitable alternative to which they can switch.

Users of injectables, IUDs and condoms are all much more likely to switch to alternative modern methods than pill users. This is because the pill is the most commonly used modern method and, therefore, switches from injectables, IUDs or condoms to another method are nearly always to the pill rather than any other type of modern method. The proportion of users of modern methods other than the pill switching to non-use while at risk of unintended pregnancy is much higher for those using injectables than for IUD or condom users.

In general, users of any modern method are unlikely to switch to a traditional method, particularly users of injectables or the IUD, and the transition probability remains fairly constant over the six-year period. However, the rate of moving to a traditional method is considerably higher among condom users than users of any other modern method. Users of traditional methods are the least likely to switch to a modern method.

The 'not at risk' category consists mainly of women who have discontinued use in order to become pregnant, although this group also contains women who have become pregnant as the result of contraceptive failure. Among those who have switched to 'non-use and not at risk', 28% have experienced a contraceptive failure (16% for pill users, 26% for users of modern methods other than the pill, and 45% for traditional users). The probability of switching to this category is highest for traditional users, a reflection of the fact that these methods are more likely to be used by couples with a low commitment to family planning who wish to have more children and also of their higher failure rates. Women using injectables or IUDs are very unlikely to discontinue use for these reasons which would suggest that these methods are perceived as more long-term and are thus tend to be used by couples who wish to limit rather than space their births. Spacers appear to be more likely to use traditional methods, condoms and, to a lesser extent, the pill.

Table 7.1: 24-month cumulative probabilities of switching by origin method

	Type of Transition					
Method	Non-use and not at risk [a]	Non-use and at risk	Another modern	Traditional	Continue	n
Pill	0.141	0.234	0.154	0.060	0.411	3843
Injectables	0.074	0.301	0.282	0.042	0.301	927
IUD	0.042	0.196	0.265	0.039	0.458	443
Condom	0.248	0.158	0.302	0.109	0.183	983
Traditional	0.329	0.088	0.208	-	0.375	1504

[a] Discontinuations because a woman wants to become pregnant, or is pregnant after contraceptive failure. Note that women who stop because of menopause, infrequent sex or marital separation are treated as censored.

Figure 7.1: Cumulative probabilities of switching from the pill use

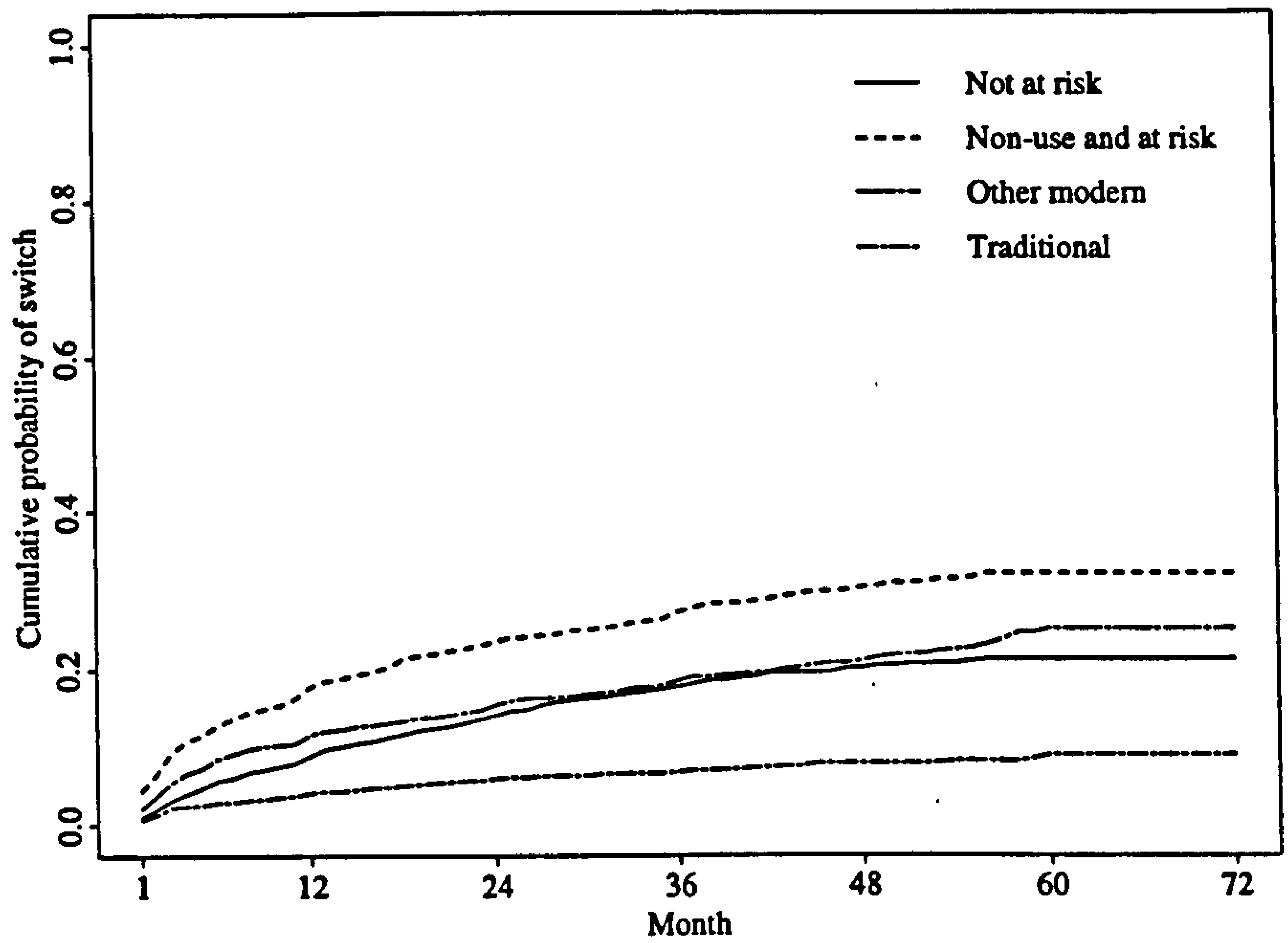


Figure 7.2: Cumulative probabilities of switching from use of injectables

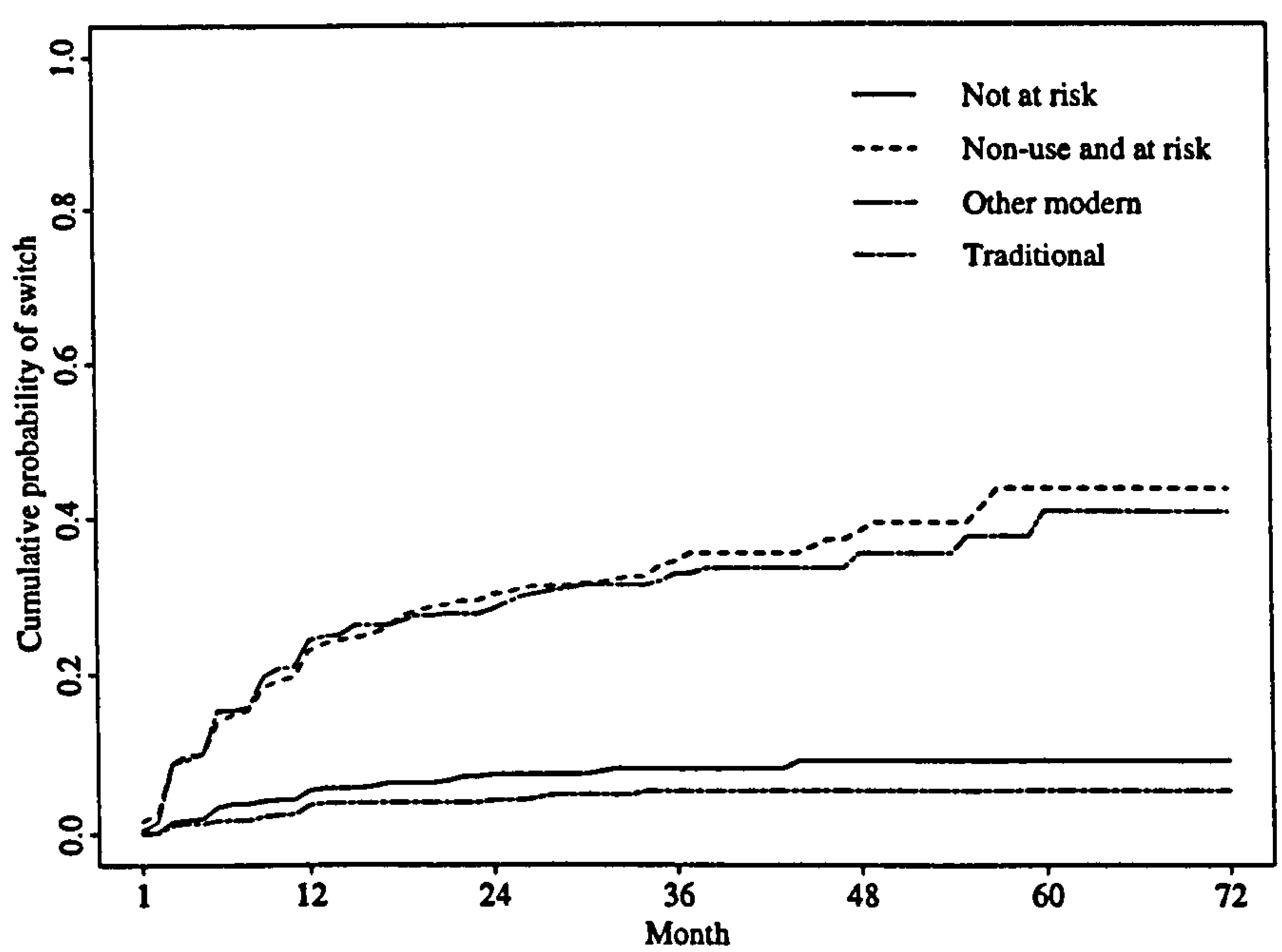


Figure 7.3: Cumulative probabilities of switching from IUD use

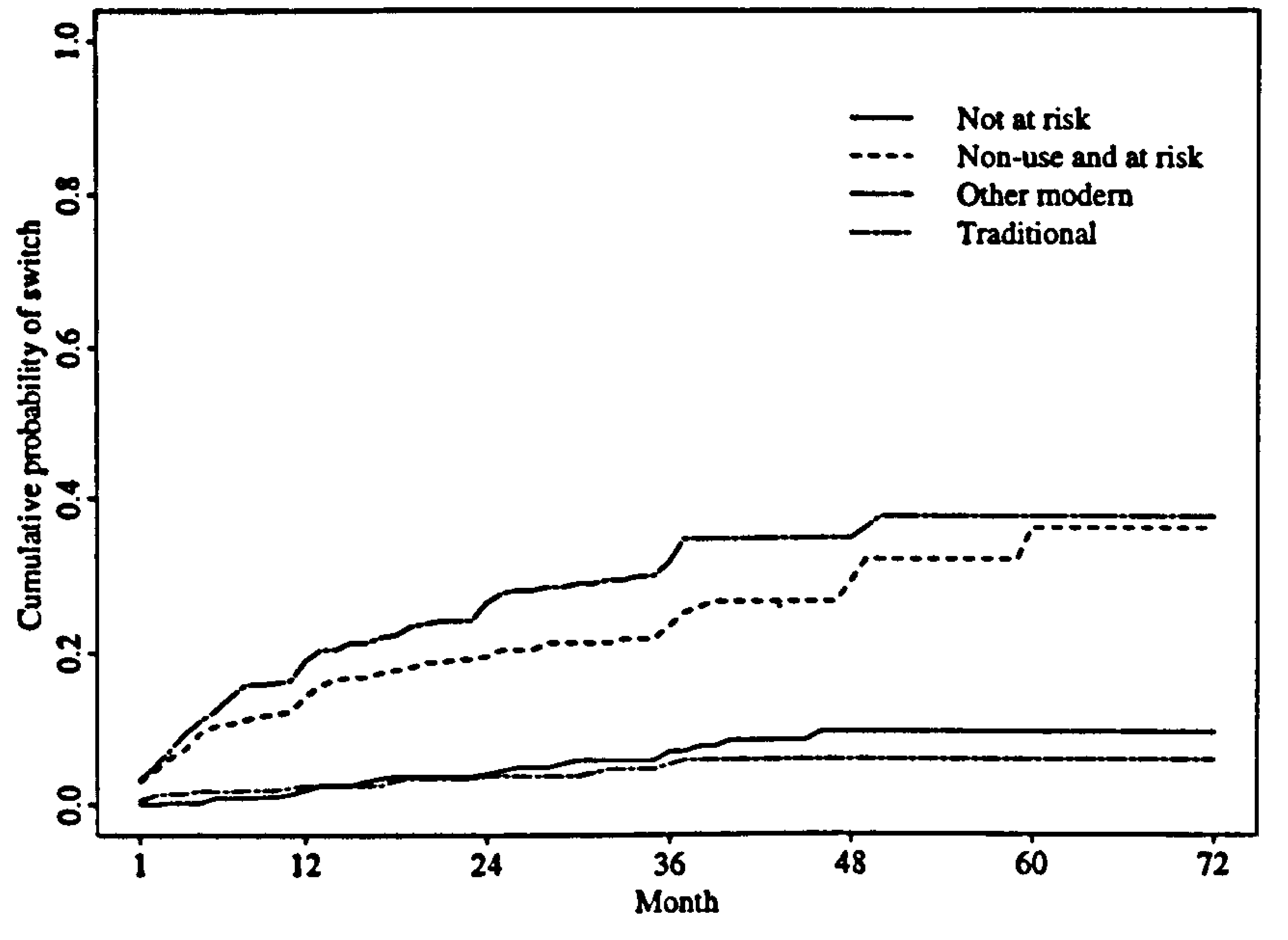


Figure 7.4: Cumulative probabilities of switching from condom use

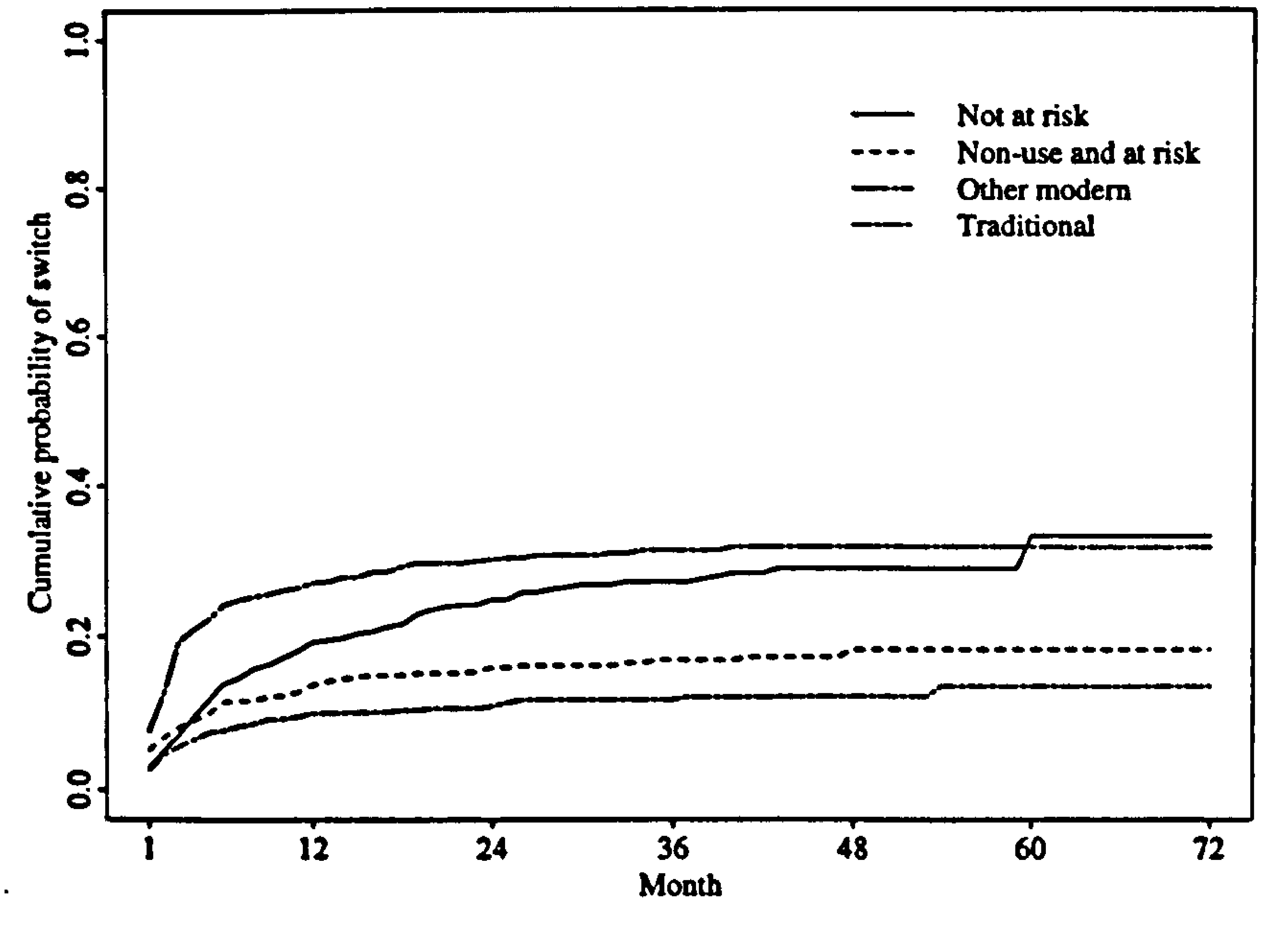
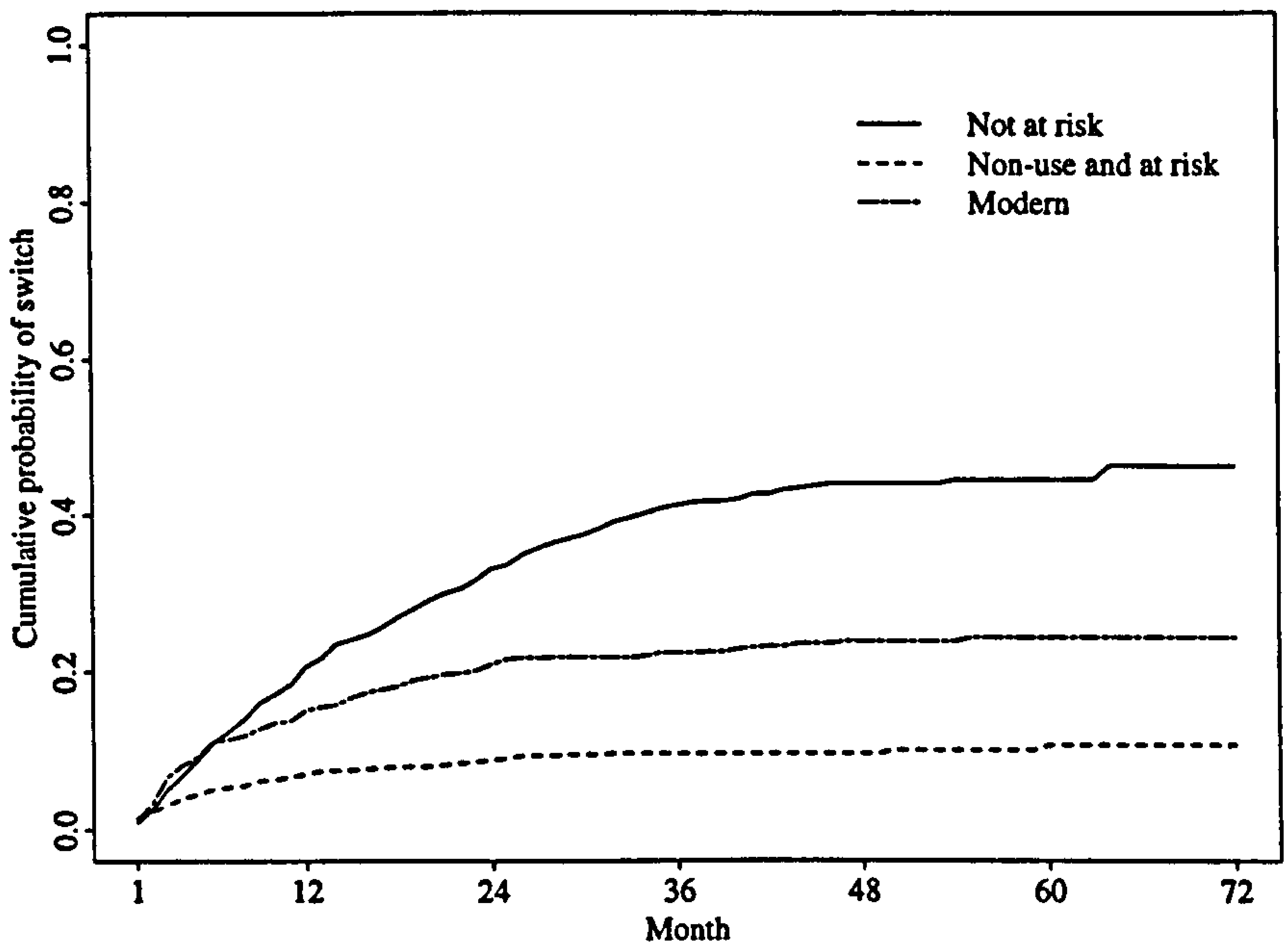


Figure 7.5: Cumulative probabilities of switching from traditional use



7.5.2 Multiple-Decrement Life Table Switching Rates by Selected Background Characteristics

Tables 7.2, 7.3 and 7.4 present 24-month cumulative switching probabilities for pill, other modern method and traditional users respectively by selected background characteristics: division, type of region of residence, education and age at start of use. These probabilities are calculated using multiple-decrement life tables.

Division

There are large differences between the administrative divisions in the risk of switching within 24 months of starting use to non-use while at risk of an unwanted pregnancy. For pill users, Chittagong has the highest rate, while Rajshahi has the lowest. A similar, though less pronounced, pattern is observed among users of other modern methods. The high rates in Chittagong could be attributed to a number of factors. Strong religious conservatism has consistently led to low levels of contraceptive use in Chittagong compared to the rest of the country. Also the difficult terrain, with hilly and remote areas and areas which are flooded during the monsoon, causes access problems for family planning workers. During 1993-94, the proportion of women who had received a visit from a fieldworker during the last six months was only 29% in Chittagong compared with 38% nationwide (Mitra et al. 1994). The effect of this is likely to be more acute in an area such as Chittagong where female mobility is low and women are heavily reliant on doorstep delivery of family planning. In contrast, Rajshahi has one of the highest fieldworker visitation rates and is the division with the highest proportion of areas with both government and NGO workers.

The proportion of pill users switching to an alternative modern method is considerably lower in Chittagong, though there are few differences between the other divisions. For users of modern methods other than the pill, the risk of switching to another modern method is highest in Dhaka and Rajshahi. The degree of switching between modern methods is likely to be affected by supply factors as well as individual characteristics and, in particular, is a reflection of the range of contraceptive options made available by the family planning services. Although the pill is widely available in all divisions, satellite clinics in Chittagong are the least likely to offer IUDs and injectables.

Among traditional users, switches to modern use are most likely in Rajshahi and least likely in Chittagong. There is little evidence of divisional differences in switches to traditional methods for users of modern methods.

Type of region of residence

Women living in rural areas are more likely to switch from using a modern or traditional method to non-use while still in need of contraception than urban women. Further, rural modern method users are less likely to switch to an alternative modern method, probably due to the more limited choice of methods in rural areas. Traditional method users in urban areas have a slightly increased risk of changing to a modern method compared to those in rural areas. However, there are virtually no urban-rural differentials in the likelihood of switching from a modern method to a traditional method.

Table 7.2: 24-month cumulative probabilities of switching for pill users, by selected background characteristics

Variable	Type of transition: Non-use and not at risk [a]	Non-use and at risk	Other modern	Traditional	Continue	n
Division						
Barisal	0.125	0.260	0.152	0.056	0.407	448
Chittagong	0.200	0.300	0.087	0.078	0.335	524
Dhaka	0.142	0.239	0.173	0.052	0.394	1195
Khulna	0.115	0.237	0.175	0.075	0.398	576
Rajshahi	0.131	0.187	0.154	0.054	0.474	1100
Type of residence						
Urban	0.163	0.191	0.195	0.070	0.381	760
Rural	0.136	0.245	0.143	0.057	0.419	3083
Education						
None	0.115	0.259	0.121	0.044	0.461	1594
Incomplete primary	0.147	0.241	0.171	0.052	0.389	776
Complete primary	0.166	0.252	0.134	0.059	0.389	450
Secondary +	0.167	0.182	0.203	0.092	0.356	1023
Age at start of use						
< 20	0.260	0.244	0.136	0.045	0.315	976
20-24	0.146	0.236	0.155	0.054	0.409	1233
25-29	0.076	0.226	0.156	0.067	0.475	850
30-34	0.054	0.227	0.183	0.079	0.457	475
35+	0.035	0.235	0.149	0.086	0.495	309
Overall	0.141	0.234	0.154	0.060	0.411	3843

[a]Discontinuations because a woman wants to become pregnant, or is pregnant after contraceptive failure. Note that women who stop because of menopause, infrequent sex or marital separation are treated as censored.

Table 7.3: 24-month cumulative probabilities of switching for users of modern method other than the pill, by selected background characteristics

Variable	Type of transition					n
	Non-use and not at risk [a]	Non-use and at risk	Another modern	Traditional	Continue	
Division						
Barisal	0.130	0.251	0.266	0.064	0.289	288
Chittagong	0.125	0.260	0.224	0.063	0.328	338
Dhaka	0.141	0.233	0.321	0.046	0.259	706
Khulna	0.137	0.219	0.217	0.096	0.331	391
Rajshahi	0.145	0.166	0.333	0.080	0.276	630
Type of residence						
Urban	0.126	0.174	0.326	0.065	0.309	545
Rural	0.141	0.232	0.274	0.069	0.284	1808
Education						
None	0.120	0.253	0.132	0.055	0.340	902
Incomplete primary	0.118	0.258	0.305	0.065	0.254	451
Complete primary	0.105	0.224	0.310	0.069	0.292	270
Secondary +	0.183	0.149	0.333	0.087	0.248	730
Age at start of use						
< 20	0.240	0.236	0.280	0.077	0.167	485
20-24	0.167	0.195	0.303	0.070	0.265	677
25-29	0.129	0.199	0.296	0.054	0.322	568
30-34	0.048	0.232	0.294	0.061	0.365	368
35+	0.023	0.273	0.220	0.090	0.394	255
Overall	0.138	0.218	0.286	0.068	0.290	2353

[a]Discontinuations because a woman wants to become pregnant, or is pregnant after contraceptive failure. Note that women who stop because of menopause, infrequent sex or marital separation are treated as censored.

Table 7.4: 24-month cumulative probabilities of switching for traditional method users, by selected background characteristics

Variable	Type of transition: Non-use and not at risk [a]	Non-use and at risk	Modern	Continue	n
Division					
Barisal	0.343	0.135	0.193	0.329	214
Chittagong	0.380	0.097	0.137	0.387	239
Dhaka	0.332	0.088	0.206	0.274	385
Khulna	0.312	0.065	0.219	0.404	258
Rajshahi	0.300	0.073	0.253	0.375	408
Type of residence					
Urban	0.340	0.067	0.243	0.350	243
Rural	0.327	0.092	0.202	0.380	1261
Education					
None	0.333	0.089	0.158	0.420	589
Incomplete primary	0.368	0.091	0.173	0.367	280
Complete primary	0.273	0.085	0.223	0.419	205
Secondary +	0.325	0.084	0.294	0.297	430
Age at start of use					
< 20	0.519	0.102	0.135	0.244	373
20-24	0.350	0.097	0.278	0.275	404
25-29	0.316	0.073	0.225	0.387	293
30-34	0.177	0.100	0.223	0.500	227
35+	0.103	0.051	0.167	0.678	207
Overall	0.329	0.088	0.208	0.375	1504

[a]Discontinuations because a woman wants to become pregnant, or is pregnant after contraceptive failure. Note that women who stop because of menopause, infrequent sex or marital separation are treated as censored.

Education

Primary level education, either incomplete or complete, does not appear to affect transitions from modern method use to non-use and at risk of an unintended pregnancy. It is only at the secondary level or higher that education starts to have an impact and decrease the probability of abandoning use. Education has no effect on transitions from traditional use to non-use while still in need of contraceptive protection. Among pill users, the chance of moving to another modern method increases with education, but there is a decrease for women with complete primary education. For users of other modern methods, women with some education are much more likely to switch to an alternative modern method than those with none at all. The likelihood of moving from a traditional method to a modern method increases monotonically with education, and the probability for women with at least secondary level education is double that for women with no education. Transitions from modern methods to a traditional method are largely unaffected by education, though there is evidence of a higher rate among secondary educated women, particularly for pill users.

Continuation rates are consistently higher for women with no education and lowest for those who have been educated to secondary level or higher. These results would suggest that educated women are more knowledgable about family planning and more informed about method choice. If they are dissatisfied with their current method they will switch to an alternative and are perhaps more willing to try a new method. A higher rate of switches to other modern methods among educated pill users was also found by DaVanzo et al. (1989) in Malaysia who suggest that it may reflect a greater concern about side-effects among educated women: although these women still have a desire to use contraception, their experience with the pill has prompted them to opt for traditional methods.

Age at start of use

Age does not appear to affect the rate of transition from pill use to non-use while at risk of an unwanted pregnancy. However, among users of other types of modern reversible methods there is a J-shaped relationship with age, with women aged 35 or more the most likely to move to non-use while at risk. In contrast, among traditional users, older women are the least likely to switch to non-use.

The risk of switching to another modern method for pill users is highest for the 30-34 age group, then decreases for women aged 35 or more. For users of modern methods other than the pill, age does not affect changes to another modern method before age 35 after which there is a drop in the transition rate. Among traditional users, the risk of switching to a modern method is low for teenagers, but increases sharply for women aged 20-24 then decreases again for women aged 35 or more. This may be because young women are less motivated to change to a more effective method as they are unlikely to have achieved their desired family size and are, therefore, not as concerned as older women about the risk of experiencing a contraceptive failure. The lower rate of switching between modern methods among older women may indicate a reluctance to experiment with new methods. They are probably more experienced in the use of contraceptives and have found a method with which they are satisfied and may, therefore, prefer to continue with their present method as they approach the end of their contraceptive careers.

For pill users, the rate of switching to a traditional method increases monotonically with age. There is a slight increase in the risk of moving to traditional use after age 35 for users of modern methods other than the pill.

Since most women in the 'not at risk' category have discontinued use in order to become pregnant, the observed inverse relationship with age for all methods is expected. This is also likely to be a reflection of higher failure rates among younger women (Mitra and Sabir 1996).

7.5.3 Subsequent Contraceptive Behaviour among Non-users

One important question is what happens to women who discontinue use and do not immediately adopt a new method even though they still have a need for contraceptive protection. These women are clearly at high risk of experiencing an unwanted pregnancy and as such should be a major concern to family planning services. In fact, the majority of women adopt a new method immediately or within one month of discontinuation. A total of 3,725 women were observed to discontinue use while still in need of contraception and 60% of those resumed method use within a month. To examine contraceptive behaviour among those who continue non-use for more than one month, a multiple-decrement life table was constructed with decrements: 1) switch to another method; 2) become pregnant; and 3) continue non-use. The results show that for a woman who does not switch to another method immediately, the probability that she does so within the next 12 months is 0.19 and the probability that she resumes use within 36 months increases to only 0.24. Therefore, unless a woman switches to another method immediately following discontinuation, she is most likely to continue non-use. One would hope that if a couple accepts the idea of contraception and starts to use a method, then they will continue to use some form of contraception as long as they have a need. However, the evidence here shows that this is not the case for a large group of women. It would appear that the family planning services have been unsuccessful in encouraging these women to maintain contraceptive use or in offering them a more suitable alternative if they are dissatisfied with their present method. As a result, the large majority become pregnant, presumably accidentally. In fact, if a woman does not start using another method within one month of discontinuation, her chance of becoming pregnant within the next six months is 0.45. This increases to 0.55 if non-use is continued for up to 12 months.

In order to understand the reasons for abandoning use, it is important to look at the contraceptive experience of these women. In particular, have method-related problems led to a woman being disillusioned with contraception to such an extent that she is discouraged from trying another method? Another possibility could be that no suitable alternative method was offered and rather than resume use of a method with which she is dissatisfied, a woman prefers to discontinue use of contraception altogether. Table 7.5 shows the contraceptive use status of women in the month following discontinuation by their reason for stopping use. Women who do not have a need for contraception are excluded from this part of the analysis, that is, women who are trying to become pregnant, pregnant through contraceptive failure, having infrequent intercourse or menopausal. Only modern reversible methods are considered as they are of prime interest to family planning services and the reasons for discontinuing traditional methods tend to be very different from those given for stopping modern method use. The major reasons for discontinuation given by modern method users are side-effects and health concerns, which account for 68% of discontinuations among women still in need of contraception. Accessibility of family planning services and availability or cost of methods account for relatively few discontinuations (5%).

Table 7.5: Percentage of use intervals by type of transition and reason for discontinuation for women in need of contraception

Method	Reason for discontinuation	Type of transition (%) Non-use and at risk[a]	Another modern	Traditional	n
Pill	Side-effects/health concerns	50.3	35.5	14.1	1244
	Access/availability/cost	60.5	32.9	6.6	76
	Other method-related[b]	33.0	55.3	11.6	103
	Other	63.9	26.1	0.1	230
	Overall	51.6	35.3	13.1	1653
Other modern	Side-effects/health concerns	40.5	50.6	8.9	674
	Access/availability/cost	41.3	41.3	17.5	63
	Other method-related	26.3	53.9	19.9	297
	Other	25.7	21.9	4.9	265
	Overall	37.9	49.9	12.2	1173

[a]Discontinuations in order to become pregnant or because of contraceptive failure

[b]Includes discontinuations because the husband disapproves, inconvenience with use of a method or because a more effective method is needed.

The majority of pill users who discontinue use because of side-effects or health concerns do not switch to another modern method and this is also true for women who stop use because of accessibility or availability problems. Women who discontinue for these reasons are much more likely to continue non-use than those who have stopped for other method-related reasons, and those who stop because of side-effects or health concerns are the most likely to switch to a less effective traditional method. Among users of modern reversible methods other than the pill, the proportion who continue non-use is greater for those who stopped because of side-effects, health concerns or accessibility problems than for any other reason for discontinuation.

7.5.4 Repeated Transitions

It is likely that some women will experience more than one method switch over the six-year observation period. For those women who have repeated switches, one might expect the durations of their use intervals to be correlated. For example, if a woman experiences problems with contraceptive use, she may change methods regularly and thus contribute several short spells of use. In order to control for this clustering, a further level, corresponding to use interval, can be specified in the hierarchical structure of the model. To examine the extent of repeated transitions over the study period, the number of intervals of contraceptive use contributed to the analysis sample per woman is presented (Table 7.6). It can be seen that although the majority of women experience only one switch, a large proportion (44%) contribute more than one use spell, accounting for 68% of all spells included in the analysis.

Table 7.6: Number of use intervals contributed to the analysis sample per woman

No. intervals	No. women	% women
1	2493	55.6
2	1284	28.6
3	481	10.7
4+	228	5.1
Total	4486	100.0

However, after disaggregating by the type of transitions made (Table 7.7), we can see that relatively few women have made the same kind of switch more than once during the six-year observation window. The transitions most likely to occur more than once are switches to non-use when the woman is not in need of contraception (mainly discontinuations in order to become pregnant) and switches to a modern method. Further analysis shows that when use intervals are grouped into method types, only 18% of pill users, 19% of users of modern methods other than the pill and 14% of traditional users contribute more than one spell of the same method type. When we examine the types of switches that occur from pill, other modern method and traditional use intervals separately, very few women are found to have experienced the same type of switch from the same method more than once. Therefore, in this case, since a separate analysis is carried out for each type of method and a competing risks framework is used to distinguish between different kinds of transition, it would appear that it is not necessary to add additional use interval-specific random effects to the model. Of course, the lack of multiple spells of the same type is likely to be due in part to the relatively short observation period. If complete contraceptive histories were available, it is likely that we would find more women experiencing repeated transitions.

Table 7.7: Number of transitions per woman by type of transition

No. transitions	No. of women by type of transition				
	Non-use and not at risk[a]	Non-use and and at risk	Modern	Traditional	Continue
0	3342	3249	3504	4164	1350
1	1059	1095	723	293	3072
2	84	122	196	25	43
3+	1	20	63	4	21
Total	4486	4486	4486	4486	4486

[a]Discontinuations because a woman wants to become pregnant, or is pregnant after contraceptive failure

7.6 Choice of Covariates

The BDHS woman and household questionnaires collected a large amount of socioeconomic and demographic data together with information on membership of women's savings groups, employment and women's status. However, in a longitudinal analysis to investigate changes in contraceptive behaviour over a period of time, care must be taken in the choice of potential covariates. Most variables are only measured at the time of the survey and although many of these, such as education and socioeconomic indicators, are likely to remain static over the observation period, there are some which we cannot assume to be constant over time. For example, since women's savings groups are a relatively recent innovation in Bangladesh, we cannot assume that a woman who was a member at the time of interview was also a member during the entire six-year period for which we have contraceptive use data. For the same reasons, much of the data collected in the service availability schedule are of limited use in a longitudinal analysis since we do not know how long these services have been available. Unfortunately, this precludes the use of much of the more detailed information on quality of services. Therefore, for this analysis the choice of covariates is restricted to variables for which we have time-dependent information, for example age or family composition, or those which we can assume have remained relatively constant throughout the six-year window.

The demographic variables considered were administrative division, type of region of residence, family composition and the woman's age at the start of the use interval. The family composition variable, calculated from the complete birth histories, is a count of the number of children alive at a given point in time and is time-dependent since the count can decrease during an interval of contraceptive use if a child dies. The variable also incorporated information on the children's sex to test whether there is any evidence of a son preference effect on transition rates. A number of socioeconomic covariates were also considered. These included woman's education level, husband's occupation, religion, and characteristics of the household. Since 99.5% of ever-married women in the sample are either Islam or Hindu, other religions were excluded from the analysis. Some household-level variables, including the source of drinking water, main building material of the house and the number of possessions owned by the household were also considered as potential covariates.

In addition to a woman's background characteristics, some covariates relating to contraceptive use were considered. For the analysis of transitions from modern methods other than the pill, the method used was included as a covariate to allow for the differences in transition rates observed between users of injectables, IUDs and condoms (Figures

7.2, 7.3 and 7.4). Also of interest is the extent to which prior contraceptive experience influences the probability of switching from the current method. Unfortunately, since contraceptive use is known only for a period of six years, it is not possible to determine the number of previous uses and the frequency with which problems have been encountered over a woman's contraceptive career. Therefore, covariates capturing the experience with the previous method only were considered: two variables were created to examine the effects of the reason for discontinuation and the duration of use of the last method on the likelihood of switching from the present method. In particular, this allows us to examine the behaviour of women who switched to another method after experiencing problems with their last method and to see whether this pattern is repeated. Similarly, a woman who used her last method for a very brief period of time may tend to use her next method for a shorter duration than those who have a history of more prolonged use.

Several variables collected in the service availability questionnaire were considered as potential covariates. The distance to the *thana* headquarters was selected as a proxy for the remoteness and accessibility of an area. As measures of the level of socioeconomic development of the cluster, two variables were used: the proportion of households in the village or mohalla living in one room and the proportion of non-pukka houses, that is houses constructed with basic building materials such as bamboo and mud. These proportions were estimated by a key informant living in the cluster, for example, a village leader or school teacher. As discussed earlier, much of the detailed information on family planning service provision is only relevant to the time of the survey and as such is unsuitable for inclusion in a longitudinal analysis. Therefore, only one family planning variable was selected: the presence of a family welfare centre (FWC) in the cluster. This information was collected from family welfare visitors who are based at FWCs. Since they were asked how long ago each FWC in the area was established, it was possible to construct a time-dependent covariate indicating the presence or absence of a FWC at any point in time during the six-year observation window. Unfortunately, it was not possible to construct a similar time-dependent variable from data collected on house-to-house visits made by family planning workers, a service which has had considerable impact on contraceptive use levels in Bangladesh (Kamal 1995b).

7.7 Results from the Multilevel Discrete-time Competing Risks Models

7.7.1 Model Fitting and Selection

Initially, models with one month between discrete time points were considered. Using this approach, each interval of contraceptive use generates a series of records, one for each month of the duration of method use (Section 4.4.2). However, this results in a huge data set: for example, the 3,826 pill spells generate 57,709 records. Further, in order to fit the multilevel competing-risks model, a multivariate data structure is required where the multinomial response for each month generates a set of binary responses, one for each response category except for the reference category (see Section 2.6.2). In the model for transitions from pill use, there are five response outcomes and therefore the number of cases is multiplied by four to give a total of 230,836 records. In addition, the number of parameters in the model is large, since in a multinomial model a separate set is estimated for each contrast of an outcome with the baseline. In an attempt to reduce the size of the data set, single-level models were fitted in which the width of the interval between discrete-time points was increased. These initial models showed that an increase to three months had almost no effect on the parameter estimates and standard errors, but a further increase to six months lead to a more marked change in some of the estimates. Models in which contraceptive use status was recorded in three month intervals were, therefore, employed.

Since the data sets generated for a discrete-time hazards analysis are so large, fitting a multilevel model is extremely slow and as a result the process of model selection is very time-consuming. For this reason, single-level models were used to aid model screening. For comparative purposes, it was decided to retain the same set of covariates in each of the models for transitions from the pill, other modern methods and traditional methods. Forward selection was used whereby a potential covariate was tested for significance in each of the three models and rejected only if it was found to be non-significant in all three models. If at any stage of the forward selection procedure a covariate was found to be no longer significant, it was subsequently removed. After finding the significant main effects, all two-way interactions between the selected covariates were tested for significance.

A number of different formulations for the duration effects were considered: a categorical representation where the hazard of switching is assumed to be piecewise-constant, log(duration), and linear and quadratic formulations. The quadratic representation, where duration and squared duration terms were incorporated in the model, was found to fit the observed hazards very closely. To test for proportionality of the hazards, all two-way interactions between duration and the other selected covariates were considered.

An additional benefit of fitting the single-level models first is that the estimates obtained can be used as starting values for the parameters in the fixed part of the multilevel model. Due to the large number of cases and model parameters, convergence is very slow and sometimes cannot be achieved unless suitable starting values are provided. If convergence problems are still encountered using the single-level parameter estimates, starting values for the random-part parameters may also be obtained by fitting a series of multilevel pairwise logistic models as an approximation to the multinomial. However, in this case, starting values for the fixed part parameters obtained from the single-level analysis were found to be sufficient for convergence to be attained. Models using the first-order MQL approximation were first fitted and after convergence these were then extended to first-order PQL followed by second-order PQL approximations (see Section 2.9 for a discussion of model fitting strategies for binary and multinomial response data).

7.7.2 Interpretation of the Fixed Effects

The parameter estimates and standard errors from the selected multilevel competing risks models for transitions from pill, other modern and traditional use are presented in Tables 7.8, 7.9 and 7.10 respectively. To aid interpretation, estimates of the hazard functions for each three-month interval for a period of two years have been calculated from the fitted model. Using formulae presented in Section 4.6.2, these are used to derive two-year cumulative probabilities of switching (Tables 7.11, 7.12 and 7.13). The effects of each covariate are considered separately, while fixing all other covariates in the model at average values calculated from the sample of contraceptive users. To calculate the averages, all three analysis samples were combined, so that the covariates were fixed at the same values in each model. Since all covariates are categorical, average values correspond to the proportions in each category. At this stage, since we are focusing on the impact of the fixed effects, the random effects for the district, cluster and woman levels are held at their average values of zero.

Age at start of use

The age at the start of the use interval was found to have a significant impact on transition rates, though its significance was considerably reduced after controlling for the number of surviving children. For all method types, the probability of a switch to non-use while not at risk of an unintended pregnancy is lower among older women: women aged 30 or more (or over 35 for pill users) are much less likely to discontinue use to become pregnant or because of contraceptive failure. Among users of modern methods other than the pill, the chance of moving to another type of modern method decreases as age increases. Since most transitions of this type are to the pill, this indicates a reluctance among older women to switch from IUD, injectables or condom use to the pill. This could be due to difficulties with pill use in the past or because these women now wish to limit rather than space their births and, therefore, prefer a more long-term method such as the IUD. For traditional users, the probability of switching to a modern method shows an inverted J-shaped relationship with age. Teenage women are less likely than those in their 20s or early 30s to switch to a modern method, probably because they are at the start of their childbearing and are less committed to avoiding an unintentional pregnancy by using an effective contraceptive method. The lowest transition rates from traditional to modern method use are experienced by women aged 35 or more. Many of these women are near the end of their reproductive careers, are likely to be having less frequent intercourse and, therefore, may not consider themselves in need of a more effective method at this stage in their lives.

Table 7.8: Estimates (and standard errors) from the multilevel discrete-time competing risks model for transitions from pill use

Variable	Type of transition							
	Non-use and not at risk/ Continue		Non-use and at risk/ Continue		Other Modern/ Continue		Traditional/ Continue	
	Est.	(S.E.)	Est.	(S.E.)	Est.	(S.E.)	Est.	(S.E.)
Constant	-5.01	(0.27)	-2.77	(0.19)	-3.00	(0.21)	-4.52	(0.35)
Duration	0.06**	(0.03)	-0.24***	(0.02)	-0.20***	(0.02)	-0.19***	(0.04)
Duration^2	-0.003	(0.002)	0.01***	(0.002)	0.01***	(0.002)	0.01***	(0.003)
Age (Base=20-24)								
< 20	-0.12	(0.12)	-0.09	(0.11)	0.003	(0.15)	0.11	(0.25)
25-29	-0.03	(0.17)	-0.003	(0.11)	-0.10	(0.14)	0.01	(0.23)
30-34	-0.03	(0.26)	0.06	(0.14)	0.03	(0.17)	0.13	(0.27)
35+	-0.61*	(0.37)	-0.03	(0.17)	-0.30	(0.21)	0.21	(0.31)
No. children (Base=3+)								
0	4.00***	(0.22)	1.43***	(0.18)	0.32	(0.25)	0.13	(0.42)
1	2.07***	(0.19)	0.62***	(0.13)	-0.21	(0.17)	-0.17	(0.28)
2	1.14***	(0.18)	0.31***	(0.11)	-0.12	(0.14)	-0.40*	(0.23)
Education (Base=none)								
Primary	0.22**	(0.11)	-0.12	(0.09)	0.33***	(0.12)	0.17	(0.19)
Secondary+	-0.51***	(0.13)	-0.58***	(0.11)	0.49***	(0.13)	0.78***	(0.20)
Division (Base=Dhaka)								
Barisal + Khulna	-0.30*	(0.16)	-0.04	(0.13)	0.05	(0.14)	0.23	(0.22)
Chittagong	0.36**	(0.16)	0.31**	(0.14)	-0.50***	(0.19)	0.17	(0.26)
Rajshahi	-0.44***	(0.15)	-0.43***	(0.13)	-0.15	(0.14)	-0.03	(0.22)
Residence and miles to thana HQ (Base=rural, ≥ 5)								
Urban	0.08	(0.13)	-0.11	(0.12)	0.30**	(0.15)	0.07	(0.22)
Rural, < 5	-0.19*	(0.11)	-0.18*	(0.10)	0.19	(0.13)	0.28	(0.18)
Religion (Base=Muslim)								
Hindu	-0.02	(0.14)	-0.41***	(0.13)	-0.25	(0.16)	0.14	(0.22)
Reason for stopping last method (Base=nonmethod-related)								
No last method	0.21	(0.15)	0.51***	(0.13)	-0.11	(0.13)	-0.04	(0.23)
Method-related	0.15	(0.17)	0.35**	(0.14)	-0.23	(0.15)	0.31	(0.24)
Random effect variances								
Woman-level σ^2_{ur}	0.001	(0.001)	0.41***	(0.09)	1.27***	(0.15)	2.25***	(0.38)
Cluster-level σ^2_{vr}	0.02	(0.05)	0.10**	(0.04)	0.17**	(0.07)	0.08	(0.15)
District-level σ^2_{wr}	0.06*	(0.04)	0.03	(0.03)	0.001	(0.001)	0.01	(0.07)

* p=0.10; ** p=0.05; *** p=0.01

Table 7.9: Estimates (and standard errors) from the multilevel discrete-time competing risks model for transitions from other modern method use

Variable	Type of transition							
	Non-use and not at risk/ Continue		Non-use and at risk/ Continue		Another Modern/ Continue		Traditional/ Continue	
	Est.	(S.E.)	Est.	(S.E.)	Est.	(S.E.)	ESt.	(S.E.)
Constant	-3.41	(0.34)	-2.95	(0.27)	-1.93	(0.23)	-5.49	(0.53)
Duration	-0.03	(0.12)	-0.15***	(0.05)	-0.20***	(0.05)	-0.06	(0.16)
Duration^2	-0.002	(0.01)	0.009**	(0.004)	0.01***	(0.004)	-0.004	(0.02)
Method (Base=injectables)								
IUD	-1.71***	(0.57)	-0.43**	(0.20)	-0.43**	(0.19)	-0.24	(0.50)
Condom	1.36***	(0.25)	0.12	(0.17)	0.62***	(0.15)	1.87***	(0.33)
Method by Duration Interaction								
IUD*Duration	0.34	(0.20)	-0.09	(0.08)	0.06	(0.08)	-0.01	(0.23)
IUD*Duration^2	-0.01	(0.01)	0.007	(0.006)	-0.004	(0.006)	0.01	(0.02)
Condom*Duration	-0.07	(0.13)	-0.27***	(0.09)	-0.31***	(0.08)	-0.40**	(0.19)
Condom*Duration^2	0.007	(0.01)	0.01***	(0.007)	0.01***	(0.007)	0.03	(0.02)
Age (Base=20-24)								
< 20	-0.28	(0.18)	0.16	(0.18)	0.39**	(0.16)	0.23	(0.29)
25-29	-0.06	(0.19)	0.11	(0.16)	-0.16	(0.14)	-0.34	(0.28)
30-34	-0.88***	(0.33)	0.27	(0.19)	-0.32	(0.17)	0.05	(0.32)
35+	-1.93***	(0.53)	0.25	(0.21)	-0.64***	(0.21)	0.13	(0.35)
No. children (Base=3+)								
0	2.27***	(0.28)	0.70**	(0.28)	-0.55**	(0.27)	0.38	(0.42)
1	1.24***	(0.23)	0.72***	(0.19)	-0.09	(0.17)	-0.10	(0.34)
2	0.42*	(0.22)	0.22	(0.16)	0.02	(0.14)	0.08	(0.27)
Education (Base=None)								
Primary	-0.39**	(0.17)	0.15	(0.12)	0.27**	(0.12)	0.12	(0.23)
Secondary+	-0.75***	(0.18)	-0.20	(0.16)	0.30**	(0.14)	0.04	(0.25)
Division (Base=Dhaka)								
Barisal + Khulna	-0.37**	(0.17)	-0.04	(0.14)	-0.40**	(0.16)	0.26	(0.25)
Chittagong	-0.41*	(0.23)	-0.04	(0.14)	-0.43**	(0.20)	-0.09	(0.33)
Rajshahi	-0.22	(0.17)	-0.42***	(0.16)	-0.03	(0.16)	0.36	(0.25)
Residence and miles to thana HQ (Base=rural, $\geq$ 5)								
Urban	-0.46***	(0.18)	-0.27	(0.16)	-0.10	(0.14)	-0.30	(0.25)
Rural, < 5	-0.10	(0.15)	-0.02	(0.13)	-0.09	(0.14)	-0.06	(0.21)
Religion (Base=Muslim)								
Hindu	0.03	(0.21)	-0.61***	(0.21)	-0.10	(0.18)	0.72***	(0.24)
Reason for stopping last method (Base=nonmethod-related)								
No last method	-0.22	(0.20)	0.75***	(0.19)	-0.10	(0.15)	0.92	(0.36)
Method-related	-0.11	(0.20)	0.26	(0.19)	-0.08	(0.14)	0.89***	(0.35)
Random effect variances								
Woman-level σ^2_{ur}	0.001	(0.001)	0.40***	(0.14)	0.75***	(0.13)	0.64*	(0.37)
Cluster-level σ^2_{vr}	0.001	(0.001)	0.09*	(0.07)	0.01	(0.06)	0.07	(0.17)
District-level σ^2_{wr}	0.001	(0.005)	0.002	(0.005)	0.06	(0.04)	0.001	(0.001)

* p=0.10; ** p=0.05; *** p=0.01

Table 7.10: Estimates (and standard errors) from the multilevel discrete-time competing risks model for transitions from traditional use

Variable	Type of transition					
	Non-use and not at risk/ Continue		Non-use and at risk/ Continue		Modern/ Continue	
	Est.	(S.E.)	Est.	(S.E.)	Est.	(S.E.)
Constant	-2.56	(0.24)	-3.24	(0.45)	-2.26	(0.29)
Duration	0.11***	(0.04)	-0.24***	(0.07)	-0.11**	(0.05)
Duration^2	-0.01***	(0.003)	0.01	(0.01)	0.001	(0.004)
Age (Base=20-24)						
<20	0.06	(0.15)	-0.13	(0.32)	-0.51**	(0.24)
25-29	-0.13	(0.16)	-0.23	(0.31)	-0.40**	(0.20)
30-34	-0.96***	(0.23)	-0.31	(0.35)	-0.68***	(0.25)
35+	-1.82***	(0.29)	-1.06**	(0.42)	-1.17***	(0.27)
No. children (Base=3+)						
0	0.78***	(0.22)	0.28	(0.43)	-0.20	(0.33)
1	0.44**	(0.18)	-0.17	(0.36)	-0.28	(0.25)
2	-0.13	(0.18)	-0.03	(0.31)	-0.13	(0.21)
Education (Base=none)						
Primary	-0.11	(0.12)	0.03	(0.24)	0.18	(0.18)
Secondary +	-0.21	(0.14)	0.14	(0.27)	0.61***	(0.19)
Division (Base=Dhaka)						
Barisal + Khulna	0.06	(0.14)	0.09	(0.31)	0.03	(0.19)
Chittagong	0.24	(0.16)	0.08	(0.34)	-0.65**	(0.27)
Rajshahi	0.02	(0.15)	-0.16	(0.32)	0.21	(0.19)
Residence and miles to thana HQ (Base=rural, ≥ 5)						
Urban	0.05	(0.15)	-0.25	(0.35)	0.24	(0.22)
Rural, <5	-0.16	(0.12)	0.30	(0.24)	0.61***	(0.16)
Religion (Base=Muslim)						
Hindu	-0.10	(0.14)	-0.33	(0.31)	0.02	(0.19)
Reason for stopping last method (Base=Nonmethod-related)						
No last method	-0.24	(0.16)	0.003	(0.31)	-0.82***	(0.19)
Method-related	-0.12	(0.17)	0.01	(0.32)	-0.42**	(0.19)
Random effect variances						
Woman-level σ^2_{ur}	0.001	(0.001)	0.60	(0.45)	0.73***	(0.19)
Cluster-level σ^2_{vr}	0.001	(0.002)	0.34	(0.27)	0.001	(0.002)
District-level σ^2_{wr}	0.001	(0.003)	0.11	(0.15)	0.002	(0.002)

* p=0.10; ** p=0.05; *** p=0.01

Table 7.11: Estimated 24-month cumulative probabilities of switching from pill use

Variable[a]	Type of transition Non-use and not at risk	Non-use and at risk	Other modern	Traditional	Continue
Age at start					
<20	0.093	0.233	0.179	0.064	0.431
20-24	0.102	0.251	0.176	0.056	0.414
25-29	0.101	0.254	0.161	0.058	0.427
30-34	0.098	0.261	0.177	0.063	0.401
35+	0.060	0.257	0.138	0.073	0.471
No. surviving children					
0	0.526	0.312	0.099	0.032	0.031
1	0.204	0.290	0.128	0.052	0.326
2	0.096	0.244	0.161	0.047	0.452
3+	0.033	0.188	0.192	0.075	0.512
Education					
None	0.100	0.303	0.132	0.046	0.419
Primary	0.120	0.262	0.177	0.052	0.389
Secondary +	0.062	0.173	0.221	0.103	0.441
Division					
Barisal + Khulna	0.079	0.254	0.192	0.069	0.406
Chittagong	0.141	0.341	0.104	0.061	0.353
Dhaka	0.105	0.263	0.181	0.054	0.397
Rajshahi	0.078	0.191	0.173	0.059	0.499
Residence and miles to thana HQ					
Urban	0.104	0.239	0.196	0.057	0.404
Rural, < 5	0.083	0.229	0.182	0.072	0.434
Rural, ≥ 5	0.100	0.272	0.149	0.054	0.426
Religion					
Muslim	0.093	0.259	0.173	0.059	0.416
Hindu	0.101	0.188	0.146	0.074	0.491
Reason for stopping last method					
No last method	0.096	0.280	0.169	0.052	0.404
Nonmethod-related	0.086	0.181	0.204	0.058	0.471
Method-related	0.094	0.245	0.155	0.076	0.429
Overall average	0.094	0.250	0.170	0.061	0.426

[a]Other covariates are fixed at the sample averages and all random effects are fixed at their average values of 0.

Table 7.12: Estimated 24-month cumulative probabilities of switching from other modern method use

Variable[a]	Type of transition: Non-use and not at risk	Non-use and at risk	Another modern	Traditional	Continue
Method					
Injectables	0.062	0.279	0.309	0.037	0.313
IUD	0.037	0.191	0.275	0.038	0.458
Condom	0.174	0.165	0.295	0.112	0.255
Age at start					
<20	0.083	0.198	0.404	0.058	0.257
20-24	0.127	0.190	0.307	0.052	0.325
25-29	0.125	0.218	0.269	0.038	0.350
30-34	0.057	0.264	0.237	0.059	0.383
35+	0.022	0.278	0.185	0.069	0.447
No. surviving children					
0	0.348	0.270	0.146	0.058	0.178
1	0.143	0.303	0.254	0.040	0.260
2	0.074	0.208	0.318	0.055	0.346
3+	0.052	0.175	0.329	0.054	0.391
Education					
None	0.124	0.225	0.249	0.051	0.351
Primary	0.078	0.246	0.306	0.054	0.315
Secondary +	0.059	0.186	0.338	0.053	0.364
Division					
Barisal + Khulna	0.078	0.248	0.243	0.060	0.371
Chittagong	0.077	0.252	0.239	0.043	0.388
Dhaka	0.100	0.234	0.328	0.041	0.300
Rajshahi	0.087	0.164	0.341	0.064	0.345
Residence and miles to thana HQ					
Urban	0.068	0.191	0.300	0.046	0.398
Rural, < 5	0.090	0.230	0.282	0.054	0.344
Rural, ≥ 5	0.096	0.228	0.298	0.055	0.323
Religion					
Muslim	0.089	0.233	0.293	0.048	0.341
Hindu	0.095	0.135	0.282	0.106	0.383
Reason for stopping last method					
No last method	0.075	0.289	0.269	0.057	0.310
Nonmethod-related	0.106	0.151	0.332	0.026	0.385
Method-related	0.091	0.189	0.298	0.061	0.360
Overall average	0.087	0.220	0.293	0.053	0.348

[a]Other covariates are fixed at the sample averages and all random effects are fixed at their average values of 0.

Table 7.13: Estimated 24-month cumulative probabilities of switching from traditional method use

Variable[a]	Type of transition: Non-use and not at risk	Non-use and at risk	Modern	Continue
Age at start				
<20	0.412	0.099	0.171	0.318
20-24	0.354	0.105	0.266	0.275
25-29	0.358	0.092	0.198	0.352
30-34	0.193	0.098	0.178	0.532
35+	0.097	0.053	0.127	0.723
No. surviving children				
0	0.449	0.109	0.158	0.284
1	0.371	0.078	0.166	0.385
2	0.231	0.095	0.206	0.468
3+	0.250	0.095	0.226	0.430
Education				
None	0.330	0.091	0.161	0.419
Primary	0.295	0.093	0.191	0.421
Secondary +	0.249	0.099	0.277	0.376
Division				
Barisal + Khulna	0.290	0.102	0.208	0.400
Chittagong	0.362	0.104	0.110	0.424
Dhaka	0.280	0.095	0.207	0.418
Rajshahi	0.278	0.079	0.249	0.395
Residence and miles to thana HQ				
Urban	0.326	0.069	0.197	0.408
Rural, < 5	0.247	0.113	0.269	0.373
Rural, ≥ 5	0.320	0.091	0.158	0.432
Religion				
Muslim	0.232	0.158	0.253	0.357
Hindu	0.219	0.119	0.270	0.392
Reason for stopping last method				
No last method	0.224	0.161	0.204	0.410
Nonmethod-related	0.231	0.130	0.376	0.264
Method-related	0.232	0.148	0.278	0.342
Overall average	0.230	0.152	0.255	0.363

[a]Other covariates are fixed at the sample averages and all random effects are fixed at their average values of 0.

Family composition

Family composition was initially measured by both the number and sex of living children. This was later simplified to the number of children as there was no evidence that the sex composition of children influenced any type of method switch. For all methods, there is a large decline in transition rates to non-use while in need of contraception as the number of children increases. This is as expected since this group consists largely of women who have stopped to become pregnant. Among users of modern methods, the risk of moving to non-use while at risk of an unintended pregnancy also shows a negative relationship with family size. This suggests that women with few children are less motivated to continue use of contraception. For these women, the consequences of an unwanted pregnancy are less severe than for women who already have three or more children. One might have expected the rate of transitions from traditional to modern methods to increase with the number of children: women who have achieved their desired family size may be more likely to move to a more effective method to avoid any further pregnancies. However, in this case, although the effect is in the expected direction, it is not statistically significant. Users of modern methods other than the pill are less likely to switch to another type of modern method if they do not have children. It is possible that this group of women prefers to continue with their current method until they plan to have a child, and perhaps wait until after their first pregnancy to switch to another method.

Education

For all methods, a woman's level of education influences her likelihood of switching. Initially, education was represented by a variable with four categories: no education, incomplete primary, complete primary and secondary or higher. Since there were no significant differences between women with incomplete and complete primary education, these categories were subsequently combined. The lower rate of transitions from modern methods to non-use and not at risk among educated women may be explained by their lower levels of fertility and lower failure rates. For pill users, the risk of switching to non-use while at risk of an unintended pregnancy is lower if the woman has been educated to at least secondary level. However, this effect is nonsignificant among users of other types of modern method. There is also evidence of a greater degree of switching between modern methods, and from traditional to modern, for educated women. Women who have received some level of education tend to be more aware and knowledgeable about family planning. They experience fewer problems with use and are therefore less likely to discontinue while still in need of contraceptive protection. It is probable that if they do become dissatisfied with a method, they are more likely to switch to a more suitable alternative since they have a greater awareness of the contraceptive options available and know where they can be obtained. There is also evidence that pill users who have at least secondary level education have an increased chance of switching to a traditional method, although this is not significant for users of other types of modern method. It is possible that this is a reflection of a greater concern about side-effects and the long-term effects of the pill. If no other type of modern method is available, they may prefer to opt for a traditional method rather than continue with the pill or abandon contraceptive use altogether.

Division

There is evidence of differences in transition rates between administrative divisions. However, since the rates for Khulna and Barisal did not differ significantly and because Barisal was part of Khulna until 1992, these two categories were amalgamated. For pill users, the risk of discontinuing use in order to become pregnant or because of contraceptive failure is highest in Chittagong. Chittagong also has the highest rate of switching from pill use to non-use while at risk of an unintended pregnancy, and the lowest rate of switching between modern methods, or from traditional to a more effective modern method. This is likely to be a reflection of the high level of religious conservatism and restrictions on woman's mobility in this area. The analysis presented in Chapter 3 found these factors to have an impact on contraceptive acceptance, and the results here suggest that they continue to have an effect on contraceptive behaviour once family planning has been adopted. Problems of accessibility are also likely to affect transition rates. These are particularly acute in Chittagong where women are heavily reliant on doorstep delivery for obtaining family planning because of their low level of mobility, and therefore pills and condoms are often the only methods available. A lack of a range of contraceptive options may lead women who are dissatisfied with the pill to abandon use even though they are still in need of contraceptive protection. Similarly, traditional users may be unable to obtain a modern method which suits their needs. In contrast, Rajshahi consistently experiences low rates of switching from modern method use to non-use while at risk of an unintended conception, high continuation rates and a high level of movement between modern methods.

Type of region of residence

The type of region of residence was initially represented by two categories: rural or urban. In order to incorporate some measure of the degree of accessibility and remoteness of an area, a cluster-level variable, distance to the *thana* headquarters, was also used. Since nearly all urban clusters are within one mile of the *thana* headquarters, this variable is only relevant in rural areas. Therefore, region was represented by three categories: 1) urban, 2) rural and within five miles of the *thana* headquarters, and 3) rural and five miles or more from the *thana* headquarters. The results show that pill users in rural areas are less likely to switch to non-use while at risk if they live in a less remote area, within five miles of the *thana* headquarters. Further, women using the pill have a greater chance of changing to another modern method if they live in an urban area or in a less remote rural area. Traditional users are almost twice as likely to adopt a modern method if they live within a five mile radius of the *thana* headquarters than women who live at a greater distance. Interestingly, traditional users living in the less remote rural areas actually have a higher chance of switching to a modern method than their urban counterparts.

Religion

For any type of modern method, the rate of switching to non-use while still in need of contraception is considerably greater for Muslims than for Hindus. Also, Hindu users of modern methods other than the pill are twice as likely as Muslims to switch to a traditional method. There is no evidence of any effect of religion on transitions between modern methods.

Reason for stopping last method

Two variables relating to a woman's experience with the method used prior to the index method were considered: the reason for discontinuing the previous method and the duration of the last use interval. The reason for discontinuation was found to have the stronger impact. A variable with three categories was constructed to indicate the reason for stopping the previous method. Since for some women the index use is their first interval, or at least their first in the six-year observation window, a separate category was created for these intervals. The comparison of interest is between the remaining two categories which distinguish between nonmethod-related and method-related reasons for discontinuation for those women who have had a previous interval of use during the study period. Method-related reasons include contraceptive failure, side-effects, health concerns, inconvenience with use, husband's disapproval and a need for a more effective method.

Pill users who experienced method-related problems with their last method have an increased chance of discontinuing use while at risk of an unwanted pregnancy compared to those who have discontinued their last method for other reasons. Among users of modern methods other than the pill, the risk of moving to a traditional method is much higher if method-related problems were experienced with the previous method. Also, traditional users are less likely to switch to a modern method if they stopped their last method for method-related reasons. These findings suggest that current contraceptive behaviour is indeed linked to past experience and, in particular, an unfavourable experience with any modern method is more likely to lead to a switch to a less effective method or even non-use if problems were also encountered with the previous method. If a woman adopted traditional use after experiencing method-related problems with her last method, in most cases a modern method, she is less likely to switch back to a modern method.

Method

In the model for transitions from modern methods other than the pill, the origin method was included as a covariate to enable the calculation of method-specific transition rates. An interaction between method and duration of use was found to be significant. As shown in Figures 7.2, 7.3 and 7.4, the effect of duration on switching behaviour is different for condom users than for users of IUDs or injectables; discontinuation tends to occur much earlier for condom users while use of IUDs and injectables is more long-term. Condom users are the most likely to switch to non-use and not at risk, in most cases because they wish to become pregnant. IUD users are unlikely to discontinue for this reason, especially within a period of two years. Women using injectables are the most likely to switch to non-use while still in need of contraception. Condom users are around three times more likely to change to a traditional method than users of IUDs or injectables.

7.7.3 Interpretation of the Random Effects

Even after controlling for a wide range of demographic and socioeconomic background variables and characteristics of previous contraceptive experience, a large amount of variation in switching risks remains unexplained. Most of this unobserved heterogeneity is at the woman level as opposed to the cluster or district level. In particular, there is evidence of a large amount of variation in transitions to non-use while at risk of an unintended pregnancy and to modern or traditional methods. There is virtually no extravariation in transition rates from method use to non-use because of contraceptive failure or because the woman wants to have a child. Among pill users, there is some evidence of unexplained cluster-level variation in transition rates to non-use while at risk of unintended pregnancy, and to other modern methods. However, there is no significant variation between clusters in transitions from other types of modern method or from traditional methods. There is no evidence of district-level variation for any type of transition. For switches with significant extravariation at the woman level, random slopes models were fitted to test whether the effects of any of the covariates in the fixed part of the model varied randomly across women. In the model for transitions from pill use, additional random effects were added to the cluster level to test whether the coefficient of any covariate varied across clusters. However, there was no evidence of random coefficients at either the woman or the cluster level.

The large amount of unexplained variation at the woman level indicates that unobserved factors operating at the individual level have an important impact on switching behaviour. The lack of extravariation at higher levels suggests that the decision to switch for a contraceptive user is mainly an individual one and not heavily influenced by community-level factors. Whereas district-level characteristics were found to have an important impact on the decision to use versus not to use contraception (see Chapter 3), the results from the present analysis show that they have no influence on the contraceptive behaviour of acceptors.

Unobserved individual-level factors which may potentially influence switching behaviour include the level of commitment and motivation to use family planning, and the experience with the current and previous methods. Although the experience with the last method has been considered in the analysis, it is likely that the decision to switch is influenced by the experience with other methods used prior to the start of the six-year window about which we have no information. These factors are most likely to affect transitions from method use to non-use while at risk of an unwanted pregnancy or from modern method use to a less efficient traditional method. If a woman has a history of difficulty with use, this may reduce her motivation to use a modern method of contraception which may lead her to abandon use or opt for a traditional method. Another possible contributing factor is the level of a woman's mobility. If her mobility is low, then the choice of methods available to her may be restricted to those which can be delivered to her door. This is likely to have the greatest impact on transitions from the pill to other types of modern method.

The inter-cluster variation in the transition rates between the pill and other modern methods is most likely to be due to factors relating to the availability and quality of the family planning services in each community. The range of contraceptive options is likely to vary across clusters; while the pill is widely available in all areas, injectables are not always available and not all clinics perform IUD insertions. The rate of switching from pills to other types of modern method depends on the number of other modern methods available, and in some clusters there may be no other option. Even if there is a choice of methods, the clinic from which they can be obtained may not be easily accessible and, therefore, the only option is to continue with the pill, abandon use or switch to a traditional method.

The cluster-level variation in switching rates from the pill to non-use while still in need of contraception could also be explained by differences in family planning provision between clusters. Most pill users obtain their supplies from family planning workers who deliver direct to the door. In some areas, however, it is not possible for the outreach workers to visit every household regularly, particularly in very remote areas which are difficult to reach. Therefore, it is possible that some women may run out of pill supplies and, if no other option is available, may be forced to abandon use. Another possible explanation for the cluster-level variation in switches from the pill to non-use is differences in the attitudes of religious leaders or other influential figures in the community. It has been shown that religiosity influences the level of modern method use (see Chapter 3) and it is likely that it would continue to have an effect on continuation and switching behaviour among those who have accepted contraception. If there is a strong negative attitude towards family planning in the community, women may be more likely to abandon use.

In order to evaluate the impact of the unobserved woman- and cluster-level variation on transition rates, two-year cumulative switching probabilities have been calculated for selected types of transition for a range of values of the random effects. Table 7.14 presents probabilities of switching from the pill to a traditional method within the first two years of use by level of education. As in the interpretation of the fixed part of the model, all other covariates are held constant at the sample averages and the cluster- and district-level random effects are held at their average values of zero. Transition probabilities are then calculated for values of the woman-level random effect u_{4jkl} ranging from $-2\hat{\sigma}_{u4}$ and $+2\hat{\sigma}_{u4}$. It can be seen that the value of the random effect u_{4jkl}, representing unobserved factors operating at the woman level, has a considerable impact on the probability of switching from pill to traditional use. As noted in the discussion of the fixed effects, if all else is equal, women with secondary level education are almost twice as likely to switch to a traditional method as those with less education. However, if a woman with no education were to have a random effect value of $\hat{\sigma}_{u4}$, then she would have a higher chance of switching to a traditional method than a woman with at least secondary level education, but who has a random effect value of 0.

Table 7.14: 24-month cumulative probabilities of switching from the pill to a traditional method by education, for a range of values of the woman-level random effect u_{4jkl}

	Random effect u_{4jkl}				
Education	$-2\hat{\sigma}_{u4}$	$-\hat{\sigma}_{u4}$	0	$\hat{\sigma}_{u4}$	$2\hat{\sigma}_{u4}$
None	0.002	0.011	0.046	0.184	0.548
Primary	0.003	0.012	0.052	0.208	0.588
Secondary +	0.006	0.024	0.103	0.366	0.785

Tables 7.15 and 7.16 present probabilities of switching from a modern method other than the pill to non-use by type of method, and from a traditional to a modern method, by reason for stopping the last method. Again, the transition probabilities are greatly influenced by the value of the woman-level random effect. In both cases, the unobserved factors have more impact than the selected observed covariate.

Table 7.15: 24-month cumulative probabilities of switching from a modern method other than the pill to non-use and at risk by type of method, for a range of values of the woman-level random effect u_{2jkl}

	Random effect u_{2jkl}				
Method	$-2\hat{\sigma}_{u2}$	$-\hat{\sigma}_{u2}$	0	$\hat{\sigma}_{u2}$	$2\hat{\sigma}_{u2}$
Injection	0.092	0.164	0.279	0.441	0.627
IUD	0.060	0.108	0.192	0.322	0.501
Condom	0.051	0.093	0.165	0.278	0.435

Table 7.16: 24-month cumulative probabilities of switching from a traditional to a modern method by reason for stopping last method, for a range of values of the woman-level random effect u_{3jkl}

	Random effect u_{3jkl}				
Reason for stopping last method	$-2\hat{\sigma}_{u3}$	$-\hat{\sigma}_{u3}$	0	$\hat{\sigma}_{u3}$	$2\hat{\sigma}_{u3}$
No last method	0.042	0.095	0.204	0.401	0.657
Nonmethod-related	0.087	0.190	0.376	0.627	0.833
Method-related	0.060	0.133	0.278	0.508	0.753

Table 7.17: 24-month cumulative probabilities of switching from the pill to another modern method by type of region of residence, for a range of values of the cluster-level random effect v_{3jkl}

	Random effect v_{3jkl}				
Region and miles to thana HQ	$-2\hat{\sigma}_{v3}$	$-\hat{\sigma}_{v3}$	0	$\hat{\sigma}_{v3}$	$2\hat{\sigma}_{v3}$
Urban	0.092	0.136	0.196	0.277	0.380
Rural, < 5	0.085	0.125	0.182	0.259	0.357
Rural, ≥ 5	0.069	0.102	0.149	0.214	0.300

To illustrate the influence of the unobserved cluster-level factors on transitions from the pill, two-year cumulative probabilities of switching to another modern method have been calculated by type of region of residence, for a range of values of the cluster-level random effect v_{3jkl}. For an 'average' cluster with a random effect value of zero, a woman living in an urban area or less remote rural area has a greater probability of switching to another modern method than a similar woman living in a rural area, five miles or more from the *thana* headquarters. However, if the latter woman is living in an 'above average' cluster, corresponding to a positive value of v_{3jkl}, then she in fact has a greater chance of moving to another modern method than a woman living in an 'average' urban or less remote rural cluster.

7.8 Summary

The aim of the analysis presented in this chapter is to investigate changes in contraceptive behaviour over a period of six years, using data from a recent survey in Bangladesh. A series of multilevel competing risks models are used to study the factors affecting a woman's risk of switching from the pill, other modern methods and traditional methods. Of most interest to family planning policy makers is the degree of switching from efficient methods to inefficient methods or to non-use while the woman is at risk of an unintended pregnancy, as these types of method change are likely to have the most impact on fertility levels. The results show that the age at the start of use, number of children, education, religion and division of residence all have an influence on switching rates. In addition, there are rural-urban differentials and differences between remote and less remote rural areas. Among modern method users, there is also evidence that a bad experience with the previous method leads to an increased chance of switching from the current model to a traditional method or abandoning use altogether.

Although a large number of background characteristics were considered in the analysis, there are likely to be some important factors which have been omitted. For example, the level of social independence has been shown to influence contraceptive choice (Chapter 3) and it is likely that this would also have an effect on continuation of use. Although some indicators of women's status are available, they were measured only at the time of the survey and were, therefore, excluded from the longitudinal analysis. To allow for these unobserved factors, a woman-specific random effect was included in the model for each type of switch. Further random effects were added to test for extravariation at the sampling cluster and district level. This model can be extended straightforwardly to handle situations where repeated switches are made by the same woman by including an extra random effect corresponding to the use interval. In this case, however, it was found that after disaggregating by type of switch and method, very few multiple switches of the same type were observed during the six-year observation window.

The results show a large amount of unobserved heterogeneity between women in the risk of switching from a modern method to non-use while at risk of an unintended pregnancy, and to another modern or traditional method. Nearly all of the extravariation in switching risks is at the woman level as opposed to the cluster or district level. For pill users, there is some evidence of variation between clusters which may be due to differences in family planning service provision. Although, there are significant differences in switching rates between division, the multilevel analysis shows that there is no variation within divisions between districts.

Chapter 8

Summary and Discussion

8.1 Introduction

Most populations in demography and the other social sciences have a hierarchical structure. For example, individuals live in households which themselves lie within neighbourhoods or communities. If such hierarchies exist, it is likely that the attitudes and behaviour of individuals will be influenced not only by their own characteristics, but by those of other individuals in the group. Group members may also share common experiences which will lead responses for individuals in the same group to be more homogeneous than responses for individuals from different groups. For instance, in the area of child health, several studies have found evidence of familial death clustering associated with certain genetic characteristics and environmental conditions which are shared by children with the same mother. In the context of fertility and related behaviour, family planning programmes are usually implemented at the community level which means that couples in the same community will have access to similar services and range of method options. As a result, couples living in the same area may exhibit similar patterns of contraceptive use and levels of fertility.

Hierarchical structures can often be artefacts of survey design as in many surveys individuals are selecting using multistage sampling procedures. In such cases, the fundamental assumption of standard regression techniques that individuals are independently distributed will often be invalid as within sampling clusters the responses of individuals will tend to be correlated. Therefore, even if the hierarchical structure is not of substantive interest itself, complex survey data need to be analysed using methodology which takes into account intraclass correlation.

Another way in which hierarchically structured data may arise is in longitudinal studies. In general, measurements made on the same individual over time will tend to be correlated. This leads to a two-level structure in which observations at successive time points (level 1) are nested within an individual (level 2). An example of longitudinal data is an event history, where an individual is observed over time until he or she experiences the event of interest. If a discrete-time analysis approach is used, the data are expanded so that each individual's duration to event occurrence generates a series of binary or multinomial responses (depending on the number of events of interest), one for each discrete time point. The reconstructed data set, therefore, has a two-level nested structure as described above. Models for unobserved heterogeneity in individuals' risks, which include random effects at the individual level, can be conveniently estimated using existing multilevel software for hierarchical discrete response data.

In this thesis, the use of multilevel modelling techniques in demography was illustrated by a series of analyses of current status and longitudinal health and family planning data. The first analysis (Chapter 3) used a multilevel multinomial model to study variations in contraceptive choice in Bangladesh between districts, and within districts between clusters. The remaining three analytical chapters used multilevel event history methodology. In Chapter 5, a multilevel discrete-time hazards model was employed to study household- and village-level variation in child immunisation uptake in rural Bangladesh. The final two analytical chapters used a multilevel approach to look at two aspects of contraceptive use dynamics—discontinuation and switching. In the first of these studies, a two-level competing risks model was to used to explore the extent of unobserved heterogeneity in women's discontinuation risks in China. In the final analytical chapter, this was extended to a four-level model to examine switching between contraceptive methods in Bangladesh, and in particular the amount of unobserved heterogeneity between districts, within districts between clusters, and within clusters between women.

In this chapter, the main results from the four empirical studies are summarised, focusing on the interpretation of the multilevel random effects. This is followed by a discussion of the limitations in the availability of contextual data for longitudinal analyses and some of the practical computing aspects which need to be considered in the analysis of large hierarchical data sets. Finally, some ideas for further work, both on the substantive and methodological side, are suggested.

8.2 Summary of the Main Results

In Chapter 3, a multilevel multinomial model was used to analyse the factors affecting contraceptive choice in Bangladesh. The results reveal that even after controlling for a range of individual-level socioeconomic, demographic and cultural characteristics, there remain substantial variations in levels of use and method choice both between districts, and within districts between clusters. In an attempt to explain some of the inter-district variation, a series of district-level variables were added to the model. Two of these—religious practice (measured by the proportion of Muslims who pray every day) and literacy—were found to have a strong influence on contraceptive choice. A district with a high proportion of strict Muslims tends to have a low level of contraceptive use, while use of modern methods is generally higher in districts with high proportions of literate women. The addition of these two variables leads to a reduction of 56% in the district-level variation for the contrast between sterilisation and non-use, and a 33% reduction for the contrast between reversible modern method use and non-use. However, there remains considerable extravariation within districts between sampling clusters. A possible explanation for this is the omission of unobserved characteristics of family planning service provision which is likely to operate at a more local level than that of the district.

Continuing the theme of contraceptive use dynamics in Chapter 6, a two-level discrete-time competing risks model was developed to analyse the factors affecting contraceptive discontinuation in China. A competing risks framework was used to distinguish between three broad types of reasons for discontinuation: failure, and method-related and nonmethod-related reasons. Even if a range of background characteristics are considered in the analysis, there is likely to remain some risk factors which have not been included, and some of these may be unobservable. It is also possible that a woman will experience more than one discontinuation during the eight-year observation period; in the presence of unobserved heterogeneity, the durations of these repeated spells may be correlated. The discrete-time model can be formulated as a multilevel multinomial model and unobserved heterogeneity in women's discontinuation risks can be incorporated by including a woman-specific random effect for each type of discontinuation. The results show significant extravariation between women for the contrasts between failure and continuation of use and between other method-related discontinuations and continuation. These unobserved woman-level covariates have a strong impact on the risk of either type of discontinuation. Possible candidates for these unobserved variables include biological factors, such as fecundity which is likely to affect a woman's risk of failure, or characteristics of the method, such as IUD quality which would affect the risk of IUD expulsion, side-effects or any other method-related discontinuation.

In the final analytical chapter (Chapter 7), we returned to contraceptive use dynamics in Bangladesh with an analysis of switching patterns over a six-year period. A multilevel competing risks framework was used to analyse the risk of switching from a method to another method (either modern or traditional) or non-use while at risk of an unintended pregnancy. As well as controlling for unobserved heterogeneity between women in their risks of changing method, the two-level competing risks model used in Chapter 6 was extended to include random effects at two higher levels of aggregation: the cluster and the district. There are pronounced differences in switching behaviour between the four administrative divisions. In particular, the religiously conservative Chittagong division shows high rates of method abandonment and low rates of switching to modern methods. In contrast to the contraceptive choice analysis, however, there is no evidence of district-level variation for any method nor for any type of switch. Within districts, the only significant cluster-level variation is among pill users; there is some variation between clusters in the risk of switching from the pill to another modern method or to non-use while in need of contraceptive protection. A plausible explanation for this are differences in the range of method options made available by local family planning clinics. If a woman experiences problems with the pill and no suitable alternative is available, she may prefer to discontinue use altogether rather than continue pill use. Further, since many pill users rely on doorstep delivery services for their supplies, infrequent visits by family planning workers may force a woman to abandon use. However, the cluster-level variation is small relative to the woman-level variation. There is overwhelming evidence of unobserved heterogeneity between women in the risk of switching for all method types. The lack of variation at higher levels suggests that the decision to switch methods is largely made by the individual and is relatively unaffected by the behaviour of others in the community. While these factors may influence the initial decision to adopt family planning, the results from the switching analysis would indicate that these are unimportant among contraceptive users. Although a range of individual-characteristics were considered, including factors relating to the experience with the previous method, there remains substantial extravariation between women which has a strong impact on switching probabilities. Clearly, there are some, perhaps unobservable, individual-level factors other than those considered in the analysis which have an important influence on changes in contraceptive behaviour.

In recent years, there have been several studies which have used random effects logistic models to control for clustering in child mortality risks for children with the same mother. In Chapter 5, multilevel techniques were used to study the extent of clustering in another health outcome—child immunisation—using data collected in four rural areas of Bangladesh by Save the Children USA. Preliminary analysis shows that households who immunise one child have a tendency to immunise subsequent children too. A three-level discrete-time event history model was used to control for clustering in immunisation uptake not only at the household level, but at the village level. The analysis reveals a large amount of variation between villages in the most remote intervention area, where development programmes began 10 years later than in the other areas. One possible reason for this is accessibility; as some of the villages are cut-off during the monsoon and difficult to reach throughout the year, visits from health workers may be less frequent than in other villages. Other contributing factors may include vaccine shortages and the attitudes of influential community figures, such as religious leaders or school teachers. However, it was found that most of the unobserved variation could be attributed to the household level. Although a wide range of child-, mother- and household-level characteristics were considered, there remains substantial variation in immunisation uptake within villages between households. Possible explanations for this include differences in health care knowledge and in the attitudes of the persons making the decisions about children's welfare. The results show that if the child's mother is not the wife of the head of the household, she has a lower chance of immunising her child. Since these mothers tend to have lower status in the household, it is likely that decisions about her child's health care would be made by an older woman who may have more traditional views regarding immunisation.

To summarise, in all four studies the use of multilevel analysis techniques has revealed a considerable amount of extravariation at various levels of nesting. The analyses of contraceptive choice (Chapter 3), and child immunisation uptake (Chapter 5), provide strong evidence that individual behaviour is affected not only by individual characteristics but by unobserved factors operating at areal levels. The other two studies illustrated the use of multilevel models to control for unobserved heterogeneity in women's contraceptive behaviour over time. Both analyses reveal significant extravariation between women which in many cases has more of an impact on behaviour than the observed covariates considered in the models. In the remaining sections of this chapter, we look at some of the substantive and methodological questions raised by the empirical work. It is possible that some of these could be addressed through the addition of longitudinal contextual data and further analyses.

8.3 Contextual Data

In each of the analytical chapters, multilevel modelling has revealed a high degree of extravariation either at the individual level, in the form of unobserved heterogeneity in event history models, or at higher levels of aggregation. In most situations, it is inevitable that a certain amount of variation will remain unexplained since some variables will be unobservable. Although there will always be some unobservables, however, it is likely that other covariates can be observed, but are omitted because they have not been measured. In the study of contraceptive switching behaviour in Bangladesh (Chapter 7), very little of the community data could be used in the analysis since most of the detailed information on family planning service availability was measured only once at the time of the survey. Therefore, although the community survey is a rich source of data for an analysis of current-status data, the cluster-level variables could not be linked to the longitudinal individual-level data collected in the calendar. It is possible that some of the inter-cluster variation in switches from pill use could be reduced by the addition of time-dependent information such as the availability of modern methods other than the pill or the frequency of visits made by family planning workers. Such information is particularly useful for the evaluation of family planning and health programmes. For example, a fully longitudinal analysis of contraceptive use dynamics would allow one to test whether improvements in family planning service provision lead to changes in individual behaviour, in terms of increased levels of use, lower discontinuation rates and lower levels of switching from modern methods to inefficient methods.

The lack of longitudinal community-level data has been recognised by Demographic and Health Surveys (DHS) who are now conducting several surveys in the same clusters as earlier surveys. When this data become available, it will be possible to link the community data from both surveys to obtain longitudinal cluster-level data. DHS have also carried out an individual-level panel study in Morocco in 1995 in which individuals in half the clusters used in the 1992 survey were reinterviewed. Although no cluster-level data were collected in the 1995 survey, the 1992 cluster-level data can be linked to the 1995 calendar data.

In some situations, it may be difficult to collect quantitative data which truly captures the unobserved characteristics. For example, in the analysis of child immunisation uptake in Chapter 5, several villages in the most recently intervened area of Nasirnagar were identified as having particularly high or low immunisation rates. Fieldworkers in the area suggested a number of possible explanations for this village-level variation. For example, it was suggested that the high immunisation rate in one village could be due in part to the co-operation of a local school teacher who encouraged parents to immunise their children. In another village, it is possible that the low rate of immunisation uptake may be explained by the negative attitude of a religious leader towards modern health care. It would be difficult to obtain a quantitative measure that reflected the attitudes of these influential community figures, especially since there may have been changes over the observation period of more than five years. In this case, although multilevel analysis has enabled us to identify those villages with significantly higher or lower immunisation coverage, qualitative research, perhaps in the form of in-depth case studies of the 'outlying' villages, would probably be required to uncover the main influences.

8.4 Suggestions for Further Work

8.4.1 Substantive Issues

Clearly, each of the four analyses presented in this thesis could be extended in a number of ways depending on the substantive questions of interest. One obvious area for further research would be to use the methodology described in Chapter 6 to analyse contraceptive discontinuation in Bangladesh. In the analysis of contraceptive switching, the main focus of interest was the factors affecting the risk of changing methods among women who had a need for contraception. Therefore, discontinuations due to failure or because the woman wanted to become pregnant were classified as 'not at risk' and were grouped together in the same response category. In particular, it would be interesting to examine whether there is clustering in discontinuation patterns at the cluster- and district-level. One might expect factors relating to the availability and quality of family planning services to have an impact on failure rates and on discontinuation rates for other method-related reasons. Alternatively, we may find that, as for switching behaviour, the risk of discontinuation is affected mainly by individual characteristics.

On a similar theme, another possible area for further research would be to extend the analysis of discontinuation in China to look at variation between provinces and, within provinces, between clusters. The work in Chapter 6 was carried out in 1993 using the precursor to MLn, ML3. As the name suggests, this earlier version could analyse hierarchical structures with up to three levels. Since the multivariate data structure required to fit a multilevel multinomial model (see Section 2.8.2) means that a two-level structure is analysed using a three-level model, random effects could only be fitted at the woman level. With the development of MLn, it is now possible to model structures with

5 levels or more. However, extending the analysis from four provinces to all 31 provinces would lead to a huge data set. Even a 10% sample would contain around 50 000 women. Given that many of these women will contribute more than one spell of contraceptive use and that the data need to be expanded considerably before a multilevel competing risks model can be fitted, it is infeasible to analyse such a large sample. It would be necessary to use a sample fraction of around 1%, although this would depend on the width of the interval between discrete-time points and the number of competing risks.

Finally, it would be interesting to extend the analysis of immunisation uptake in rural Bangladesh (Chapter 5) to examine the amount of clustering at the household and village level in other health outcomes. For example, Save the Children have also collected data on mortality, antenatal care and, as part of a growth monitoring programme, indices of children's nutritional status. Pebley et al. (1996) examined the extent of clustering at the family and community level in prenatal and delivery care as well as childhood immunisation in Guatemala. They found that there was less clustering at the community level for immunisation than for the other health outcomes, which they suggest is due to the nature of the immunisation programme; while EPI is nationwide in coverage, other health initiatives tend to be on a much smaller scale which means that there is more scope for variability in their implementation between areas than for the large-scale immunisation campaigns. Multilevel techniques could be used to determine whether this is also the case for the four study areas in Bangladesh.

8.4.2 Methodological Issues

There a number of methodological issues concerning the use of multilevel models and their estimation which warrant further investigation. An important issue, which has come to light relatively recently, is the adequacy of the estimation procedures currently available for the estimation of clustered binary response data. Since the simulation study of Rodriguez and Goldman (1995), which found that both ML3 and VARCL can severely underestimate random effect variances, the procedure used by ML3, marginal quasi-likelihood (MQL), has been improved. Further simulation studies have shown that a new estimation procedure, penalised quasi-likelihood (PQL), yields estimates of the variance components which are very close to their 'true' values (Goldstein 1985) and has since been implemented in MLn. However, Goldstein (1995) remarks that in many situations the MQL estimates will be adequate and it is not clear under what circumstances MQL will underestimate. Rodriguez and Goldman (1995) suggest that MQL will produce underestimates when cluster size is small and when the random effect is 'large enough to be interesting'. Another situation in which MQL is thought to perform poorly is when the response probabilities are extreme. In each of the analyses presented in this thesis, a first-order MQL approximation was fitted first and the estimates obtained were used for starting values for the first-order PQL approximation. Finally, these were extended to a second-order PQL model. In each case, the first-order MQL estimates were compared with those obtained from the improved second-order PQL approximation. In general, the results for the two models were very close. This would suggest that, in the cases considered here, the MQL approximation is adequate. This is in spite of the fact that the response probabilities in event history models tend to be very small (since we are modelling the probability of an event occurring in a small interval of time) and that in some cases, particularly in the analysis of contraceptive switching, the random effect variances were large. Clearly more research, in the form of simulation studies, is needed in this area so as to advise practitioners on the estimation procedure most appropriate for their application. Although, if using MLn, the best option would be to fit both MQL and PQL models, it is not always possible to achieve convergence using the second-order PQL approximation (Goldstein 1995). In such cases the iterative bootstrap (Section 2.7.3) and Gibbs sampling (Section 2.7.7) are attractive alternatives. However, they are highly computer-intensive which may discourage some researchers with large data sets from using them.

Another methodological issue, raised in the analysis of immunisation uptake in Chapter 5, concerns the number of clusters with more than one observation that are required to enable a multilevel model to be estimated. Multilevel models have been used in several contexts where cluster sizes are very small and the proportion of clusters with only one observation is high, for example in studies of familial clustering in child mortality. In the immunisation study, a model was initially fitted in which there were separate random effects for the mother and for the household. After controlling for variation at the household level, the variation at the mother level was estimated very close to zero. This could indicate that there is no variation between mothers in the same household. However, since only 10% of households contain more than one mother, it seems likely that in this case there is a problem of power as there is not enough information to disentangle the mother and household effects. Some practical guidelines on this issue would be useful.

The final methodological point concerns the estimation of multilevel multinomial models. Currently, multilevel multinomial models can only be fitted in MLn. However, the approach used by MLn can quickly become impractical for large data sets, especially if there are a large number of response categories. In Section 2.8.2, the reconstruction of the data set required to fit the multinomial model was described. Basically, the data need to be restructured so that each multinomial response generates a series of binary responses, one for each of the response categories minus one. In addition, a new set of covariates need to be created, one for each contrast of a response category with the baseline, to allow a separate set of parameters to be estimated for each contrast. This enables the model to be fitted as a special case of a multivariate response model. However, although it is reasonably straightforward to restructure the data into this form using MLn commands, it is extremely time consuming, and more importantly can lead to the creation of very large data sets. To be specific, if the original data set contains n cases and p covariates and the multinomial response has s categories, the generated data matrix will be of dimension $(s-1)n \times (s-1)p$. If the original data set is large, a multinomial analysis is only possible on a PC with a large amount of RAM or on a Unix platform. Fortunately, it was possible to carry out the analyses presented in this thesis using a test versions of ML3 and MLn which have been compiled on a Sun Sparc station and a Cray supercomputer respectively. Clearly, however, these options are not available to many researchers. It would be useful to investigate whether a more practical approach to model estimation is possible.

References

Aitkin, M. and Longford, N. (1986) Statistical modelling issues in school effectiveness studies (with discussion). *Journal of the Royal Statistical Society*, A, 149, 1-43.

Ali, K. (1993) Antenatal care in Bangladesh: evidence from the Bangladesh Fertility Survey 1989 data. Working paper, Bangladesh Institute of Development Studies.

Ali, M. and Cleland, J. (1996) Contraceptive discontinuation in six developing countries—a cause-specific analysis. *International Family Planning Perspectives*, 21, 92-97.

Allison, P.D. (1982) Discrete-time methods for the analysis of event history data. In *Sociological Methodology* (ed. S. Leinhardt). San Francisco: Jossey-Bass.

Allison, P.D. (1987) Introducing a disturbance into logit and probit regression models. *Sociological Methods and Research*, 15, 355-374.

Alker, H.S. (1969) A typology of ecological fallacies. In *Quantitative Ecological Analysis* (ed. M. Dogan and S. Rokkan). Mass: MIT Press.

Amin, S. (1993) The Expanded Programme on Immunisation in Bangladesh. Paper presented at the IUSSP Seminar on the Evaluation of Health Interventions, Belo Horizonte, Brazil.

Amin, S., Ali, K. and Streatfield, K. (1993) Vaccine uptake in Bangladesh: an analysis of the Expanded Program on Immunization. In *Secondary Analysis of the Bangladesh Fertility Survey* (ed. J.C. Cleland et al.). Dhaka: National Institute of Population Research and Training (NIPORT).

Amin, S. and Hossain, S. (1995) Women's reproductive rights and the politics of fundamentalism: A view from Bangladesh. *The American University Law Review*, 44, 1319-1344.

Andersen, E.B. (1990) *The Statistical Analysis of Categorical Data*. Berlin: Springer-Verlag.

Anderson, D.A. and Aitkin, M. (1985) Variance component models with binary response: interviewer variability. *Journal of the Royal Statistical Society*, B, 47, 203-210.

Barry, J.T., Francis, B.J. and Davies, R.B. (1989) *SABRE Users Guide.* Research Report 89-02, Centre for Applied Statistics, Lancaster University.

Bartholomew, D. (1987) *Latent Variable Models and Factor Analysis.* London: Griffin.

Becker, S. and Sosa, D. (1992) An experiment using a month-by-month calendar in a family planning survey in Costa Rica. *Studies in Family Planning*, 23, 386-391.

Begg, C.B. and Gray, R. (1984) Calculation of polychotomous logistic regression parameters using individualised regressions. *Biometrika*, 71, 11-18.

Ben-Akiva, M. and Lerman, S. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand.* Cambridge: MIT press.

Bilsborrow, R.E. (1985) Collecting community-level data for fertility analysis. In *The Collection and Analysis of Community Data* (ed. J. Casterline). Voorburg, Netherlands: International Statistical Institute.

Blanchet, T. (1991) An anthropological study of maternal nutrition and birth practices in Nasirnagar, Bangladesh. Report, Save the Children USA, Dhaka.

Blossfeld, H. and Hamerle, A. (1989) Unobserved heterogeneity in hazard rate models: a test and an illustration from a study of career mobility. *Quality and Quantity*, 23, 129-141.

Bongaarts, J. and Potter, R.G. (1983) *Fertility, Biology and Behaviour: An Analysis of the Proximate Determinants.* New York: Academic Press.

Bongaarts, J. and Rodriguez, G. (1990) A new method for estimation of contraceptive failure rates. In *Measuring the Dynamics of Contraceptive Use: Proceedings of the United Nations Expert Group Meeting.* New York: United Nations.

Bracher, M. and Santow, G. (1992) Premature discontinuation of contraception in Australia. *Family Planning Perspectives*, 24, 58-65.

Breslow, N.E. and Clayton, D.G. (1993) Approximate inference in generalized linear models. *Journal of the American Statistical Association*, 88, 9-25.

Bryk, A.S., Raudenbush, S.W., Seltzer, M. and Congdon, R. (1988) *An Introduction to HLM: Computer Program and User's Guide* (2nd edition). Chicago: University of Chicago.

Bryk, A.S and Raudenbush, S.W. (1992) *Hierarchical Linear Models: Applications and Data-analysis Methods.* Newbury Park: Sage.

Caldwell, J.C. (1980) Mass education as a determinant of the timing of the fertility decline. *Population and Development Review*, 6, 225-255.

Casterline, J. (1985) Community effects on fertility. In *The Collection and Analysis of Community Data* (ed. J. Casterline). Voorburg, Netherlands: International Statistical Institute.

Casterline, J. (1987) The collection and analysis of community data: the WFS experience. In *The World Fertility Survey: an Assessment* (ed. J. Cleland and C. Scott). Oxford: Oxford University Press.

Choe, M.K. and Zablan, Z.C. (1991) Contraceptive discontinuation and failure rates in the Philippines: estimates from the 1986 contraceptive prevalence survey. In *Measuring the Dynamics of Contraceptive Use: Proceedings of the United Nations Expert Group Meeting.* New York: United Nations.

Cleland, J. and Mauldin, W.P. (1991) The promotion of family planning by financial payments: the case of Bangladesh. *Studies in Family Planning*, 22, 1-18.

Cleland, J. and Streatfield, K. (1992) The demographic transition: Bangladesh. Technical Report, Staff Reference Series no. 1/92, Programme Planning Unit, Dhaka: UNICEF.

Cleland, J., Phillips, J.F., Amin, S. and Kamal, G.M. (1994) *The Determinants of Reproductive Change in Bangladesh: Success in a Challenging Environment.* Washington, DC: The World Bank.

Conaway, M.R. (1990) A random effects model for binary data. *Biometrics*, 46, 317-328.

Cox, D.R. (1972) Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, B, 34, 187-220.

Curtis, S.L., Diamond, I. and McDonald, J.W. (1991) Birth interval effects and healthy families in Brazil. In *Proceedings of the Demographic and Health Surveys World Conference Proceedings*, vol. II, (ed. IRD/Macro International inc.).

Curtis, S.L., Diamond, I. and McDonald, J.W. (1993) Birth interval and family effects on postneonatal mortality in Brazil. *Demography*, 30, 33-43.

Curtis, S.L. and Hammerslough, C.R. (1995) *Model Further Analysis Plan: Contraceptive Use Dynamics.* Calverton, Maryland: Macro International Inc..

Curtis, S.L. and Steele, F. (1996) Variation in familial neonatal mortality risks in four countries. *Journal of Biosocial Science*, 27, 95-106.

Das Gupta, M. (1987) Selective discrimination against female children in rural Punjab, India. *Population and Development Review*, 13, 77-100.

Das Gupta, M. (1990) Death clustering, mother's education and the determinants of child mortality in rural Punjab, India. *Population Studies*, 44, 489-505.

DaVanzo, J., Rebouisson, D., Starbird, E., Tan, B.A. and Hadi, S.A. (1989) Contraceptive method switching over women's reproductive careers. *Journal of Biosocial Science* II (Suppl.), 95-116.

Davies, R.B., Elias, P. and Penn, R. (1992) The relationship between a husband's unemployment and his wife's participation in the labour-force. *Oxford Bulletin of Economics and Statistics*, 54, 145-171.

Delfs, R. (1990) Population growth threatens to get out of control: the fertility factor. *Far Eastern Economic Review*, July 19.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) A maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*, B, 76, 341-353.

Dempster, A.P., Rubin, R.B. and Tsutakawa, R.K. (1981) Estimation in covariance components models. *Journal of the American Statistical Association*, 76, 341-353.

Diamond, I., McDonald, J.W. and Shah, I.H. (1986) Proportional hazards models for current status data: application to the study of differentials in weaning in Pakistan. *Demography*, 23, 607-20.

Efron, B. and Tibshirani, R.J. (1993) *An Introduction to the Bootstrap*. London: Chapman and Hall.

Egger, P.J. (1992) Event History Analysis: Discrete-Time Models including Unobserved Heterogeneity with Applications to Birth History Data. PhD thesis, Department of Social Statistics, University of Southampton.

Elias, P. (1994) Job-related training, trade union membership, and labor mobility—a longitudinal study. *Oxford Economic Papers—New Series*, 46, 563-578.

Enberg, J., Gottscalk, P. and Wolf, D. (1990) A random-effects logit model of work-welfare transitions. *Journal of Econometrics*, 43, 63-75.

Entwisle, B, Hermalin, A.I., Kamnuansilpa, P., Chamratrithirong, A. (1984) A multilevel model of family planning availability and contraceptive use in rural Thailand. *Demography*, 21, 559-574.

Entwisle, B., Mason, W.M. and Hermalin, A.I. (1986) The multilevel dependence of contraceptive use on socioeconomic development and family planning program strength. *Demography*, 23, 199-216.

Entwisle, B., Casterline, J.B. and Sayed, H.A.A. (1989) Villages as contexts for contraceptive behaviour in rural Egypt. *American Sociological Review*, 54, 1019-1034.

Foster, A.D. (1993) Household partition in rural Bangladesh. *Population Studies*, 47, 97-114.

Freedman, R. (1985) Summary observations on the WFS seminar on collection and analysis of data on community and institutional factors. In *The Collection and Analysis of Community Data* (ed. J. Casterline). Voorburg, Netherlands: International Statistical Institute.

Gelfand, A.E. and Smith, A.F.M. (1990) Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 86, 79-86.

Goldman, N., Moreno, L. and Westoff, C.F. (1989) *Peru Experimental Study: An Evaluation of Fertility and Child Health Information.* Columbia, Maryland: Office of Population Research, Princeton University, Princeton, New Jersey and Institute for Resource Development (IRD)/Macro Systems Inc.

Goldstein, H. (1986) Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73, 43-56.

Goldstein, H. (1987) *Multilevel Models in Educational and Social Research.* New York: Oxford University Press.

Goldstein, H. (1989) Restricted unbiased IGLS estimation. *Biometrika*, 76, 622-23.

Goldstein, H. (1991) Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, 78, 45-51.

Goldstein, H. (1995) *Multilevel Statistical Models* (2nd edition). London: Edward Arnold.

Goldstein, H. (1996) Consistent estimators for multilevel generalized linear models using an iterated bootstrap. *Multilevel Modelling Newsletter*, vol.8, no.1, Institute of Education, London.

Goldstein, H. and Healy, M.J.R. (1995) The graphical presentation of a collection of means. *Journal of the Royal Statistical Society*, A, 158,175-177.

Goldstein, H. and Rasbash, J. (1995) Improved approximations for multilevel models with binary responses. Working paper, Institute of Education, London.

Goldstein, H. and Spiegelhalter, D.J. (1996) League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society*, A, 159 (to appear).

Grady, W.R., Hirsch, M.B., Keen, N. and Vaughan, B. (1983) Contraceptive failure and continuation among married women in the US, 1970-75. *Studies in Family Planning*, 14, 9-19.

Grady, W.R., Hayward, M.D. and Florey, F.A. (1988) Contraceptive discontinuation among married women in the US. *Studies in Family Planning*, 19, 227-235.

Grady, W., Hayward, M., Billy, J. and Florey, F. (1989) Contraceptive switching among currently married women in the United States. *Journal of Biosocial Science* II (Suppl.), 119-132.

Greenhaulgh, S. (1986) Shifts in China's population policy 1984-86: views from the central, provincial and local levels. *Population and Development Review*, 12, 491-515.

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

Guo, G. (1993) Use of sibling data to estimate family mortality effects in Guatemala. *Demography*, 30, 15-32.

Hamill, D.N., Tsui, A.O. and Thapa, S. (1990) Determinants of contraceptive switching behaviour in rural Sri Lanka. *Demography*, 27, 559-578.

Hammerslough, C. (1984) Characteristics of women who stop using contraceptives. *Family Planning Perspectives*, 16, 14-18.

Hardee-Cleaveland, K. and Bannister, J. (1988) Fertility and policy implementation in China 1986-88. *Population and Development Review*, 14, 245-286.

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

Hermalin, A. (1979) Multivariate areal analysis. In United Nations, Department of International Economic and Social Affairs, Population Division, *Manual IX: The Methodology of Measuring the Impact of Family Planning Programmes on Fertility.* New York: United Nations. (Population studies no. 66).

Hill, D.H., Axinn, W.G. and Thornton, A. (1993) Competing hazards with shared unmeasured risk factors. *Sociological Methodology*, 23, 245-277.

Holt, D. (1985) Cross-level analyses: discussion. In *The Collection and Analysis of Community Data* (ed. J. Casterline). Voorburg, Netherlands: International Statistical Institute.

Huq, N.M. and Cleland, J. (1990) *Bangladesh Fertility Survey 1989.* Main report. Dhaka: National Institute of Population Research and Training (NIPORT).

Huq, M. (1991) (ed.) *Near Miracle in Bangladesh.* Dhaka: University Press.

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

Islam, M.A. (1994) Multistate survival models for transitions and reverse transitions—an application to contraceptive use data. *Journal of the Royal Statistical Society*, A, 157, 441-455.

Jejeebhoy, S. (1990) Measuring contraceptive use-failure and continuation: an overview of new approaches. In *Measuring the Dynamics of Contraceptive Use: Proceedings of the United Nations Expert Group Meeting.* New York: United Nations.

Kalbfleisch, J.D. and Prentice, R.L. (1980) *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Kamal, N. (1995a) Areal variations in use of modern contraceptives in rural Bangladesh. PhD thesis, London School of Hygiene and Tropical Medicine.

Kamal, N. (1995b) The role of government family planning workers and health centres as determinants of contraceptive use in Bangladesh. Paper presented at the Annual Meeting of the British Society of Population Studies, Durham.

Kamal, N. and Sloggett, A. (1993) The influence of religiosity, mobility and decision making on contraceptive use. In *Secondary Analysis of the Bangladesh Fertility Survey Data* (ed. J.C. Cleland et al.). Dhaka: National Institute of Population Research and Training (NIPORT)

Koenig, M.A. and Simmons, R. (1992) Constraints on supply and demand for family planning: evidence from rural Bangladesh. In *Family Planning Programmes and Fertility* (ed. J.F. Phillips and J.A. Ross). Oxford: Oxford University Press.

Kost, K. (1993) The dynamics of contraceptive use in Peru. *Studies in Family Planning*, 24, 109-119.

Kuk, A.Y.C. (1995) Asymptotically unbiased estimation in generalized linear models with random effects. *Journal of the Royal Statistical Society*, B, 57, 395-407.

Liang, K. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 45-51.

Longford, N.T. (1987) A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74, 817-827.

Longford, N.T. (1988) *VARCL - Software for Variance Components Analysis of Data with Hierarchically Nested Random Effects (Maximum Likelihood).* Princeton, New Jersey: Educational Testing Service.

Longford, N.T. (1994) Logistic regression with random coefficients. *Journal of Computational Statistics and Data Analysis*, 17, 1-15.

McCullagh, P. and Nelder, N.A. (1989) *Generalized Linear Models* (2nd edition). London: Chapman and Hall.

Maddala, G.S. (1988) *Introduction to Econometrics.* New York: Macmillan.

Madise, N. and Diamond, I. (1995) Determinants of infant mortality in Malawi—an analysis to control for death clustering within families. *Journal of Biosocial Science*, 27, 95-106.

Manton, K.G., Stallard, E. and Vaupel, J.W. (1986) Alternative models for the heterogeneity of mortality risks among the aged. *Journal of the American Statistical Association*, 81, 635-644.

Mason, W.M., Wong, G.Y. and Entwisle, B. (1983) Contextual Analysis through the multilevel linear model. In *Sociological Methodology* (ed. S. Leinhardt). San Francisco: Jossey-Bass.

Mauritsen, R.H. (1984) Logistic Regression with Random Effects. PhD thesis, Department of Biostatistics, University of Washington.

Mitra, S.N., Ali, M.N., Islam, S., Cross, A.R. and Saha, T. (1994) *Bangladesh Demographic and Health Survey 1993-1994*. Calverton, Maryland: National Institute of Population Research and Training (NIPORT), Mitra and Associates, and Macro International Inc.

Mitra, S.N. and Sabir, G. (1996) Contraceptive use dynamics in Bangladesh: discontinuation, switching and failure. In *Secondary Analysis of the Bangladesh and Demographic Health Survey*. Dhaka: NIPORT (to appear).

Moreno, L. and Goldman, N. (1991) Contraceptive failure rates in developing countries: evidence from the Demographic and Health Survey. *Family Planning Perspectives*, 17, 44-49.

Mussallam, B.F. (1983) *Sex and Society in Islam: Birth Control before the Nineteenth Century*. New York: Cambridge University Press.

Narendranathan, W. and Elias, P. (1993) Influences of past history on the incidence of youth unemployment: empirical findings for the UK. *Oxford Bulletin of Economics and Statistics*, 55, 161-186.

Narendranathan, W. and Stewart, M.B. (1993) How does benefit effect vary as unemployment spells lengthen? *Journal of Applied Econometrics*, 8, 361-381.

Neuhaus, J.M., Hauck, W.W. and Kalbfleisch, J.D. (1992) The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*, 79, 755-762.

Notestein, F.W. (1945) Population—the long view. In *Food for the World* (ed. T. Schultz). Chicago, University of Chicago Press.

Pebley, A.R. and Stupp, P.W. (1987) Reproductive patterns and child mortality in Guatemala. *Demography*, 24, 43-60.

Pebley, A.R. and Amin, S. (1991) The impact of public health interventions on sex differentials in childhood mortality in rural Punjab, India. Working paper no. 24, the Population Council.

Pebley, A.R., Goldman, N. and Rodriguez, G. (1996) Prenatal and delivery care and childhood immunization in Guatemala—Do family and community matter? *Demography*, 33, 231-247.

Peng, X. (1991) *Demographic Transition in China: Fertility Trends since the 1950s.* Oxford: Clarendon Press.

Poston, D.L., Jr (1986) Patterns of contraceptive use in China. *Studies in Family Planning*, 17, 217-227.

Poston, D.L., Jr (1992) Epilogue: China in 1990. In *The Population of Modern China* (ed. D.L. Poston and D. Yaukey). New York: Plenum Press.

Poston, D.L., Jr and Yu, M. (1985) Quality of life, intellectual development and behavioural characteristics of single children in China: evidence from a 1980 survey in Changsha, Hanou province. *Journal of Biosocial Science*, 17, 127-136.

Rahman, M.M., Haaga, J.G., Maru, R.M. and Hossain, M.B. (1993) Access of rural women to family planning service: a study in Bangladesh. Paper presented at Annual Meeting of the Population Association of America, Cincinnati.

Rasbash, J. and Woodhouse (1995) *MLn Command Reference.* London: Institute of Education.

Rashid, M.A. and Ali, M.N. (1993) Determinants of low contraceptive use in Chittagong division. In *Secondary Analysis of the Bangladesh Fertility Survey Data* (ed. J.C. Cleland et al.). Dhaka: National Institute of Population Research and Training (NIPORT)

Raudenbush, S.W. and Bryk, A.S. (1986) A hierarchical model for studying school effects. *Sociology for Education*, 59, 1-17.

Robinson, W.S. (1950) Ecological correlations and the behaviour of individuals. *American Sociological Review*, 15, 351-357.

Rodriguez, G. and Goldman, N. (1995) An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society*, A, 158, 73-89.

Samosir, O.B. (1994) Contraceptive use in Indonesia. PhD thesis, Department of Social Statistics, University of Southampton.

Sastry, N. (1996) Community characteristics, individual and household attributes, and child survival in Brazil. *Demography*, 33, 211-229.

Schirm, A., Trussell, J., Menken, J. and Grady, W.R. (1982) Contraceptive failure in the US: the impact of social, economic and demographic factors. *Family Planning Perspectives*, 14, 68-80.

Shahidullah, M. and Chakraborty, N. (1993) Factors affecting the use of contraception in Bangladesh: A multivariate analysis. *Asia-Pacific Population Journal*, 8, 19-30.

Sinha, D., Tanner, M.A. and Hall, W.J. (1994) Maximisation of the marginal likelihood of grouped survival data. *Biometrika*, 81, 53-60.

Skinner, C.J., Holt, D. and Smith, A.M.F. (1989) *Analysis of Complex Surveys*. Chichester: Wiley.

Spiegelhalter, D., Thomas, A., Best, N. and Gilks, W. (1994) *BUGS (Bayesian inference Using Gibbs Sampling) Version 0.030 Manual*. Cambridge: MRC Biostatistics Unit.

Statistics and Epidemiology Research Corporation (1991) *EGRET: Reference Manual and Addendum.* Seattle: SERC.

Stiratelli, R., Laird, N. and Ware, J.H. (1984) Random-effects models for serial observations with binary response. *Biometrics*, 40, 961-971.

Streatfield, K., Singarimbun, M. and Diamond, I. (1990) Maternal education and child immunisation. *Demography*, 27, 447-455.

Suchedina, Z. (1990) Islam, procreation and the law. *International Family Planning Perspectives*, 16, 107-11.

Thisted, R.A. (1988) *Elements of Statistical Computing: Numerical Computation.* New York: Chapman and Hall.

Tien, H.Y., Zhang, T., Ping, Y., Li, J. and Liang, Z. (1992) China's demographic dilemmas. *Population Bulletin*, 47.

Trussell, J. and Kost, K. (1987) Contraceptive failure in the United States: a critical review of the literature. *Family Planning Perspectives*, 14, 68-80.

Tsui, A.O., Thapa, S., Hamill, D. and DeSilva, V. (1989) Contraceptive method change among rural Sri Lankan women. *Journal of Biosocial Science* II (Suppl.), 133-148.

Tutz, G. (1995) Competing risks models in discrete time with nominal or ordinal categories of response. *Quality and Quantity*, 29, 405-420.

United Nations (1989) *Case Studies in Population Policy: China.* Population policy paper no. 20.

Van de Pol, F., Langeheine, R. and de Jong, W. (1991) *PANMARK User Manual: Panel Analysis using Markov Chains, Version 2.2.* Netherlands Central Bureau of Statistics, Department of Statistical Methods, Voorburg.

Vaupel, J.W., Manton, K.G. and Stallard, E. (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16, 439-454.

Wang, D.L. (1994) Contraceptive use dynamics in China. PhD thesis, Department of Social Statistics, University of Southampton.

Wang, D.L. and Diamond, I. (1995) The impact on fertility of contraceptive failure in China in the 1980s. *Journal of Biosocial Science*, 27, 277-284.

Westoff, C.F., Goldman, N. and Moreno, L. (1990) *Dominican Republic Experimental Study: An Evaluation of Fertility and Child Health Information.* Columbia, Maryland: Office of Population, Princeton University, Princeton, New Jersey and IRD/Macro Systems Inc..

Williams, D.A. (1982) Extra-binomial variation in logistic linear models. *Journal of the Royal Statistical Society*, C, 31, 144-148.

Wong, G.Y. and Mason, W.M. (1985) The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association*, 80, 513-524.

Yamaguchi, K. (1991) *Event History Analysis.* Newbury Park, CA: Sage.

Zeger, S.L., Liang, K.Y. and Albert, P.S. (1988) Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 44, 1049-1060.

Zeger, S.L. and Karim, M.R. (1991) Generalized linear models with random effects: a Gibbs Sampling approach. *Journal of the American Statistical Association*, 86, 79-86.

Zenger, E. (1993) Siblings neonatal-mortality risks and birth spacing in Bangladesh. *Demography*, 30, 477-488.