January, 1987

ANALYSIS OF CATEGORICAL DATA ON PREGNANCY OUTCOME

by

R M PICKERING

being a thesis submitted for the degree of Doctor of Philosophy of the University of Glasgow

ERRATA

Page 63,	Eqn 1	$f_Y(y; \theta, \phi)$	ϕ not Φ
Page 68,	Eqn 2	$W_i = \{a_i(\phi)\, b''(\theta_i)\}^{-1} \left\{\frac{d\mu}{d\eta}\right\}_i^2$	
Page 64,	for Normal	$c(\cdot,\cdot) = -\frac{1}{2}\left\{\frac{y^2}{\phi} + \ln(2\pi\phi)\right\}$	
Page 70,	Eqn 1	$D(y;\hat{\mu}^{(2)})/\phi - D(y;\hat{\mu}^{(1)})/\phi$	
Page 71,	line 6	$g^*(\mu)$	
Page 79,	Eqn 4	$a_1 - a_2 = \psi^{-1}(p_{2j}) - \psi^{-1}(p_{1j})$	$1 \le j < k$
	Eqn 5	$1 - F(\theta) = \left\{1 + (1/u)\exp\left[-\left(\frac{m(\theta) - a}{b}\right)\right]\right\}^{-u}$	$(u>0)$
Page 80,	Eqn 1	$\psi^{-1} = -\ln(u) + \text{logit}\{(1-p_j)^{1/u}\}$	$1 \le j < k$
Page 93,	line 17	$\beta_{0j} = \beta^*_{0j} + \log(\pi_j/\pi_k)$	
Page 96,	Eqn 2	$l_P(\theta) = \frac{(T\theta)^d e^{-\theta T}}{d!}$	$\propto l(\theta)$
Page 181,	Eqn 2	$\ln(b_1) - \ln(b_2) = \ln\{\text{logit}(p_{2j}) / \text{logit}(p_{1j})\}$	
Page 174,	line 12	decreased not increased	
Page 179,	line 10	exclude marital status	
Page 181,	Eqn 3	$p_{2j} = \frac{p_{1j}^{\exp(\Delta)}}{(1-p_{1j})^{\exp(\Delta)} + p_{1j}^{\exp(\Delta)}}$	

Page 183, Difference between logits for ≥ 3 and 0-2 previous livebirths

CONTENTS

PAGE

PAGE

PAGE

LIST OF FIGURES

List of Tables

PAGE

PAGE

PAGE

ACKNOWLEDGEMENTS

My thanks are due to Dr GD Murray for his advice and encouragement throughout the period of my research. I am also indebted to the members of the Social Paediatric and Obstetric Research Unit who have provided a supportive background during my study. The data which made the analyses possible were kindly supplied by Dr SK Cole, Information Services Division, Common Services Agency, Edinburgh. The Social Paediatric and Obstetric Research Unit is funded by the Scottish Home and Health Department and the Greater Glasgow Health Board.

DECLARATION

The results of Chapters 4 and 5 have been published, in part, by Pickering & Forbes (1984), Pickering, Murray & Forbes (1986) and Pickering (1987a). A description of the results concerning low birthweight have been submitted for publication to 'The Journal of Epidemiology and Community Health' (Pickering, 1987b).

SUMMARY

In this thesis we examine the risk of low birthweight, preterm delivery and changes in the distribution of birthweight standardised for gestational age related to ten maternal, socio-demographic and biological covariates available from routinely collected maternity discharge records. The information is considered in the form of categorised variables and is analysed using regression models adapted for use with categorical data. A second part of the thesis describes a classification of neonates based on a latent class model for eleven variables measured during the neonatal period.

Chapter 1 contains a review of the epidemiology of birthweight and gestational age, focusing on their relationship with the ten covariates and also with perinatal mortality.

In Chapter 2, the SMR2 and SMR11 data collection schemes are described, and the variables used in the regression analyses and the neonatal classification are defined.

Chapter 3 introduces the binary logistic regression model and several approaches to examining the fit of the model. Ordered logistic regression models for polytomous response data and, in particular, the proportional hazards model for polytomous survival times are described. More general models, where the ordering is relaxed, are considered. A comparison is drawn with various methods related to fitting regression models to categorical data that were not used.

The first section of Chapter 4 summarises the main findings relating to the three perinatal outcomes. Later sections describe the results in more detail and examine the impact of changes in model assumptions for each outcome considered

Chapter 5 introduces and describes the latent class model for neonates. One to six class models are explored but problems with multiple maxima in the likelihood arose when four or more classes were fitted. The impact of missing data, the stability of the classification between 1978 and 1980, and the effect of repeating the analysis on random halves of the data are exmined for the three class model.

Chapter 6 discusses the use of SMR2 data in epidemiology, the results of the regression analyses and the cost of fitting regression models to categorical data. The neonatal classification is discussed in the second part of Chapter 6.

GLOSSARY OF TERMS AND ABBREVIATIONS

ABORTION - A pregnancy terminated before the 28th week of gestation where the foetus showed no sign of life.

GESTATIONAL AGE - Measured in completed weeks since the first day of the last menstrual period.

INDUCED ABORTION - A pregnancy deliberately terminated before the 28th week of gestation where the foetus shows no sign of life.

LIVEBIRTH - An infant of any gestational age which showed signs of life.

LGA - Large for Gestational Age, an infant above the 90th percentile of the birthweight distribution for its gestational age.

LMP - Last Menstrual Period.

MULTIPARAE - A woman in her second or greater pregnancy terminating in a live or still birth

NEONATE - An infant in the first month of life.

PARITY - The number of previous live or still births.

PERINATAL MORTALITY RATE - Stillbirths plus first week deaths per 1,000 live or still births.

PRETERM DELIVERY - A birth before the 37th week of gestation.

PRIMIPARAE - A woman in her first pregancy terminating in a live or still birth.

SGA - Small for Gestational Age, an infant below the 10th percentile of the birthweight distribution for its gestational age.

SMR2 - Scottish Morbidity Record 2, the maternal discharge record.

SMR11 - Scottish Morbidity Record 11, the neonatal discharge record.

SPONTANEOUS ABORTION - A pregnancy terminating spontaneously before the 28th week of gestation where the foetus showed no signs of life.

STILLBIRTH - An infant born after the 28th week of pregnancy which showed no signs of life.

INTRODUCTION

At the discharge of every woman from Scottish obstetric services a document, the Scottish Morbidity Record 2 (SMR2), containing information about the mother and a summary of the health of her infant, is completed. Over 99 per cent of Scottish births take place in hospital and, during the fifteen years since its inception, the SMR2 data collection scheme has gradually been accepted by all hospitals so that its coverage in recent years is almost complete. A separate document, the Scottish Morbidity Record 11 (SMR11), covers the period between the birth and eventual discharge of a liveborn infant. The SMR11 scheme is a more recent innovation and not all hospitals return this information. The two sources of data provide an opportunity for epidemiological investigation of pregnancy outcome and neonatal morbidity in an annual population of approximately 65,000 maternities in Scotland.

The relationship between three perinatal outcomes, birthweight, gestational age and birthweight standardised for gestational age and ten covariates from the SMR2 scheme, the sex of the infant, the height, the age, the socio-economic status and the obstetric history of the woman, are examined in a series of regression analyses. The data set is large, and a considerable reduction results from categorising the outcome variables to form binary and polytomous responses, and each covariate into two to four groups. Logistic regression models are used to analyse binary response data, and ordered logistic and proportioanl hazards models are considered for the polytomous responses. By modifying these latter models it is possible to describe a variety of changes in the distributional form of the response variable.

The first stage of the study examines the risk of birthweight below 2,500 gms, 2,000 gms, 1,500 gms and 1,000 gms in a series of binary logistic regression models. Several tests are made of the adequacy of the main effects logistic models. The significance of interactions are tested; residuals from the models are compared to simulated residuals; observed probabilities of birthweight below each cutpoint are plotted against fitted probabilities; and goodness-of-link tests of the adequacy of the logistic link within a two parameter family of alternative link are performed. The association between the covariates and birthweight may be explained by association with either gestational age or birthweight standardised for gestational age, and these two mediating outcomes are the subject of the remaining regression analyses.

In the second stage, preterm gestational ages are treated as polytomous foetal survival times and are related to the covariates in a proportional hazards model. The proportionality assumption is investigated by plotting the logarithm of the weekly hazards of delivery for each level of the covariates; by including linear time dependent terms; and by allowing the hazards to vary freely for each preterm week of gestation. Covariate interactions are investigated in a binary regression model of preterm delivery. The approach of modelling the hazard of delivery is compared to a polytomous logistic regression model in which the covariates are related to the probability of delivery below each week of gestation, and to a binary logistic model of preterm delivery.

In the third stage, birthweight standardised for gestational age in categories defined by seven percentiles of the birthweight distribution at each gestational age is examined in a logistic regression model for ordered polytomous response data. The model assumes that the standardised birthweight distribution is stochastically ordered with respect to the covariates and the odds ratios of birthweight below all seven percentiles are constrained to be equal. The model is tested by relaxing the constraint so that odds ratios for each covariate in turn vary over the percentiles, while the odds ratios for other covariates remain constrained. These results are compared to the results of a series separate binary logistic regressions models for birthweight below each percentile, where odds ratios for all covariates vary simultaneously.

The analysis of gestational age was the first part of the study to be carried out and was based on one year of data, 1981. When the analyses of birthweight and birthweight standardised for gestational age were performed, data for 1982 were available and data for 1980 were also included to form a three year study period. Because fitting the models was expensive in computer time the analysis of gestational age was not repeated for the three year period. Before 1980 some of the information on the SMR2 scheme was less accurate and earlier years are not considered in any of the regression analyses.

CHAPTER 1 : EPIDEMIOLOGY OF BIRTHWEIGHT AND GESTATIONAL AGE

1.1 Outcome Variables

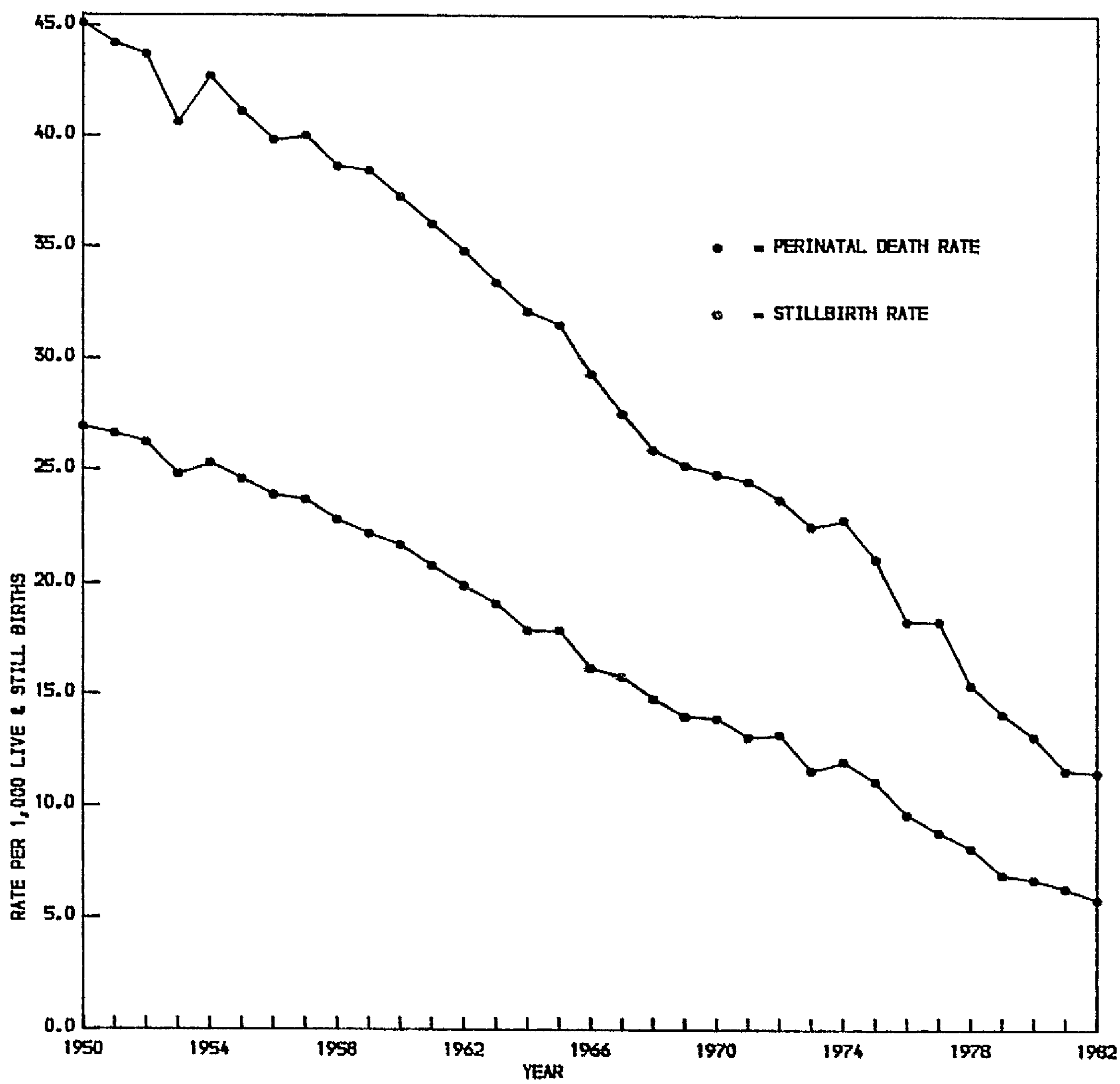
1.1.1 Introduction

The last thirty years have been characterised by continuing falls in the stillbirth, first week death and perinatal death rates, as illustrated by the Scottish rates for 1950 to 1982 in Figure 1.1 (Registrar General Scotland, 1983). In 1950 the stillbirth rate was 26.9 and the perinatal death rate was 45.1 per 1,000 live and still births. After a steady decline in perinatal deaths, the stillbirth rates during the years of this study, 1980, 1981 and 1982 were 6.5, 6.3 and 5.8, and the perinatal death rates were 13.1, 11.6 and 11.5 respectively. Perinatal death is now a rare event in Scotland and other Western nations, and attention has turned to other measures of infant health.

One of the most frequently studied pregnancy outcome variables is birthweight, providing a summary measure of foetal growth and the general health of the infant at the time of delivery. Birthweight is strongly influenced by the length of gestation, and these two outcome variables, along with a third, birthweight standardised for gestational age, are the subject of the regression analyses that are described in Chapters 2, 3 and 4. The association between perinatal mortality and birthweight and gestational age is described for the Scottish population in section 1.1.2 and 1.1.3. Some general remarks about the measurement of gestational age and birthweight standardised for gestational age, and several difficulties surrounding the interpretation of birthweight standardised for gestational age are discussed. Later sections of Chapter 1 review the literature

FIGURE 1.1

SCOTTISH STILLBIRTH & PERINATAL DEATH RATES 1950-82 REGISTRAR GENERAL



concerning the relationship between maternal, socio-economic and biological covariates and pregnancy outcome.

1.1.2 Birthweight

Table 1.1 demonstrates the strong association between perinatal mortality and birthweight, using rates calculated from Scottish singleton infants in 1980 to 1982. The perinatal death rate for infants of birthweight less than 1,000 gms was 758 per 1,000 live and still births, this comprised a stillbirth rate of 281 per 1,000 births, and a first week death rate of 663 per 1,000 live births. The perinatal death rate fell to 391 for singleton infants of birthweight 1,000 to 1,499 gms, 148 in the 1,500 to 1,999 gms band, 35 in the 2,000 to 2,499 gms band, and remained in single figures for all higher birthweight bands, with only a slight increase for infants of birthweight greater than 4,500 gms. All three mortality rates for infants of unknown birthweight were high. Conversely, singleton infants of birthweight less than 1,000 gms comprised 0.2 per cent of births but 16 per cent of perinatal deaths, while singleton infants of birthweight less than 2,500 gms comprised 5.8 per cent of births, but 63 per cent of perinatal deaths.

In spite of the continuing fall in perinatal mortality there has been little change in the distribution of birthweight, Table 1.2 (Forbes & Mckellar, 1985). These figures are based on data produced by individual Health Authorities prior to 1974, and on SMR2 data in later years. The 1973 distribution from SMR2 is also available and shows a similar pattern to later years. It can be seen that, before 1974, the rate of birthweight below 2,500 gms varied between 6.6 and 7.5 per cent, but there did not appear to be a trend in the rates. From 1976 to 1982, ignoring the missing

TABLE 1.1

SCOTTISH SINGLETON STILLBIRTH, FIRST WEEK DEATH AND PERINATAL DEATH RATES BY BIRTHWEIGHT, 1980-82

Birthweight	Singleton Births Percentage Distribution	Stillbirths, rate per 1000 live and stillbirths	First week deaths, rate per 1000 live births	Perinatal deaths, rate per 1000 live and stillbirths
< 1000 gms	.2	280.6	663.2	757.7
1000-1499 gms	.6	182.5	250.1	390.8
1500-1999 gms	1.1	88.7	64.8	147.7
2000-2499 gms	3.9	23.2	11.9	34.8
2500-2999 gms	17.5	5.2	2.7	7.9
3000-3499 gms	38.9	2.1	1.1	3.2
3500-3999 gms	28.6	1.3	.8	2.2
4000-4499 gms	7.9	1.1	1.4	2.6
≥ 4500 gms	1.2	3.0	1.3	4.3
Unknown	.2	62.8	245.0	292.5
TOTAL (Number)	100.0 (201,624)	5.9 (1,192)	5.2 (1,032)	11.0 (2,224)

From Scottish Health Statistics (1981, 1982, 1983)

TABLE 1.2

SCOTTISH BIRTHWEIGHT PERCENTAGE DISTRIBUTION, TOTAL BIRTHS 1963–1982

	YEAR																			
Birthweight	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982
≤1000 gms	.5	.5	.5	.5	.4	.4	.4	.4	.4	.4	.4	.3	.3	.3	.3	.3	.2	.3	.3	.3
1001–1500 gms	.8	.9	.8	.8	.7	.8	.7	.7	.7	.8	.8	.6	.7	.6	.6	.6	.7	.7	.7	.7
1501–2000 gms	1.6	1.5	1.5	1.4	1.4	1.3	1.3	1.4	1.3	1.4	1.4	1.1	1.3	1.3	1.2	1.2	1.3	1.3	1.3	1.3
2001–2500 gms	4.7	4.7	4.6	4.3	4.6	4.6	4.2	4.5	4.3	4.8	4.7	4.3	4.8	4.6	4.5	4.5	4.4	4.3	4.3	4.3
>2500 gms	92.5	92.4	92.6	93.1	92.9	93.0	93.4	93.1	93.3	92.7	92.8	74.1	86.9	90.6	90.7	91.5	90.3	90.6	91.6	91.7
Unknown	–	–	–	–	–	–	–	–	–	–	–	19.7	6.1	2.6	2.6	2.0	3.1	3.0	1.9	1.8
Number of Births	105,775	105,168	102,495	98,125	97,926	96,212	91,273,	88,571	87,883	79,597	75,265	70,955	68,709	65,584	62,905	64,883	68,850	69,374	69,506	66,586

From Forbes, J.F. and McKellar, D. (1985)

cases, the proportion of infants with birthweight below 2,500 gms varied between 6.6 and 6.8 per cent. This stability in rates is surprising, since there have been considerable changes in patterns of fertility, in relation to, for example, parity and maternal age. Alberman (1981) noted that between 1953 and 1975 the proportion of English infants below 2,501 gms remained steady (between 6.4 and 7.0 per cent), in spite of an increase in maternal smoking between 1958 and 1970 which might be expected to result in a reduction in mean birthweight of 20 gms, or a 1 per cent increase in birthweight below 2,501 gms.

Although Scottish birthweight specific perinatal death rates have declined for all birthweight categories, Forbes et al (1982) found that the greatest reduction (52 per cent) between 1970 and 1978 occurred amongst infants weighing over 2,500 gms. The decline for infants weighing under 1,500 gms (16 per cent) was much lower. Because of the higher rates of perinatal mortality amongst low birthweight infants, marginal changes in the birthweight distribution accounted for 13 per cent of the reduction in perinatal mortality between 1970 and 1978. In 1970 infants weighing under 2,500 gms accounted for 58 per cent of perinatal deaths, by 1978 they accounted for 65 per cent. Further improvement in perinatal survival is increasingly dependent either upon reductions in the rate of low birthweight or in continued falls in perinatal mortality amongst low birthweight infants. For these reasons the aetiology of birthweight and antecedents of low birthweight in particular are of great interest.

1.1.3 Gestational Age

Table 1.3 demonstrates the increasing rates of preterm delivery amongst singleton infants of decreasing birthweight. Approaching 95 per cent of infants with birthweight below 1,500 gms were delivered preterm. The figure fell to 79 per cent for infants of birthweight 1,500 to 1,999 gms, 39 per cent for infants of birthweight 2,000 to 2,499 gms and, overall, the preterm rate was 5 per cent for all singleton deliveries. Table 1.4 shows the risk of stillbirth, first week death and perinatal death suffered by singleton Scottish infants of differing gestational age. The three sets of gestation specific mortality rates dropped markedly between the periods 28-30, 31-33, 34-36 weeks of gestation and later deliveries. As with the birthweight specific mortality rates, the highest rates were only experienced by a small proportion of the population.

Gestational age is measured in completed weeks since the first day of the last menstrual period (LMP). Ovulation and conception are assumed to take place two weeks before the first day of the next expected menstrual period which is normally, two weeks after LMP. If the woman is not sure of her dates and for a variety of other reasons it is often difficult to obtain an accurate measure of gestation. Even when the date of LMP is available it may, for example when the menstrual cycle is prolonged or irregular, be inappropriate as a measure of gestation. One further difficulty is that monthly bleeding during early pregnancy may be confused with menstruation. Hall et al (1985a) list several reasons for absence of information on LMP, which included conception immediately following a previous pregnancy or during oral contraceptive use, and continual bleeding. Gestational age can also be estimated from ultrasound

TABLE 1.3

SCOTTISH SINGLETON PRETERM DELIVERY RATES BY BIRTHWEIGHT, 1980-82

Birthweight	Preterm Delivery per 1000 Live and Stillbirths
$<$1000 gms	931.3
1000-1499 gms	946.6
1500-1999 gms	790.0
2000-2499 gms	386.6
2500-2999 gms	81.7
3000-3499 gms	13.4
3500-3999 gms	3.5
4000-4499 gms	1.9
$\geq$ 4500 gms	2.2
TOTAL	50.6

TABLE 1.4

SCOTTISH SINGLETON STILLBIRTH, FIRST WEEK DEATH AND PERINATAL DEATH RATES BY GESTATIONAL AGE (≥ 28 weeks), 1980-82

Gestational Age (completed weeks)	Singleton Births Percentage Distribution	Stillbirths, Rate per 1000 Live and Stillbirths	First Week Deaths, Rate per 1000 Live births	Perinatal Deaths Rate per 1000 Life and Stillbirths
28-30	.5	170.2	257.0	383.4
31-33	.9	127.2	79.1	196.3
34-36	3.7	9.5	15.4	50.9
37-38	14.1	6.7	3.4	10.3
39-42	79.4	1.8	1.3	3.2
≥ 43	1.3	7.6	3.5	11.1
TOTAL	100.0	5.9	3.8	9.7
(Number)	(197,602)	(1,160)	(752)	(1,912)

From SMR 2 1980-82

measurements before the 20th week of gestation and from clinical observations soon after birth.

Buerkens et al (1984) found that infants of mothers with unknown LMP had higher risks of low birthweight which could, it was suggested, be a result of the poor socio-economic and demographic status of their mothers. Similarly Hall et al (1985a) found that the characteristics of women with uncertain LMP were generally less favourable than those with approximate or certain LMP. In a companion article (1985b), the same authors reported that uncertainty of gestation was a better predictor of several adverse pregnancy outcomes than social class, maternal age, marital status, educational status, smoking history and parity, and that it appeared to act independently of these maternal factors. They suggested that uncertainty of gestation may be identifying some additional behavioural or physical trait, and that exclusion of such cases from an analysis would bias the results. In the regression analyses of this study, births of uncertain gestational age were included in the analysis, but the role of uncertainty of gestation as a predictor was not investigated.

The majority of epidemiological investigations have examined gestational age dichotomised into preterm (defined variously as birth before the 36^{th}, 37^{th} or 38^{th} week of gestation) and term or later deliveries. Since early work in the 1960's (Shapiro et al, 1962; French & Bierman, 1962; and Mellin, 1962), some studies have used life table techniques to examine the distribution of gestational age. More recently, foetal life tables have been compiled by Bakketeig et al (1978) for competing outcomes, including foetal loss, and delivery preceding a first month or first year death. Some studies using life tables seek to explain the association between prenatal care and premature delivery. The situation is complicated by the reduced length of time women delivering preterm have had to initiate prenatal care. Terris & Glasser (1974) found that the percentage of mothers having their first prenatal visit in each month of gestation amongst those still undelivered at the start of the month was substantially the same for premature and mature births, and concluded that the association between prenatal care and premature delivery was spurious. This was not the conclusion of Harris (1982) who found that mothers who received prenatal care did have lower risk of preterm delivery but could not eliminate the possibility that this was due to association of both variables with underlying maternal factors.

1.1.4 Birthweight Standardised for Gestational Age

Although preterm deliveries form a large proportion of low birthweight infants (Table 1.3), the low weights of a substantial group is not explained by shortened gestation. Low intrauterine growth amongst these infants can result from various factors (Chiswick, 1985), some cause symmetrical impairment of growth, others assymmetrical impairment with affected infants tending to be long and thin. Infants of low birthweight due to shortened gestation and the group with abnormally low birthweight for their gestational age, as a whole, have different prognoses during the perinatal period and later. Neligan et al (1976) identified two groups, one of preterm infants born before the 37^{th} completed week of gestation, the other of infants born too small, below the 10^{th} percentile of the birthweight distribution for their gestational age. This latter group was further subdivided into infants below the 5^{th} percentile for comparison to the preterm infants who formed approximately five per cent of the population. At six to seven years of age the groups were compared with respect to various language, IQ, behavioural, neurological and size measurements. In all but one instance there were increasingly adverse scores from the controls, the infants born between the 10^{th} and 5^{th} percentiles to the infants below the 5^{th} percentile of the birthweight distribution. On all scores the five per cent of preterm infants scored better than the five per cent with birthweight below the 5^{th} percentile. The authors reviewed several other reports and concluded that both groups of infants experienced increased risk of subsequent impaired development but that the group of infants below the 5^{th} percentile were at greater risk. A preterm infant, they suggested, with otherwise normal intrauterine growth pattern can develop normally if he can avoid neonatal complications but a child whose birthweight is below the 5^{th} percentile has already suffered seriously impaired growth and his later development may also be affected.

The most commonly used approach to the identification of infants with an abnormally low rate of foetal growth is comparison to birthweight standards. Thomson, Billewicz & Hytten (1968) presented some of the first British birthweight standards based on over 50,000 births in Aberdeen during the years 1948-64. More recently, standards based on Aberdeen data have been published by Carr-Hill & Pritchard (1985), and on Scottish data by Smalls & Forbes (1983). Birthweight standards can be constructed by assuming the birthweight distribution at each gestational age is Normal, and estimating age specific means and standard deviations so that percentiles for each gestational age can be deduced using standard Normal deviates. A second method of preparing birthweight standards, involves the direct estimation of percentiles of the distribution at each gestation. Standardisation for other factors that are known to affect the birthweight distribution is often required. In the first approach this may be done by calculation of increments to be added to the mean of the birthweight distribution for each standardising factor. Alternatively, and for either approach, separate tables within appropriate sub-groups of the obstetric population can be prepared. Large numbers of births need to be available in order to obtain reliable estimates of the birthweight distributions at low, infrequently occurring, gestations.

Birthweight standards accurately reflect only the population from which they were calculated. The birthweight distribution varies across countries, races, populations at different heights above sea level and many other factors. If there are secular trends in the distribution of birthweight and gestational age the standards may require frequent revision. For example, Carr-Hill & Pritchard (1985) noted that between 1951-55 and 1976-80 the proportion of preterm births fell in Aberdeen, within this group however, there was a trend to lower birthweight. There are several points for consideration when deciding which standards are most appropriate for a given application. For clinical decision making it may be sensible to standardise for known correlates of birthweight that are not thought to be associated with impaired development. For research purposes the choice depends on which variables are of primary interest. If the standards are controlled for factors other than gestational age, any interaction between the primary and controlled variables cannot be examined, and results might be misleading if the variables of primary interest are associated with birthweight through the standardised variables.

Birthweight standardised for gestational age amongst preterm infants cannot be interpreted as a measure of normal intrauterine growth. The birthweight of an infant born preterm may be very different from that of a foetus of the same gestational age that survives in utero to later pregnancy. Birthweight percentiles are often based on all livebirths, but the fact that an infant is delivered preterm, is itself an indication of a poor pregnancy outcome. Birthweight was found to be substantially lower within gestational age groups amongst infants who died from several causes as neonates, than amongst control infants who survived the neonatal period (Naeye & Dixon, 1978). The causes of neonatal death were chosen because they were not excluded from earlier growth standards. A better approximation to intrauterine growth may be obtained by removing deaths, and other critical conditions from the standard population, but this would leave few infants at early gestations. Standardising for gestational age is not necessarily beneficial for prediction. Hellier & Goldstein (1979) showed that while infants with extreme values of both gestational age and birthweight standardised for gestational age experienced increased risk, birthweight alone was a better predictor of perinatal mortality than either. The causal hypotheses behind the construction of standards of birthweight for gestational age was questioned by Wilcox (1981). If achieved birthweight prompts delivery, then measuring extreme gestational age for given birthweight would be a more natural construct, and variables defined in this way have been considered in several studies (Hoffman et al, 1974; Paneth et al, 1982). Wilcox also illustrated how a hypothetical factor associated with both lower birthweight and gestational age may result in higher birthweight for preterm deliveries but lower birthweight for term deliveries.

1.2 Maternal, Socio-economic and Biological Correlates of Perinatal Outcome

1.2.1 Introduction

Three major surveys, the National Survey of Health and Development (Joint Committee, 1948), the British Perinatal Mortality Survey (Butler & Alberman, 1969), and the British Births Survey (Chamberlain et al, 1975) have established patterns of risk of low birthweight and preterm delivery associated with maternal, socio-economic and biological covariates in Britain. The most detailed analysis was performed on data from the second survey, in which information on all births in England, Wales and Scotland during the week of the 3^{rd}-9^{th} March 1958 was collected for administrative, obstetric and socio-biological research. This section comprises a review of risk factors of low birthweight, preterm delivery and the birth of an SGA or LGA (small- or large-for-gestational-age) infant, identified in the report of the 1958 survey. Later sections contain a more detailed description of the association between pregnancy outcome and the covariates considered in this study.

An examination of the rates of birthweight below 2,501 and 2,001 gms in the 1958 survey showed that infants of both teenagers and women aged over 35 years were more likely to be of lower birthweight, as were the infants of single women, and substantial gradients of increasing risk were associated with decreasing social class and maternal height. Mean birthweight increased with maternal age, parity, social class, and height, while lower mean birthweight was observed amongst infants of smokers and single women. In an analysis of covariance of birthweight adjusted for various maternal characteristics including smoking and pre-eclampsia (an adverse condition which may arise during pregnancy), the impact of maternal age and social class was reduced to borderline significance. Peters et al (1983) compared identical analyses of covariance of the 1958 and 1970 survey data, and found that the same four factors, height, parity, smoking and pre-eclampsia, were predominant in both years.

Women of age less than 20 years or greater than 34 years in the 1958 survey had higher rates of preterm delivery than the other age groups, as did women with none or two or more births, compared to women with only one previous birth. Increasing rates of preterm delivery were associated with decreasing social class and maternal height. Several covariates were related to birthweight controlled for gestational age. Analyses of variance including gestational age as a covariate indicated that male infants had higher mean birthweight than females, and increasing mean birthweight was noted with increasing parity, maternal height and higher social class. Interactions of each covariate with gestational age were tested and were mostly non significant. The distributions of the maternal covariates for infants less than two standard deviations below the mean birthweight for their gestational age were compared to the distributions for all infants. Higher percentages of primiparae, mothers aged less than 20 years, and mothers in social classes IV and V and of short stature were observed for the light for dates infants.

1.2.2 Sex of Infant

Several investigators have noted a reversal of risks associated with sex of infant, male infants having higher mean birthweight than female infants, but female infants having longer gestations. Hall & Carr-Hill (1982) reviewed several reports of increased rates of preterm delivery amongst male infants and investigated possible explanations. The higher rates amongst males did not appear to be due to induction or elective caesarean section, because the boys were delivered spontaneously more often than the girls. It was suggested that the greater weight of male infants advanced their production of hormones which may effect the onset of labour.

1.2.3 Social Class

Social class is usually based on paternal occupation, and, except for certain occupations involving toxic substances, is unlikely to have a direct causal influence on pregnancy outcome. Illsley & Mitchell (1984) explain the biological significance of social class as a summary variable for a wide range of background characteristics of the mother and family, including per capita income, area of residence, housing conditions, educational level and nutritional and health status, all of which are highly intercorrelated. It is difficult to identify specific social factors in the aetiology of poor outcome. Individual correlations between these factors and poor perinatal outcome are largely reduced to insignificance when controlled for the general measure of social class based on parental occupation. The importance of social class lies in its ability to encapsulate the exposure of the mother to past experiences during childhood and adult life that have affected her biological functioning as a reproductive agent.

1.2.4 Maternal Height

Epidemiological studies have repeatedly demonstrated a strong relationship between maternal height and birthweight, and, in some reports, preterm delivery. Although there is an obvious genetic explanation of the relationship between maternal height and birthweight, other causal mechanisms could also be at play. Baird (1985) reviewed stillbirth rates amongst social class IV-V mothers from 1950 to 1980. He suggested that food rationing policy, started during the war, had produced better nourished mothers and laid the foundations for the improvement in the stillbirth rate amongst their offspring born in the 1960's. He also examined histories of women in Aberdeen who delivered a low birthweight infant, which could not be explained in terms of obstetric complications, maternal disease or foetal deformity. The mothers and grandmothers of unexplainably low birthweight infants were more likely to be short, of light body build and of poor health than the mothers and grandmothers of explainably low birthweight infants. A concentration of the grandmothers of unexplainably low birthweight infants were born between 1930 and 1932, the years when the depression was at its worse. These findings suggest that maternal height may, in part, be a reflection of poor nourishment during infancy, and, more generally, of family background and social environment.

1.2.5 Obstetric History

A previous spontaneous abortion or perinatal loss, and the birthweight and gestational age of the immediately preceding sibling have been identified as risk factors for poor perinatal outcome in many studies. For example, birthweight below one standard deviation below mean birthweight in the previous pregnancy was associated with increased risk of an SGA infant but decreased risk of an LGA infant, while an elder sibling's birthweight above one standard deviation above mean birthweight was associated with decreased risk of an SGA infant but increased risk on an LGA infant (Ounsted et al, 1985). Van den Berg & Oechsli (1984) found increased risk of preterm delivery following foetal death, preterm delivery, or low birthweight. Amongst 3,502 British women doctors, mean birthweight of infants to women whose only previous pregnancy had ended in a spontaneous abortion did not differ significantly from either first births or second births following a livebirth, but mean birthweight following two spontaneous abortions was significantly lower, and mean birthweight following three spontaneous abortions was lower still (Alberman et al, 1980). Birthweight in pregnancy preceding a spontaneous abortion was also examined by Alberman et al, and was slightly lower than preceding a livebirth, suggesting that women whose pregnancies ended in a spontaneous abortion were at some form of disadvantage in the earlier pregnancy.

Bakketeig, Hoffman & Harley (1979) examined the tendency to repeat a specific birthweight and gestational age combination in successive pregnancies. A grid of birthweight in 500 gms intervals and gestational age in three weekly-intervals was used to define the pregnancy outcomes. The relative distribution of second births across the compartments of the grid for women whose first birth was in a specified area of the grid was calculated by dividing the relative frequency for the specified women by the overall relative frequency of second births. In all cases the highest frequency ratios were observed in the compartments of the grid close to the birthweight and gestational age of the specified earlier births. In particular, women with a previous

SGA or LGA infant had a relatively high chance of repeating their previous pregnancy outcome. The authors also examined third pregnancies following specific perinatal outcomes in the first and second pregnancy. Each outcome examined (pre- and post-term delivery, low and heavy birthweight) was most likely to occur in a third pregnancy following two similar pregnancies and least likely to occur if neither of the two previous pregnancies ended with the outcome. If one but not both of the first two pregnancies had ended with the outcome, it was most likely to recur in the third pregnancy if the second rather than the first pregnancy was similar.

In a subsequent article, Bakketteig & Hoffman (1983) investigated the risk of perinatal mortality related to the tendency to repeat birthweight and gestational age. Amongst second births in birthweight categories ≤1,500 gms and 1,501-2,000 gms, infants were at lowest risk of perinatal death if their elder siblings were in the same birthweight category, and at higher risk if the elder sibling was either lighter or heavier. For the higher birthweight categories of second births, the infant had lowest risk if its elder sibling was of slightly lower birthweight. The analysis was also repeated for birthweight standardised for gestational age in five categories. A similar pattern emerged with, however, one exception. Second births above the 90th percentile were at highest risk of perinatal death if their elder sibling was also above the 90th percentile.

1.2.6 Previous Induced Abortion

Although there is a large international literature concerning the relationship between induced abortion and subsequent reproduction no general pattern of risk has emerged.

Hogue et al (1982), in a comprehensive review of reports on the sequelae of medically induced abortion, concluded that there was no risk of shortened gestation but that low birthweight was more frequent after abortion performed by dilatation and curretage under general anaesthesia. However, the studies were carried out in many countries where clinical practice varied, and induced abortion had different legal status and degrees of cultural acceptability. Within individual countries there may have been increased risk following induced abortion which were not augmented by similar findings elsewhere.

Two study designs are used for examining risk following induced abortion. Firstly, in an abortion cohort design, a group of women who are identified at the time of an induced abortion are followed prospectively. The comparison group in this design may be defined as women who did not have an induced abortion or women delivering a full term birth. The effective size of these studies when examining pregnancy outcome (rather than fertility), is not the number of women recruited, but the number who become pregnant within the follow-up period. In the second, pregnancy cohort designs, any previous induced abortions are noted amongst women who become pregnant within the period of the study. These designs are usually of limited value when studying fertility and early spontaneous abortion because information on women failing to acheive a further pregnancy is not available and women may not report all abortions.

Harlap et al (1979) used a life table to examine the risk of first and second trimester (the first and second three months of pregnancy respectively) spontaneous abortion following an induced abortion. The study was based on 32,000 members of a health plan who attended their first prenatal visit between 1974 and 1976 in

Northern California. No increase in risk was found for multiparae after adjusting for maternal age. Primiparae with a history of induced abortion did suffer increased risk, particularly during the second trimester and the relative risk following two or more induced abortions was significant. Although it was not known where, or by what method previous pregnancies had been terminated, in 1973, hospitals within the health plan changed from performing terminations by dilatation and currettage to the use of laminaria tents. The excess of risk in primiparae was found mainly amongst women with an induced abortion performed before 1973.

Most studies of pregnancy following induced abortion using British data rely on small numbers of women. In a hospital based study of 211 secundagravidae whose first pregnancy ended in a legal termination, Richardson & Dixon (1976) used as a control group women whose first pregnancy had ended in spontaneous abortion. There were significantly more first and second trimester abortions and preterm deliveries amongst cases than controls. Another British study (Mackenzie et al, 1977) compared 204 pregnancies in women (contacted through a postal questionaire with a 69 per cent response rate) who had a history of abortion induced by prostaglandins with a control group of women with no previous induced abortions. The study revealed increased risk of spontaneous abortion following induced abortion, but no difference in birthweight or gestational age in third trimester deliveries. More recently, Frank et al (1985) have reported initial results from an abortion cohort study of 745 pregnancies to women whose index pregnancy ended in an induced pregnancy, and 1,339 control pregnancies to women whose pregnancy was unplanned.

There was some suggestion of increased risk of spontaneous abortion, stillbirth, preterm delivery and low birthweight amongst women with a previous induced abortion, but, possibly through lack of sufficient numbers in the study, none of these results was significant.

1.2.7 Differential Fertility and Parity Effects

Through use of the pill and other contraceptive methods women can exercise considerable control over their fertility. Billewicz (1973) demonstrated a trend in fertility following unsuccessful (defined as a spontaneous abortion or perinatal death) and successful pregnancies. Women of a given pregnancy order who had no further pregnancies had a higher percentage of previous successful outcomes than those who continued to a further pregnancy. The lowest rate of unsuccessful outcome was experienced by women in their final pregnancy. This suggests that women with poor reproductive histories compensated for previous losses by initiating a further pregnancy, and were thus more likely to reach higher pregnancy orders than women with good histories. Frank et al (1985) even found a higher continuation rate amongst women whose foetal loss had been an unplanned pregnancy. Interpregnancy intervals are shorter following a stillbirth than a livebirth (Resseguie, 1973; Bjerkedal & Erikson, 1983). Leridon (1976), however, found a different pattern in a study of two groups of women, one near Paris the other in Martinique. In both groups pregnancy continuation rates were higher following livebirths than stillbirths. In Martinique there was little control over fertility, and the result could be due to higher rates of sterility following a poor outcome. The results from women near Paris suggest that poor outcome may discourage some women from further pregnancy. Strobino et al

(1980) found that the interval between ceasing birth control and conception was 50 per cent longer for a group of women with a history of spontaneous abortion than in women with no such history. These latter results do not, necessarily, contradict the reproductive compensation theory, which, involving the time between a previous delivery and the following conception or birth, should be largely influenced by a woman's desire for a further pregnancy.

Self selective differentials in fertility together with the tendency to repeat a poor pregnancy outcome, have a major impact on the interpretation of cross-sectional and longitudinal studies of perinatal outcome with respect to parity. Cross-sectional analyses tend to reveal a J-shaped pattern of risk associated with increasing parity. The first pregnancy is found to be at high risk, the second and third at lower risk and thereafter the risk increases with increasing parity. The patterns of risk found in longitudinal studies are typified by two studies, of perinatal mortality and preterm birth (Bakketeig & Hoffman, 1979; 1981). Risk was found to increase with increasing completed sibship size; within sibships of a given size, however, the risk decreased with parity and was lowest for the final birth. In the cross-sectional studies a larger proportion of women with poor pregnancy histories make up a disportionate number of the women at higher pregnancy orders thus raising their apparent level of risk. While in the longitudinal analyses of completed sibships the lowest rates for final pregnancies reflect the tendency to complete childbearing with a successful pregnancy.

In a computer simulation where all pregnancies had equal chance of ending in a foetal loss, Golding et al (1983) studied bias in pregnancy order effects from cross-sectional and longitudinal forms of analysis. All pregnancies were subject to a desired family size, simulated for each woman from a given distribution, and all women were further assumed to suffer a constant rate of sterility following each pregnancy. The simulation showed that while the longitudinal approach revealed a decreasing risk of foetal loss with increasing pregnancy order, it was the cross-sectional approach that replicated the true constant rate of foetal loss at each pregnancy order. Roman (1984) expanded the assumption of constant risk of foetal loss at each pregnancy order and assumed a mixture of two groups both with constant but different rates of foetal loss. This led to an apparent variation in rates at each pregnancy order, and by futher manipulations of the distribution of desired family size the cross-sectional approach produced the traditional J-shaped pattern of risks with increasing pregnancy order. The results of one simulation study (Wilcox & Gladen, 1982) suggested that the increasing risk of spontaneous abortion associated with increasing gravidity could be explained by an underlying risk that was constant for each woman but varied in the population according to a Beta distribution, a distribution of desired family size and an effect due to increasing maternal age.

It is not just the number of previous adverse pregnancies that is associated with current outcome. The order in which outcomes occur is also important. Leridon (1976) and Roman et al (1978) suggest that the risk of foetal loss should be studied for specific combinations of previous pregnancy outcomes. Both studies showed that the lowest rates of foetal loss at any pregnancy order were amongst women with no previous losses, while all the previous pregnancies of the group experiencing the highest rates had ended in foetal loss. Yudkin & Baras (1983) controlled for parity, the presence of previous poor outcomes and a short interpregnancy interval in a logistic regression model. The greatest risk of poor pregnancy outcome was associated with first birth, the risk fell for second births and thereafter it increased, but not significantly, with increasing parity. In this study risks in pregnancies following three or more livebirths are controlled for previous perinatal death and spontaneous abortions, and might be expected to be similar to Yudkin & Baras's results.

1.2.8 Maternal Age

There is considerable ambiguity concerning the interpretation of maternal age effects on perinatal outcome. First, an inescapable association between maternal age and parity, obscures the separate effects of these two ageing processes. To avoid the problems of self-selection to higher pregnancy orders and maternal age, Resseguie (1974) examined pregnancies to Amish women, amongst whom, it is thought, successful and unsuccessful reproducers alike continue to higher pregnancy order. Minimum rates of spontaneous abortion and stillbirth were experienced by Amish women in their early thirties, but the rates were not statistically different from those of women in their early twenties. The minimum risk of foetal loss was thus experienced considerablly later than traditional analyses have suggested. Results for pregnancy order were not reported.

The interrelationship of maternal age and secular trends in risk also confuses the interpretation of age effects. Baird (1974) examined births in Aberdeen in five year groups in the period 1948-72 in relation to the year of mother's birth. The risk of low birthweight amongst mothers of different ages in the five year time periods coincided fairly accurately with the hypothesis that mothers born between 1929-37, the years of the industrial depression, should be at highest risk. Resseguie (1976) has shown that decreasing stillbirth rates can seriously distort the maternal age association in cross-sectional analysis. Curves displaying the stillbirth rates in first births to mothers at increasing ages were based on the prospective experience of cohorts of women aged 20 in each of the years 1951 to 1967. The curves revealed that the minimum risk of stillbirth was experienced at around 25 years of age within each of the cohorts. Examining the same data but plotting stillbirth rates with increasing maternal age cross-sectionally for each of the years 1951 to 1967 suggested that the minimum risk was experienced by younger women aged around 21 years. A similar disparity between the age of minimum risk was also found for second and third births. In choosing to delay childbearing, women also postponed the calendar year of their pregnancies, and the older women in the cross-sectional analyses belonged to earlier cohorts which were at higher risk. The study presented here is not likely to be affected by this type of distortion since the rate of low birthweight has not changed appreciably within the last thirty years.

A further paper by Resseguie (1977) investigated the decision to begin childbearing related to a woman's educational status. The age of minimum risk of stillbirth in first pregnancy increased with increasing educational class, women experienced minimum risk when aged less than 20 years in the lowest educational class, but approaching 30 years in the highest educational class. The overall pattern of association with maternal age did not correspond to that experienced by women of any one educational class. The fact that the patterns of risk varied with educational status led the author to suggest that a social and not a biological process was being represented by maternal age. Women who had their first pregnancy substantially later than the mean for their educational class experienced higher risk of stillbirth and may represent a stillbirth-prone group who would have higher risk at any age.

1.2.9 Predictive Scores for Perinatal Outcome

The association of maternal and socio-economic variables with perinatal outcome has led to their inclusion in risk assessment systems. Fortney & Whitehorne (1982) review a number of risk scores for perinatal mortality, preterm delivery and low birthweight. The criteria used for comparison were sensitivity (the percentage of cases corrrectly classified), specificity (the percentage of non-cases correctly classified), and false positive and negative rates. The percentage of women classified as at risk by the various scores ranged from 12 to 69 per cent, whilst the sensitivity varied between 25 and 97 per cent and the false positive rate between 17 and 96 per cent, demonstrating that these scores may not perform well. In a paper reviewing risk scores for various pregnancy outcomes, Newcombe & Chalmers (1981) examine the contribution to predictive power of variables measured at different stages of pregnancy. Their criterion for comparison was the estimated probability that if two pregnancies are chosen at random, one delivering preterm and the other delivering term, the former will have received the more unfavourable score. Scores with highest prediction probability at given stages of pregnancy were tabulated. Amongst the scores for use in early pregnancy, the best performer had probability 0.75, predicting stillbirth, at 6 months the best performance probability was 0.80, for low birthweight, the best prenatal predictor, of neonatal death, had probability 0.95 and an intrapartum (during delivery) predictor of perinatal death had probability 0.96. These figures demonstrate that predictive power improves as the pregnancy advances and the presence or absence of adverse factors in later stages of the current pregnancy can be incorporated in the score. Predictive scores formulated for use in early pregnancy rely on detailed information concerning maternal illness and obstetric history. The SMR2 scheme is limited in these area, and is thus unlikely to provide a satisfactory risk score for perinatal outcome.

CHAPTER 2 : SCOTTISH MORBIDITY DATA AND THE DEFINITION OF STUDY VARIABLES

2.1 The Scottish Morbidity Record 2 (SMR2)

2.1.1 Introduction

The SMR2 scheme was introduced in the late 1960's to unify the collection of administrative and diagnostic information at the discharge of women from hospital maternity services. Clerical staff at hospitals abstract information from each woman's medical record onto a form designed for ease of transfer to computer. Completed SMR2 documents are sent to the Information Services Division of the Common Services Agency in Edinburgh which collates the data for administrative uses by the Scottish Home and Health Department and for distribution to local Health Boards. Data for the annual obstetric population is available for research purposes and arrives, in our case, on a magnetic tape containing individual information from each woman who delivered a stillbirth during or after the 28th week of gestation or a livebirth.

2.1.2 Format of the SMR2 Document

A copy of the SMR2 document used during 1980-82 is included in Appendix 1. The first section of the record contains general information about the mother, her age, marital status and her own occupation as well as that of her husband. A summary of the woman's obstetric history comprising the number of previous pregnancies, spontaneous and induced abortions, perinatal deaths, caesarean sections and children now living makes up the second section. A section on the current pregnancy contains maternal height, LMP, estimated gestational age, certainty of gestation, administative and other information. The record of labour comprises the number of births, the outcome of the current pregnancy (stillbirth, first week death or first month death), birthweight, sex of the infant, and further details of the labour. The final two sections contain the type of care the infant received, and provision for coding type of operation and clinical diagnostic information pertaining to the mother and her pregnancy. These two sections are not considered further here.

2.1.3 Coverage

The coverage of the SMR2 scheme can be established by comparison with the Registrar General's report of births in Scotland, the most accurate estimate of births available. The coverage figures in this section refer to the number of SMR2 births as a percentage of the Registrar General's total births. Cole (1980) reported an SMR2 coverage of 63.4 per cent in 1969, the first year for which the SMR2 scheme was in operation, by 1977, the last year Cole reviewed, the figure had risen to 99.6 per cent. Between 1980 and 1982, the years on which the following analyses are based, the coverage was around 98 per cent (Table 2.1).

2.1.4 Validity

The process of collecting the information for the SMR2 scheme introduces the possibility of various types of error in the completed document. Clinical staff may fail to observe symptoms, or only record severe cases of a clinical condition depending on hospital practice. The measurement of continuous variables, such as birthweight, may involve error. Some SMR2 information is elicited from the mother who may misunderstand the question, fail to recall salient facts or give deliberately misleading replies. Cole (1980) recognised this possibility and,

TABLE 2.1

COVERAGE OF SMR2 1980-82

YEAR	(1) SMR2 Births	(2) Registrar General Births*	(1) As Percentage of (2)
1980	67748**	69355	97.7%
1981	68515***	69490	98.6%
1982	65585***	66582	98.5%
1980-82	201848	205427	98.3%

* Figures from Table A1.1 of the Annual Report of the Registrar General Scotland 1983.

** 1980 SMR2 data based on singleton deliveries. Total births estimated using Registrar General's figures for twins and triplets.

*** 1981-82 SMR2 data based on all maternities. Total births estimated using Registrar General's figures for twins and triplets.

noting a 1.2 per cent excess of legitimate births on SMR2 compared to the Registrar General's tables, observed that a woman's statement of her marital status was rightly taken at face value. Macintyre (1978), in a study of interviews at an antenatal clinic where some information recorded on SMR2 is collected, describes how the attitudes of the staff or their construction of questions, particularly when interviewing women of different social class, can influence the reply given by the interviewee.

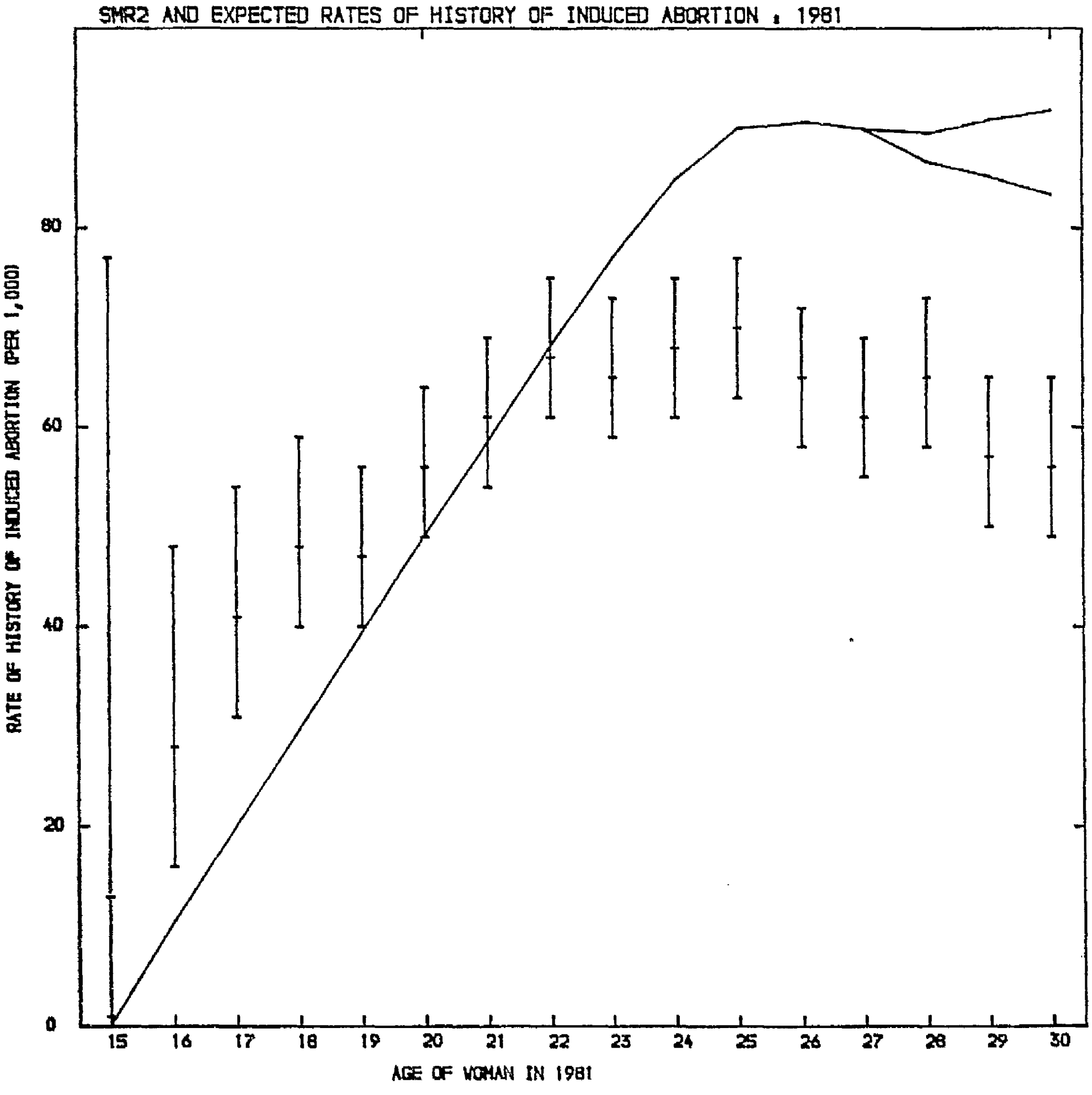
Transcription error when SMR2 information is abstracted from unstructured medical records by clerical staff may also be anticipated. On receipt of completed SMR2 forms, the Information Services Division carries out a series of computer checks before the data are released. Internal inconsistencies, such as reported cases of women suffering from exclusively male conditions, are queried with the originating hospital. These and other examinations of the validity of SMR2 information have lead to attempts to improve and standardise coding practices across hospitals.

Cole (1980) reported a comparison of 38 items of information describing 1,000 individuals collected, firstly, during the Scottish Survey of Perinatal Deaths and again in the SMR2 scheme. About 90 per cent of the items agreed between the two sources. The discrepancies were in a few items where the meaning of the question may have been confusing, and could have resulted from errors on either the SMR2 document or in the survey. More precisely defined items had a higher agreement rate of about 98 per cent. Finally, comparing SMR2 with data from the Registrar General, Cole found that there was good agreement between the maternal age distribution from the two sources.

2.1.5 Validity of History of Induced Abortion

One item of information on SMR2, the woman's statement of her history of induced abortion, may be particularly prone to error. Some impression of the accuracy of the SMR2 rates of previous induced abortion, in 1981 for example, can be obtained by comparison with age-specific rates of induced abortion in the general population during the years preceding 1981. Assuming that women in the obstetric population experienced the population rates of induced abortion in earlier years, that young women, aged under fifteen years, did not have induced abortions and that the probability of a woman having no previous abortions was the product of the probabilities of women her age having no abortions in each of the previous years, expected rates of a history of one or more induced abortions in 1981 were calculated for women in yearly age cohorts. Prior to 1969 induced abortion was illegal in Britain and two sets of figures were prepared for older women with this in mind. First, the 1969 age-specific rates were applied in earlier years and, secondly, rates of induced abortions were assumed to be zero before 1969. Figure 2.1 gives the results of the comparison for women aged 15 to 30 in 1981. Amongst women aged under twenty in the obstetric population there were more reported induced abortions than expected, while fewer women aged over 23 reported a history of induced abortion than expected. The comparison for women aged over thirty, not given here, showed that SMR2 rates of a history of induced abortion were also lower than expected and that the difference increased with age. Overall, however, the comparison is reassuring. One assumption underlying the expected rates can be seen from Figure 2.1 to be unjustified, since observed rates show that some women aged under fifteen do report a history of

FIGURE 2.1


SMR2 AND EXPECTED RATES OF HISTORY OF INDUCED ABORTION : 1981
RATE OF HISTORY OF INDUCED ABORTION (PER 1,000)
0
20
40
60
80
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
AGE OF WOMAN IN 1981

induced abortion. The assumption that the proportion of women with no induced abortions in the obstetric population is the product of age-specific annual rates of no induced abortion in the general population is probably also unjustified. Women still childbearing in their thirties may have had fewer pregnancies in earlier years, while young mothers may be more fertile and hence have higher chances of a history of induced abortion.

2.1.6 Distribution of Birthweight and Gestational Age by Certainty of Gestation

The SMR2 document contains an estimate of gestational age made by the obstetrician on the basis of LMP, ultrasound and clinical judgement. These estimates are made for all cases and cases with gestations ≤ 27 and ≥ 43 were grouped before the data were made available. Table 2.2 contains the crosstabulation of birthweight, in 500 gm intervals, by estimated gestation. With the exception of a few isolated cells, with low frequencies, the majority of the data follow the expected pattern, lying on the diagonal from low weight, short gestational age infants to heavy and late infants. Within each gestational age the distributions are broadly symmetrical.

The confidence with which the obstetrician was able to estimate gestational age is recorded on SMR2 as certainty of gestation. Tables 2.3 and 2.4 show the crosstabulations of birthweight by gestational age for infants of certain and uncertain gestations separately. There is a tendency for births of uncertain gestational age to have either longer or shorter gestations than births with certain gestational age. For example, 1.2 per cent of births where gestational age is uncertain have estimated gestations below 32 weeks, compared to 0.6 per cent of

TABLE 2.2

BIRTHWEIGHT DISTRIBUTION (500 gm INTERVALS) FOR GESTATIONS ⩾ 28 WEEKS: TOTAL BIRTHS, 1980-82

	<500 gms	500-999	1000-1499	1500-1999	2000-2499	2500-2999	3000-3499	3500-3999	4000-4499	⩾ 4500	TOTAL
28	2 (.7)	71 (25.9)	170 (62.0)	21 (7.7)	6 (2.2)	2 (.7)	1 (.4)	0	0	1 (.4)	274 (.1)
29	0	36 (14.4)	160 (64.0)	46 (18.4)	3 (1.2)	1 (.4)	2 (.8)	2 (.8)	0	0	250 (.1)
30	0	34 (8.3)	158 (41.0)	156 (40.5)	20 (5.2)	11 (2.9)	4 (1.0)	2 (.5)	0	0	385 (.2)
31	2 (.5)	23 (5.5)	126 (30.3)	204 (49.0)	40 (9.6)	11 (2.6)	7 (1.7)	2 (.5)	0	1 (.2)	416 (.2)
32	1 (.2)	27 (4.1)	120 (18.1)	305 (45.9)	162 (24.4)	38 (5.7)	10 (1.5)	1 (.2)	0	0	664 (.3)
33	2 (.3)	8 (1.2)	78 (11.2)	251 (36.2)	276 (39.7)	63 (9.1)	14 (2.0)	2 (.3)	0	0	694 (.4)
34	1 (.1)	6 (.5)	74 (5.6)	307 (23.2)	625 (47.1)	260 (19.6)	39 (2.9)	13 (1.0)	1 (.1)	0	1326 (.7)
35	0	3 (.2)	32 (1.8)	210 (11.7)	736 (41.1)	610 (34.0)	167 (9.3)	32 (1.8)	2 (.1)	0	1792 (.9)
36	0	1 (.0)	40 (1.0)	193 (4.6)	1145 (27.4)	1827 (43.8)	791 (19.0)	145 (3.5)	27 (.6)	3 (.1)	4172 (2.1)
37	0	1 (.0)	24 (.3)	149 (2.0)	1000 (13.2)	3076 (40.7)	2498 (33.0)	707 (9.3)	92 (1.2)	18 (.2)	7565 (3.8)
38	0	2 (.0)	8 (.0)	131 (.6)	1188 (5.8)	5959 (29.3)	8777 (43.1)	3633 (17.9)	570 (2.8)	74 (.4)	20342 (10.4)
39	0	4 (.0)	3 (.0)	61 (.2)	1006 (2.7)	7461 (20.2)	16309 (44.1)	9831 (26.6)	2017 (5.5)	276 (.7)	36968 (18.8)
40	1 (.0)	6 (.0)	6 (.0)	66 (.1)	1196 (1.5)	10882 (13.9)	32246 (41.2)	25730 (32.9)	7196 (9.2)	981 (1.3)	78310 (39.8)
41	0	2 (.0)	3 (.0)	15 (.0)	282 (.8)	3404 (9.6)	12954 (36.6)	13395 (37.8)	4607 (13.3)	731 (2.1)	35393 (18.0)
42	0	0	2 (.0)	2 (.0)	42 (.7)	586 (9.7)	2058 (33.9)	2270 (37.4)	942 (15.5)	170 (2.8)	6072 (3.1)
⩾43	0	0	6 (.3)	20 (1.1)	47 (2.5)	276 (14.6)	800 (42.2)	558 (29.4)	173 (9.1)	16 (.8)	1896 (1.0)
TOTAL	9 (.0)	224 (.1)	1010 (.5)	2137 (1.1)	7774 (4.0)	34467 (17.5)	76677 (39.0)	56323 (28.7)	15627 (8.0)	2271 (1.2)	196519 (100.0)

TABLE 2.3

BIRTHWEIGHT DISTRIBUTION (500 gm INTERVALS) FOR GESTATIONS ≥ 28 WEEKS: BIRTHS OF CERTAIN GESTATION, 1980-82

	<500 gms	500-999	1000-1499	1500-1999	2000-2499	2500-2999	3000-3499	3500-3999	4000-4499	≥ 4500	TOTAL
28	2 (1.1)	49 (27.2)	113 (62.8)	13 (7.2)	0	2 (1.1)	1 (.6)	0	0	0	180 (.1)
29	0	31 (17.5)	116 (65.5)	26 (14.7)	0	0	2 (1.1)	2 (1.1)	0	0	177 (.1)
30	0	22 (9.0)	102 (41.6)	104 (42.4)	11 (4.5)	4 (1.6)	1 (.4)	1 (.4)	0	0	245 (.2)
31	1 (.3)	18 (5.9)	99 (32.4)	144 (47.1)	28 (9.2)	8 (2.6)	5 (1.6)	2 (.7)	0	1 (.3)	306 (.2)
32	1 (.2)	21 (4.7)	77 (17.1)	216 (48.1)	107 (23.8)	19 (4.2)	7 (1.6)	1 (.2)	0	0	449 (.3)
33	2 (.4)	6 (1.2)	56 (11.1)	187 (37.0)	197 (39.0)	46 (9.1)	10 (2.0)	1 (.2)	0	0	505 (.3)
34	0	3 (.3)	63 (6.8)	206 (22.4)	436 (47.3)	182 (19.8)	21 (2.3)	10 (1.1)	0	0	921 (.6)
35	0	3 (.2)	25 (1.9)	144 (11.2)	542 (42.2)	434 (33.8)	118 (9.2)	16 (1.2)	1 (.1)	0	1283 (.8)
36	0	0	24 (.8)	138 (4.8)	804 (27.8)	1296 (44.9)	507 (17.5)	99 (3.4)	18 (.6)	3 (.1)	2889 (1.8)
37	0	0	16 (.3)	101 (1.8)	747 (13.1)	2300 (40.3)	1908 (33.4)	548 (9.6)	75 (1.3)	15 (.3)	5710 (3.6)
38	0	1 (.0)	6 (.0)	92 (.6)	903 (5.6)	4642 (29.0)	6959 (43.4)	2894 (18.1)	471 (2.9)	63 (.4)	16031 (10.0)
39	0	4 (.0)	2 (.0)	41 (.1)	808 (2.6)	6183 (19.7)	13822 (44.1)	8475 (27.1)	1739 (5.6)	244 (.8)	31318 (19.5)
40	1 (.0)	6 (.0)	5 (.0)	46 (.1)	899 (1.4)	8777 (13.5)	26707 (41.0)	21692 (33.3)	6142 (9.4)	839 (1.3)	65114 (40.6)
41	0	1 (.0)	3 (.0)	11 (.0)	225 (.7)	2840 (9.3)	11107 (36.5)	11582 (38.1)	4010 (13.2)	650 (2.1)	30429 (19.0)
42	0	1 (.0)	1 (.0)	1 (.0)	31 (.7)	449 (9.7)	1556 (33.7)	1730 (37.4)	726 (15.7)	128 (2.8)	4622 (2.9)
≥43	0	0	0	0	1 (.5)	30 (13.7)	79 (36.1)	76 (34.7)	31 (14.2)	2 (.9)	219 (.1)
TOTAL	7 (.0)	165 (.1)	708 (.4)	1470 (.9)	5739 (3.6)	27212 (17.0)	62810 (39.2)	47129 (29.4)	13213 (8.2)	1945 (1.2)	160398(100.0)

TABLE 2.4

BIRTHWEIGHT DISTRIBUTION (500 gm INTERVALS) FOR GESTATIONS ≥28 WEEKS: BIRTHS OF UNCERTAIN GESTATION, 1980-82

	<500 gms	500-999	1000-1499	1500-1999	2000-2499	2500-2999	3000-3499	3500-3999	4000-4499	≥ 4500	TOTAL
28	0	21 (23.3)	54 (60.0)	8 (8.9)	6 (6.7)	0	0	0	0	1 (1.1)	90 (.3)
29	0	5 (7.6)	39 (59.1)	18 (27.3)	3 (4.5)	1 (1.5)	0	0	0	0	66 (.2)
30	0	12 (8.9)	56 (41.5)	48 (35.6)	8 (5.9)	7 (5.2)	3 (2.2)	1 (.7)	0	0	135 (.4)
31	1 (.9)	5 (4.6)	27 (25.0)	58 (53.7)	12 (11.1)	3 (2.8)	2 (1.9)	0	0	0	108 (.3)
32	0	6 (2.8)	42 (19.9)	86 (40.8)	55 (26.1)	19 (9.0)	3 (1.4)	0	0	0	211 (.6)
33	0	2 (1.1)	21 (11.6)	62 (34.3)	75 (41.4)	16 (8.8)	4 (2.2)	1 (.6)	0	0	181 (.5)
34	1 (.3)	3 (.8)	11 (2.8)	95 (24.4)	183 (46.9)	75 (19.2)	18 (4.6)	3 (.8)	1 (.3)	0	390 (1.2)
35	0	0	7 (1.4)	64 (12.9)	188 (37.8)	172 (34.6)	49 (9.9)	16 (3.2)	1 (.2)	0	497 (1.5)
36	0	1 (.1)	16 (1.3)	55 (4.4)	335 (26.6)	520 (41.2)	280 (22.2)	46 (3.6)	8 (.6)	0	1261 (3.7)
37	0	1 (.1)	8 (.4)	46 (2.5)	250 (13.9)	754 (41.8)	573 (31.7)	154 (8.5)	16 (.9)	3 (.2)	1805 (5.3)
38	0	1 (.0)	2 (.0)	39 (.9)	283 (6.7)	1285 (30.6)	1766 (42.1)	714 (17.0)	98 (2.3)	10 (.2)	4198 (12.4)
39	0	0	1 (.0)	20 (.4)	192 (3.5)	1249 (22.7)	2424 (44.1)	1308 (23.8)	271 (4.9)	28 (.5)	5493 (16.2)
40	0	0	1 (.0)	20 (.2)	293 (2.3)	2055 (16.0)	5385 (41.9)	3919 (30.5)	1026 (8.0)	139 (1.1)	12838 (37.9)
41	0	1 (.0)	0	4 (.1)	56 (1.2)	557 (11.5)	1808 (37.2)	1777 (36.5)	582 (12.0)	79 (1.6)	4864 (14.4)
42	0	0	1 (.1)	1 (.1)	11 (.8)	136 (9.5)	492 (34.4)	533 (37.2)	216 (15.1)	42 (2.9)	1432 (4.2)
≥43	0	0	0	2 (.8)	4 (1.7)	28 (11.6)	97 (40.2)	88 (36.5)	20 (8.3)	2 (.8)	241 (.7)
TOTAL	2 (.0)	58 (.2)	286 (.8)	626 (1.9)	1954 (5.8)	6877 (20.3)	12904 (38.1)	8560 (25.3)	2239 (6.6)	304 (.9)	33810 (100.0)

those with certain gestations, whereas, 4.9 per cent of births of uncertain gestational age are estimated to have gestations above 41 weeks compared with 3.0 per cent of certain gestations. Similarly, births of uncertain gestational age have higher rates of birthweight below 2,500 gms (8.7 per cent) compared to those of certain gestation (5.0 per cent). A trend to lower birthweight amongst uncertain gestations is apparent within separate weeks. For example, at 38 weeks 7.6 per cent of births of uncertain gestational age were below 2,500 gms compared with 6.2 per cent of births of certain gestations, and at 40 weeks the figures were 2.5 per cent of uncertain gestations compared to 1.5 per cent of certain gestations.

These trends in low birthweight parallel the findings related to LMP of Hall et al (1985 a), reviewed in section 1.1.2, that women of uncertain LMP generally had less favourable outcomes. In the analyses described here births of uncertain gestational age were included, but certainty of gestation was not considered as a covariate.

2.2 Definition of Response Variables and Covariates in the Regression Analyses of SMR2 Data

2.2.1 Introduction

The relationship between maternal, socio-economic and biological covariates and perinatal outcome were examined in a series of regression analyses of SMR2 data. Only singleton births were included since perinatal outcome amongst multiple births is generally less favourable and atypical of the experience of the population as a whole. No one risk factor was the focus of attention and the covariates were given equal status in the analysis. Some aspects of obstetric history were only relevent to multiparae, and the analyses were performed separately for primiparae and multiparae.

2.2.2 Perinatal Outcome Variables

(1) Birthweight

The analysis of birthweight focussed on the lower tail of the distribution. Models were considered for the probabilities of birthweight below four cutpoints, 1,000 gms, 1,500 gms, 2,000 gms and 2,500 gms. Stillbirths may have died some time before delivery, during which time they may have lost weight. For this reason the 0.65 per cent of births to primiparae and the 0.56 per cent of births to multiparae that were stillborn were excluded from the analysis.

(2) Gestational age

Gestational age was grouped into ten categories, the first corresponding to births in the 28^{th} week of gestation, the second, births in the 29^{th} week, up to the tenth category which represented births in the 37^{th} or later weeks of pregnancy. The final category covers all term and posterm deliveries and the analysis thus concentrates on risks experienced throughout the preterm period. The reasons for treating the preterm period separately are discussed in chapter 4. The analysis was restricted to deliveries of gestational age 28 weeks or more. SMR2 information was only available for livebirths before this period and these form an unrepresentative sample of pregnancies terminating in earlier weeks.

(3) Birthweight standardised for gestational age

A categorical variable measuring birthweight standardised for gestational age was constructed by comparing each infant's birthweight to standard percentiles of the birthweight distribution at the infant's gestational age. The standardisation was based on Scottish singleton livebirths from the SMR2 scheme in 1975 to 1979 for the 28^{th} to 42^{nd} week of gestation (Smalls & Forbes, 1983). The small number of births before this period were compared to percentiles for the 28^{th} week, and births born after the period were compared to percentiles for the 42^{nd} week. Percentiles were prepared directly from the birthweight distribution and were smoothed across gestational ages using third order polynomial regression, weighted according to the frequency at each gestational age. Standard values of the 5^{th}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} and 95^{th} percentiles of the birthweight distribution were available (see Appendix 2). The outcome variable considered in the regression study utilised all eight categories. Separate standards were available for primiparae and multiparae, and for male and female infants.

2.2.3 Covariates

The variables chosen from the SMR2 scheme for inclusion as covariates in the study were, with the exception of sex of infant, characteristics of the mother which would not normally change during the course of her current pregnancy. The sex of the infant, though unknown, is also fixed from conception. Maternal age was included as a covariate although it could be argued that the variable is to some extent dependent on the outcome of the current pregnancy, since a woman delivering preterm will be slightly younger at the time age is coded on SMR2 than an equivalent mother of a full term infant. This dependency has been ignored in the analysis. Variables giving medical diagnoses made during pregnancy or after delivery and those describing the care received by the woman were not considered.

Table 2.5 presents the covariates, their categorisation, and their distribution for singleton livebirths with complete information. The first row contains figures relating to sex of infant. This covariate was excluded from the analysis of birthweight standardised for gestational age because sex was a standardising variable. Two socio-economic variables, the mother's marital status and her social class were included in the study. Marital status was coded single versus married, widowed, divorced, separated or other. Social class is based on the Registrar General's scale I-V derived from the occupation of the husband or father of the child and was categorised I-II (professional and managerial), III (clerical and skilled manual) and IV-V (semi- and unskilled manual). Social class is reported as unknown on SMR2 when there is insufficient information to assign a social class or when the occupation is stated as the armed forces. This category represents a substantial portion of both primiparae (23.3 per cent) and multiparae (19.7 per cent), and these cases were not excluded from the analysis. Cole (1983) compared the rate of unknown social class on SMR2 in 1977 (15.8 per cent) with that from the Registrar General's data on the obstetric population (1.2 per cent). The comparison showed that the SMR2 scheme is considerably less successful than the Registrar General in obtaining occupational data.

The next two variables describe demographic features of the population, maternal height was coded <150 cm, 150-164 cm and ≥165 cm. The first category, <150 cm, identifies very short women who are known to experience increased risk of poor pregnancy outcome. The categorisation of maternal age also identifies two groups of women, <18 years and ≥35 years, who are thought to experience increased risks during pregnancy, the

TABLE 2.5

COVARIATES, THEIR CATEGORIES AND THE FREQUENCY DISTRIBUTION FOR PRIMIPARAE & MULTIPARAE in 1980-82*

COVARIATE	CATEGORIES	FREQUENCY DISTRIBUTION (%)	
		PRIMIPARAE	MULTIPARAE
Sex of Infant	Male	51.4	51.2
	Female	48.6	48.8
Marital Status	Married**	85.5	97.6
	Single	14.5	2.4
Social Class	I-II	17.4	18.4
	III	38.6	39.6
	IV-V	20.6	22.3
	Unknown***	23.3	19.7
Maternal Height	<150 cm	3.3	3.8
	150-164 cm	71.0	72.0
	≥165 cm	25.7	24.2
Maternal Age	<18 yrs	6.2	.2
	18-24 yrs	55.6	29.2
	25-34 yrs	36.3	62.4
	≥35 yrs	1.9	8.2
Previous Spontaneous Abortion	0	91.0	79.9
	1	7.6	15.6
	≥2	1.3	4.5
Previous Induced Abortion	0	93.9	94.0
	≥1	6.2	6.0
Previous Caesarean Section	0	-	90.7
	≥1	-	9.3
Previous Perinatal Death	0	-	95.3
	≥1	-	4.7
Previous Livebirths***	0-2	-	88.5
	≥3	-	11.5
TOTAL*		100.0% (n = 80,276)	100.0% (n = 106,362)

* Livebirths with complete data on the covariates.

** Married, Widowed, Divorced, Separated, Other.

*** Based on Registrar General's Scale I-V of fathers' occupation. Cases not recorded (U.K.) reflect either unemployment or insufficient information on occupations.

**** Previous livebirths surviving the first weeks of life.

other two categories, 18-24 years and 25-34 years comprise the majority of the population.

The remaining covariates describe the woman's obstetric history. A history of spontaneous abortion was examined by identifying two groups of women, those with one previous abortion and those with $\geq$2. There were very few women with $\geq$2 previous induced abortions and this covariate was dichotomised 0 versus $\geq$1, as were two other aspects of obstetric history, previous perinatal death and caesarean section. The final covariate, number of previous livebirths, was calculated as total pregnancies minus previous abortions and perinatal deaths, and, more accurately, would be described as the number of previous livebirths surviving the first week of life. This covariate was categorised 0-2 and $\geq$3. By considering a high number of previous births surviving the first week of life rather than high parity, the risk associated with $\geq$3 births should not be confounded by compensation for previous perinatal losses.

In the regression analysis one category of each covariate was chosen as a reference and the risks are presented in comparison with this category. In general, the reference was chosen to be the category with highest frequency in the population. The one exception was maternal age, where the reference category corresponded to the most frequent category for primiparae but not multiparae. The reference categories for the covariates were as follows, sex of infant, male; marital status, married; social class, III; maternal height, 150-164 cm; maternal age, 18-24 years; previous spontanous abortions, 0; previous induced abortions, 0; previous caesarean section, 0; previous perinatal deaths, 0; previous livebirths, 0-2.

Risks of birthweight below 2,500 gms and 1,500 gms, of preterm delivery and of the birth of an SGA or LGA infant for each covariate are given in Table 2.6 for the study population of 1980-82. It can be seen that there are considerable variations in risk associated with some covariates. The table shows unadjusted patterns of risk. In chapter 4 the extent to which the association with each covariate can be explained by intercorrelation with the others is examined.

2.2.4 Missing Data

An estimate of gestational age was available in all cases, but a few births (0.16 per cent amongst primiparae and 0.18 per cent amongst multiparae) did not have birthweight coded on the SMR2 document and were excluded from the analysis of birthweight and birthweight standardised for gestational age.

Table 2.7 gives the numbers and rates of missing data for each of the covariates in the study expressed as a percentage of total liveborn singleton deliveries. Information on maternal height was missing for approximately five per cent of both primiparae and multiparae; and on marital status for less than 0.1 per cent. Although there is no provision for coding unknown obstetric history, 26 SMR2 documents contained a series of values 9 relating to numbers of previous abortions, caesarean sections and perinatal deaths and were excluded from the analysis. After excluding births with missing information on birthweight or at least one covariate (excepting social class) data were available for 95 per cent of singleton livebirths to both primiparae and multiparae.

TABLE: 2.6

POPULATION DISTRIBUTION, PERCENTAGE BIRTHWEIGHT: <2500gm, <1500 gm, PRETERM, SGA & LGA SINGLETON DELIVERIES, PRIMS & MULTIPS 1980-82 *

COVARIATES	PRIMIPARAE					MULTIPARAE				
	Birthweight < 2500 gm	Birthweight < 1500 gm	Preterm ≤ 36th week	≤ 10th percentile	> 90th percentile	Birthweight < 2500 gm	Birthweight < 1500 gm	Preterm ≤ 36th week	≤ 10th percentile	> 90th percentile
Sex of Infant										
Male	5.9	.8	5.8	-	-	4.1	.5	4.5	-	-
Female	6.9	.8	5.1	-	-	5.1	.5	4.0	-	-
Marital Status										
Married	6.0	.7	5.0	8.8	10.4	4.5	.5	4.2	8.5	11.0
Single	8.4	1.0	7.9	11.0	7.8	8.7	.8	8.5	13.8	6.7
Social Class										
I-II	4.7	.6	4.4	6.6	11.3	2.7	.3	3.0	5.5	13.4
III	6.1	.7	5.2	8.9	10.7	4.4	.4	4.0	8.2	11.2
IV-V	7.2	.9	5.8	12.7	9.3	5.4	.5	4.8	10.1	9.7
Unknown	7.3	.9	6.5	10.0	8.6	5.9	.6	5.4	10.8	9.1
Maternal Height										
≤150 cm	14.4	1.2	9.1	19.8	3.3	10.3	.8	5.9	19.1	4.0
150-164 cm	6.8	.8	5.7	10.0	8.4	4.9	.5	4.4	9.4	9.3
≥165 cm	4.2	.6	4.5	5.3	15.4	2.8	.4	3.6	4.9	16.4
Maternal Age										
<18	8.4	1.3	9.3	9.1	10.1	12.2	1.4	13.2	12.1	8.4
18-24	6.4	.9	5.3	9.7	9.4	5.3	.5	4.9	10.2	8.5
25-34	5.8	.7	4.9	8.2	10.8	4.1	.4	3.8	7.9	11.6
≥35	9.7	1.4	8.4	8.4	12.9	5.6	.7	5.3	8.8	13.3
Prev. Spontaneous Abortion										
0	6.2	.7	5.3	9.1	10.0	4.2	.4	3.9	8.4	10.8
1	7.7	1.2	6.6	9.2	10.7	5.2	.6	5.1	9.0	11.4
≥2	11.7	2.2	9.5	5.0	12.1	8.7	1.5	7.6	11.9	9.9
Prev. Induced Abortion										
0	6.3	.7	5.3	9.1	10.1	4.5	.4	4.1	8.6	10.9
≥1	7.5	1.1	7.2	8.8	9.7	6.1	.8	6.3	9.2	10.4
Prev. Caesarean Section										
0	-	-	-	-	-	4.5	.4	4.2	8.6	10.7
≥1	-	-	-	-	-	5.6	.6	4.8	9.0	12.4
Prev. Perinatal Death										
0	-	-	-	-	-	4.3	.4	4.0	8.5	6.0
≥1	-	-	-	-	-	10.4	1.7	9.7	11.7	8.9
Prev. Livebirth										
0-2	-	-	-	-	-	4.5	.5	4.1	8.5	10.7
≥3	-	-	-	-	-	5.6	.5	6.9	9.6	12.2
TOTAL	6.4	.8	5.4	9.1	10.1	4.6	.5	4.3	8.7	10.9

* Livebirths with complete data on all variables.

TABLE 2.7

MISSING COVARIATE INFORMATION IN LIVEBORN SINGLETON BIRTHS TO PRIMIPARAE & MULTIPARAE: 1980-82

	PRIMIPARAE		MULTIPARAE	
Sex of Infant	4	(0.00%)	1	(0.00%)
Marital Status	56	(0.07%)	99	(0.09%)
Social Class	20216	(23.92%)	22711	(20.28%)
Maternal Height	4175	(4.94%)	5541	(4.95%)
Maternal Age	0		13	(0.01%)
Previous Spontaneous Abortion	0		0	
Previous Induced Abortion	0		0	
Previous Caesarean Section	0		0	
Previous Perinatal Death	0		0	
Previous Livebirth	0		0	
	n = 84500	(100.00%)	n = 111984	(100.00%)

Births with maternal height unknown comprised the largest source of exclusions from the analysis, and the rate of low birthweight and preterm delivery were investigated for these infants. First, amongst births to primiparae of unknown height 11.1 per cent had birthweight below 2,500 gms compared with 6.8 per cent amongst births to primiparae of known height. Births to primiparae of unknown height were also more likely to be born preterm with a risk of 12.0 per cent, compared to 5.8 per cent amongst primiparae of known height. Similarly births to multiparae of unknown height experienced a preterm delivery rate of 8.0 per cent and a risk of birthweight below 2,500 gms of 7.6 per cent compared with the preterm delivery rate of 4.6 per cent and risk of birthweight below 2,500 gms of 4.9 per cent amongst multiparae of known height. These differentials between the study population and the excluded cases represent a potential source of bias in the findings related to maternal height and the other covariates.

2.3 Contingency Table Form of the SMR2 Data

After the covariates were categorised it was possible to summarise the data from individual women into a concise form for the purpose of fitting regression models. During 1980 to 1982 there were 84,668 liveborn singleton births to primiparae. One or more of the covariates or birthweight was missing in 4392 cases and the remaining 80,276 complete cases were considered in the analysis. The seven covariates relating to primiparae could occur in 1152 (the product of the number of levels of each covariate) combinations. Only 658 combinations were in fact observed. The absence of many possible combinations in the population reflects the severity of some of the covariate categories and also intercorrelation between the covariates.

The summarised data file on birthweight for primiparae is shown in Table 2.8. The frequencies with which each covariate combination were observed are given in the second column and the birthweight distribution is given in the final five columns. The covariate columns give, in order, marital status, the number of previous spontaneous and induced abortions, maternal height, sex of infant, maternal age and social class. A value of 1 corresponds to the reference category of each covariate. Higher values in the covariate columns correspond to other categories in descending order in Table 2.5. The summarised data file was created by reading through the SMR2 data for each year. A new combination was added to the data file or the frequency of an existing combination increased by one, for each primiparae with complete data. Covariate combinations with high frequencies were likely to be observed early on and come near the top of the summarised data file. The most frequently occurring combination (number 17) corresponded to 5178 women (6.5 per cent). Moving down the data file the covariate combinations were observed less often. None of the final hundred combinations was observed for more than five women. In Table 2.9 first and last covariate combinations in the corresponding summarised birthweight data for the 106,362 multiparae with complete data are given. Ten covariates were considered in the analyses of multiparae, and these could occur in 5264 possible combinations of which 2096 were observed.

Summarised data files were created in an identical fashion for the analyses of gestational age and birthweight standised for gestational age. A description of all the data files is given in Table 2.10. The first row describes the data files for the

TABLE 2.8

SUMMARY BIRTHWEIGHT DATA FOR PRIMIPARAE 1980-82

Cell No.	Cell frequency	Co-variates	Birthweight Distribution				
			<1000g	1000-1499g	1500-1999g	2000-2499g	≥2500g
1	859	1113213	1	3	9	27	819
2	1265	2111114	4	6	20	60	1175
3	172	2111121	0	3	4	7	158
4	483	1113112	1	0	1	10	471
5	323	2113114	2	3	4	11	303
6	4972	1111211	11	20	31	242	4668
7	758	2111124	2	5	16	37	698
8	1	1212143	0	1	0	0	0
9	208	1121131	1	1	4	8	194
10	17	1212113	1	1	0	2	13
11	173	2111221	0	5	3	12	153
12	617	2111113	3	5	7	24	578
13	1618	1113211	5	6	12	45	1550
14	78	1311131	1	1	1	6	69
15	19	1111122	0	0	0	1	18
16	3315	1111131	5	24	31	125	3130
17	5178	1111111	8	23	40	227	4880
18	2987	1111113	7	16	36	123	2805
19	2836	1111213	5	18	27	151	2635
20	59	1111244	0	1	1	6	51
21	1366	1113131	0	4	9	26	1327
22	218	1112111	1	1	4	20	192
23	37	1311234	1	0	0	2	34
24	1206	1111133	3	8	20	65	1110
25	796	1113114	0	7	7	33	749
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
634	1	2122121	0	0	1	0	0
635	1	2321213	0	0	0	0	1
636	1	1122244	0	0	0	0	1
637	1	1312133	0	0	0	0	1
638	1	1323133	0	0	0	1	0
639	1	2123231	0	0	0	0	1
640	1	1123244	0	0	0	0	1
641	1	1122142	0	0	0	0	1
642	1	2211142	0	0	0	0	1
643	1	1223144	0	0	0	0	1
644	2	1223114	0	0	0	0	2
645	1	1123223	0	0	0	0	1
646	1	2313114	0	0	0	0	1
647	1	2213134	0	0	0	0	1
648	2	2221113	0	0	0	0	2
649	1	2223112	0	0	0	0	1
650	1	1212223	0	0	0	0	1
651	1	2223113	0	0	0	0	1
652	1	2211143	0	0	0	0	1
653	1	2123142	0	0	0	0	1
654	1	1322111	0	0	0	0	1
655	1	1221142	0	0	0	0	1
656	1	1312134	0	0	0	0	1
657	1	2313113	0	0	0	0	1
658	1	1312112	0	0	0	0	1

TABLE 2.9

SUMMARY BIRTHWEIGHT DATA FOR MULTIPARAE 1980-82

Cell No.	Cell frequency	Co-variates	Birthweight Distribution				
			<1000g	1000-1499g	1500-1999g	2000-2499g	≥2500g
1	2323	1111111113	5	8	19	63	2228
2	273	1111121241	1	0	3	13	256
3	903	1111113111	1	2	6	11	883
4	71	1211113114	1	0	0	2	68
5	30	1311211131	2	0	1	4	23
6	132	1211113233	0	3	1	4	124
7	34	1112121134	1	0	0	0	33
8	27	1112112214	1	1	0	2	23
9	245	1111211231	2	6	5	10	222
10	32	1121121133	1	0	0	3	28
11	18	1311113111	0	1	1	0	16
12	724	1111113234	2	1	3	13	705
13	5740	1111111131	1	10	29	130	5570
14	2324	1111111233	1	4	19	80	2220
15	763	1111113134	1	2	2	8	750
16	66	1211211231	2	0	0	6	58
17	2880	1111111211	3	3	15	105	2754
18	5505	1111111231	6	9	41	182	5267
19	66	1111211214	0	2	0	9	55
20	5	2311111133	1	0	0	0	4
21	611	1111113113	2	2	2	11	594
22	13	1311211234	1	1	0	2	9
23	2	1321212213	1	0	0	0	1
24	1831	1111111214	4	4	19	71	1733
25	1154	1211111131	2	1	11	24	1116
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
2072	1	1312121132	1	0	0	0	0
2073	2	2122111134	0	0	0	0	2
2074	1	2121111124	0	0	0	0	1
2075	1	2221113234	0	0	0	0	1
2076	1	1222111144	0	0	0	0	1
2077	1	1222112131	0	0	0	0	1
2078	1	1112213243	0	0	0	0	1
2079	1	1321121212	0	0	0	0	1
2080	1	1121213212	0	0	0	0	1
2081	1	1221113112	0	0	0	0	1
2082	1	1322113232	0	0	0	0	1
2083	1	1312212213	0	0	1	0	0
2084	1	2112113134	0	0	0	0	1
2085	1	1121213224	0	0	0	0	1
2086	1	2111211113	0	0	0	0	1
2087	1	2121123113	0	0	0	0	1
2088	1	1211121224	0	0	0	0	1
2089	1	1212113244	0	0	0	0	1
2090	1	2121123114	0	0	0	0	1
2091	1	1212223243	0	0	0	0	1
2092	1	1211121212	0	0	0	0	1
2093	1	2111113121	0	0	0	0	1
2094	1	1312112134	0	0	0	0	1
2095	1	1312121214	0	0	0	0	1
2096	1	1312111244	0	0	0	0	1

TABLE 2.10

CONTINGENCY TABLES FOR EACH ANALYSIS

Response Variable	Number of Categories of Response Variable	Years of Study	PRIMIPARAE		MULTIPARAE		Notes
			Number of Births	Number of Distinct Covariate Combinations	Number of Births	Number of Distinct Covariate Combinations	
Birthweight	2	1980-82	80,276	658	106,362	2096	1. Singleton livebirths. 2. Analysis repeated for birthweight $\leqslant$1000 gms, $\leqslant$1500 gms, $\leqslant$2500 gms
Preterm Gestational Age	10	1981	27,014	514	36,770	1521	1. Singleton deliveries of gestational age $\geqslant$28 completed weeks.
Birthweight Standardised for Gestational Age	8	1980-82	80,279	368	106,363	1276	1. Singleton livebirths. 2. Birthweight also standardised for sex of infant.
1. Term and Post-term	8		75,900	356	101,852	1233	
2. Preterm Deliveries	8		4,379	237	5,511	603	

birthweight analyses described above. After an initial examination of birthweight standardised for gestational age based on all deliveries, more detailed results were obtained for preterm deliveries and term and post term deliveries separately.

2.4 The Scottish Morbity Record 11 (SMR11)

2.4.1 Description of the SMR11 Document

Information covering the period between birth and discharge from neonatal services is collected in the SMR11 data scheme. An SMR11 file is only opened for a livebirth, it includes information collected during the infant's stay in labour and other wards and is closed when the infant is discharged home, transferred to a general medical ward, discharged to foster care, dies, or the file may be closed for other reasons. If an infant is transferred to a neonatal ward in another hospital the document should also be passed on and information collected until eventual discharge. Hospitals have a choice between completing a full or an abbreviated version of the document, the abbreviated version was used here.

Appendix 1 contains a copy of the abbreviated SMR11 document used in 1980. The information collected on SMR11, includes identification data, details of the birth, several types of care the infant may have received, a summarised medical record, provision for including up to nine pathological conditions the infant may have suffered and up to two operations. Finally administrative discharge data are collected.

2.4.2 Coverage and validity

The SMR11 scheme is more recent than SMR2 and has not been so widely accepted. There remain several hospitals that do not participitate in the scheme. In 1980 the coverage, within participating hospitals, was 99 per but these births comprised only 76 per cent of the livebirths reported by the Registrar General. A further difference between the two schemes is that the SMR11 documents are completed directly by medical, rather than clerical staff and there should be less chance of transcription and other errors. Less validation work has been carried out on the SMR11 than the SMR2 scheme and coding policies may differ across hospitals. For example, congenital malformation rates vary across hospitals, primarily because of differences in coding policy (Cole, 1983).

2.4.3 Definition of Variables in the Neonatal Classification

The classification was based on eleven categorical variables, listed in the first column of Table 2.11. Birthweight was divided into the four categories $\leq$1,500 gms, 1,501-2,000 gms, 2,001-2,500 gms and >2,500 gms. Birthweight standardised for gestational age was dichotomised as SGA and otherwise. The Apgar score (Apgar, 1953) is a dichotomisation of a widely adopted scale used to asses the general well being of infants at birth. Recurrent apnoea, jaundice and convulsions were dichotomised present and absent. Three medical interventions, resuscitation (absent, present but not intubation, intubation), assisted ventilation and tube feeding were also included in the analysis. Finally the status (alive, dead) of the infant and age at discharge (<3 days, 4-10 days and >11 days) from neonatal services were included.

In 1980, 672 births were recorded as having one of a number of serious congenital abnormalities and were excluded from the analysis as their treatment is atypical. One or more of the eleven variables was missing in a further 6596 births and the main analysis considered only the remaining 45,426 complete

TABLE 2.11

CLASSIFICATION VARIABLES, CATEGORISATION AND POPULATION DISTRIBUTION FOR COMPLETE CASES: SMR11 1980

VARIABLE	CATEGORIES	FREQUENCY (%)
Birthweight	>2500 gms	93.5
	2001–2500 gms	4.6
	1501–2000 gms	1.3
	≤1500 gms	0.7
Birthweight Standardised for Gestational Age	≥10th Percentile	90.0
	<10th Percentile	10.0
Apgar Score at 5 min.	≥7	99.5
	<7	0.5
Resuscitation	None	87.5
	Intermediate*	9.1
	By Intubation	3.4
Assisted Ventilation after 30 min.	Absent	99.2
	Present	0.8
Recurrent Apnoea	Absent	99.5
	Present	0.5
Jaundice	Absent	69.9
	Present**	30.1
Convulsions	Absent	99.7
	Present	0.3
In Tube Feeding	Absent	97.1
	Present	2.9
Dead at Discharge	Absent	99.6
	Present	0.4
Age at Discharge	<3 days	12.9
	4–10 days	79.5
	>11 days	7.6
TOTAL		100.0% (n = 45426)

* Mask + Intermittent Positive Pressure Ventilation, drugs only, other.

** >86 µmol/litre bilirubin

cases. The eleven variables defined a contingency table in which there were 9216 possible outcomes, although only 600 were observed amongst the complete cases. The two most frequently observed outcomes accounted for 44 per cent and 19 per cent of the cases.

2.5 Linked SMR2-SMR11 Data

There is some overlap in the information collected on the two records which can be exploited to compile a linked SMR2-SMR11 data file. One of the objectives of the Information Services Division when introducing the SMR11 scheme was to promote the development of such a linkage, but lack of resources has delayed the routine production of a linked data file. The following results are taken from Smalls et al (1987) describing a linkage of the data for 1981-82, and demonstrating the accuracy of information in the two schemes.

The SMR2 and SMR11 files were matched on six items of information; hospital of birth, date of birth, initial of surname, maternal reference number, the sex of infant and birthweight (to within 50 gms). Stillbirths, which were only available on SMR2, and multiple births, for which there may be several SMR11 documents corresponding to one SMR2 document, were excluded from the linkage. Any cases that were not uniquely matched with all six items identical on the two files were investigated for a possible match on only five out of the six items. Mother-child pairs with two or more items differing were not considered.

The results of the linkage can be seen in Table 2.12. Approximately 96 per cent of singleton livebirths born in hospitals participating in the SMR11 scheme were matched. Amongst cases where a match on only five out of the six items was achieved, a difference in maternal reference number accounted for the majority in 1981 and a large proportion in 1982. Maternal reference number involves a large number of meaningless digits and it is likely to have a relatively high transcription error rate.

Apgar score at five minutes was available on both the SMR2 and SMR11 schemes but was not used in the linkage. The percentages of linked cases having identical Apgar scores are given amongst cases that matched on all six items, on all but the reference number and on all but any other matching item in Table 2.13. In 1981 approximately 95 per cent and in 1982 approximately 92 per cent had identical Apgar scores and these figures did not vary substantially over match qualities. In both years 0.28 per cent of cases were considered to have different Apgar Scores which suggests a 0.14 per cent error rate for Apgar score on either scheme.

TABLE 2.12 NUMBER OF LINKED CASES IN 1981 AND 1982

	1981	1982
Total SMR 2 records	67890	64966
Total SMR 11 records	54300	55400
Linkable* SMR 2 records	53386	54726
Linkable* SMR 11 records	53224	54401
Quality of Unique Match		
Matched on all 6 items	44979	49041
Not matched on maternal reference no.	4600	1093
Matched on hospital of birth and 4 of remaining 5 items	1671	2078
TOTAL LINKED	51250	52212
Percentage linkable* SMR 2	96.0	95.4
Percentage linkable* SMR 11	96.3	96.0

* All live singleton births in hospitals participating in SMR11 scheme.

TABLE 2.13

CONCORDANCE OF APGAR SCORES BETWEEN SMR2 and SMR11

APGAR SCORE AT 5 MINUTES

Quality of Match*	1981 Equal	1981 Not equal, but both in Range 8-10	1981 Not Equal	Total
A	94.44	5.29	0.27	87.77
B	96.59	3.13	0.28	8.97
C	95.21	4.31	.48	3.26
TOTAL	94.65	5.07	0.28	100.00%

n = 51250

Quality of Match	1982 Equal	1982 Not equal, but both in Range 8-10	1982 Not Equal	Total
A	92.30	7.44	0.06	93.93
B	91.58	7.96	0.46	2.09
C	91.87	7.46	0.67	3.98
TOTAL	92.27	7.45	0.28	100.00%

n = 52212

Percentage of total linked file for each year

* A = Matched on all 6 items

B = Not matched on maternal reference no.

C = Matched on hospital of birth and 4 of remaining 5 items

CHAPTER 3 : REGRESSION MODELS FOR CATEGORICAL DATA

3.1 Introduction

3.1.1 Notation

This chapter describes a variety of regression models and techniques for use with categorical data. The first four sections introduce models that were used to examine the relationships between birthweight, gestational age and birthweight standardised for gestational age and the ten covariates defined in section 2.2.3. The final section describes some alternative approaches and, where appropriate, draws a comparison with methods that were used.

The following notation has been used throughout. A response variable Z depends on several covariates in a (p+1)x1 vector of explanatory variables $\mathbf{x}=(x_0,x_1,\ldots,x_p)^T$. The response can be continuous or categorical, but the development here is primarily concerned with binary data, where Z takes values 1 or 0 representing positive and negative results respectively, and ordered categorical data measuring contiguous intervals on a continuous underlying scale where Z takes values 1 up to k, the number of response categories. In the models for scalar Z the first component of $\mathbf{x}$, x_0, is set identically equal to 1 and represents the mean value of Z when $x=(1,0,0,\ldots,0)^T$. The remaining components of x may be continuous or they may represent the levels of a categorical covariate. A categorical covariate taking r levels is included in the model by defining a set of r-1 dummy variables. One level is taken to be the reference while for each of the other levels the corresponding dummy variable takes value 1 when an individual falls in that category and 0 otherwise. The response variable depends on the covariates via a linear combination, $\sum_{s=0}^{p} x_s\beta_s = \mathbf{x}^T\beta$, of the covariates and a (p+1)x1 vector of unknown parameters, β. The covariate vector can be expanded to include any function of $\mathbf{x}$, for example quadratic or higher order powers of continuous covariates, but in the models considered here the components of the parameter vector, β, always enter the model linearly.

The parameter vector is estimated from a sample of N observations. When both the Z and $\mathbf{x}$ are categorical, and particularly for large N, there may be many individuals with identical covariate values and a concise indexation of the observations can be obtained by grouping cases with distinct covariate combinations. Defining a response variable, Y_i, for each distinct covariate combination, $1 \leq i \leq I$, the data are listed as $(y_i, N_i, \mathbf{x}_i)$, where the random variable Y_i takes value y_i, the proportion of individuals with binary response Z=1 amongst the N_i individuals for whom $\mathbf{x}=\mathbf{x}_i$. Quantity I is the number of distinct covariate combinations and $\sum_{i=1}^{I} N_i = N$. When the data are polytomous the response for each covariate combination is vector valued and can be defined in several ways. In this development the response is described in terms of the cumulative relative frequencies over categories of the response. Finally, the (p+1)xI matrix $\mathbf{X}=(\mathbf{x}_1,\ldots,\mathbf{x}_i,\ldots,\mathbf{x}_I)$ contains the covariate information from the entire sample.

3.1.2 Generalised Linear Models

This section introduces a class of regression models for scalar response data described by Nelder & Wedderburn (1972). Results for the class as a whole are given, but later particular emphasis is laid on the logistic regression model which is a member of the class.

There are three properties that define a generalised linear model; these are the error structure, the linear component and the link function. First, the probability density function of the response Y is given by

$$f_Y(y;\theta,\phi) = \exp[\{y\theta-b(\theta)\}/a(\phi)+c(y,\phi)],$$

where θ is an unknown parameter, and the parameter ϕ may be known or unknown. When ϕ is known, and in some cases when ϕ is unknown, such distributions are members of the exponential family. By varying the functions a(.), b(.) and c(.,.), many commonly used probability distributions can be achieved (for example the three described in Table 3.1).

Following the development of McCullagh & Nelder (1983), letting $L(\theta,\phi;y)$ represent the logarithm of the likelihood from a single observation, y, and probability density function given by the above formula, and using the standard results

$$E(\partial L/\partial\theta) = 0$$

and

$$E(\partial^2 L/\partial\theta^2) = -E(\partial L/\partial\theta)^2,$$

the following forms for the mean and variance of Y can be obtained. First,

$$\partial L/\partial\theta = \{y-b'(\theta)\}/a(\phi),$$

where a prime denotes differentiation with respect to θ, so that

$$E(Y) = \mu = b'(\theta)$$

and

$$\partial L/\partial\theta = \{y-\mu\}/a(\phi).$$

Then

TABLE 3.1

Specifications for Three Common Distributions

	Binomial	Poisson	Normal
Range of Y	$\frac{0(1)N}{N}$	$0(1)\infty$	$(-\infty,\infty)$
$a(\cdot)$	$1/N$	1	ϕ
$b(\cdot)$	$\ln(1+e^{\theta})$	e^{θ}	$\frac{1}{2}\theta^2$
$c(\cdot,\cdot)$	$\ln[\binom{N}{Ny}]$	$-\ln y$	$-\frac{1}{2}(\frac{y}{\phi}\ln(2\Pi\phi))$
$E(Y)=\mu=b'(\theta)$	$e^{\theta}/(1+e^{\theta})$	e^{θ}	θ
$Var(Y)=a(\phi)b''(\theta)$	$\mu(1-\mu)$	μ	ϕ

$$E(\partial^2 L/\partial\theta^2) = -E[b''(\theta)/a(\phi)] = -b''(\theta)/a(\phi)$$

and

$$-E(\partial L/\partial\theta)^2 = -Var(Y)/a(\phi)^2,$$

so that

$$Var(Y) = a(\phi)b''(\theta).$$

The variance of Y is a function of θ (and hence μ) but also of the term $a(\phi)$.

In Table 3.1 the functions a(.), b(.) and c(.,.) and the mean and variance of Y are given for three member probability distributions. The first column describes the Binomial distribution for use with data in the form of proportions, and the mean represents the proportion of individuals with a positive response in a sample of N. The log likelihood in terms of μ for Binomially distributed data is given by

$$L(\mu;y) \propto Ny\ln[\mu/(1-\mu)]+N\ln(1-\mu),$$

ignoring the final term $c(y,\phi)=\ln\binom{N}{Ny}$.

The above derivation refers to a single observation (or distinct covariate combination) the mean of which is a function of one parameter θ. When the data consist of a Ix1 response vector, **Y** taking values **y**, parameter θ is replaced by a Ix1 parameter vector, $\boldsymbol{\theta}$, with one component, θ_i, corresponding to each observation, Y_i, $1 \leq i \leq I$. Sufficient generality is usually achieved by restricting $a(\phi)$ to be of the form

$$a_i(\phi) = \phi/w_i,$$

where the w_i are prior weights, known in advance, and ϕ is a scalar (for example, σ^2 in a Normal regression model).

The number of parameters can be reduced, and the model simplified, by considering each θ_i to be a function of a $(p+1)\times 1$, $(p+1)<I$, vector of parameters β related to the covariate vector. In particular θ_i is assumed to be a function of the linear combination $\mathbf{x}_i^T\beta$. The linear combination $\eta = \mathbf{x}^T\beta$ is the second component of a generalised linear model and later sections deal with the selection of covariates to be included in η.

The third component of a generalised linear model is the link function, g(.), a monotonic increasing function describing the relationship between the linear combination η_i and μ_i

$$\eta_i = g(\mu_i).$$

The link function can be chosen to suit different applications, but particular functions are commonly used in conjunction with each of the distributions in Table 3.1. In the case of the Binomial distribution the mean μ is constrained by the inequalities $0<\mu<1$, and a desirable property of a suitable link function is that it should map the unit interval (0,1) onto $(-\infty,\infty)$, so that η can take arbitarily small or large values. Three functions commonly used in this context are

1. logit

$$\eta = \log\{\mu/(1-\mu)\};$$

2. probit

$$\eta = \Phi^{-1}(\mu)$$

where $\Phi(.)$ is the Normal cumulative distribution function;

3. complementary log-log

$$\eta = \log\{-\log(1-\mu)\}.$$

When the logit link function is used with Binomial error structure the log likelihood, in terms of β, is given by

$$L(\beta;\mathbf{y}) = \sum_{i=1}^{I} \{N_i y_i \mathbf{x}_i^T\beta - N_i \ln[1+\exp(\mathbf{x}_i^T\beta)]\}.$$

For this model the parameter θ is equal to the linear component of the model

$$\theta_i = \eta_i, \qquad 1 \leq i \leq I.$$

The function that produces the above equality for a probability distribution is known as the canonical link function, and the combination of a probability distribution with its canonical link function produces a simple form (of $\mathbf{x}^T\beta$ multiplied by y) for the term in the likelihood including both the parameter β and $\mathbf{y}$. The canonical link function for the Normal distribution is given by $\eta = \mu$, and that for the Poisson distribution is given by $\eta = \ln(\mu)$.

3.1.3 Fitting Generalised Linear Models in GLIM

Regression models for binary data were fitted here using GLIM (Baker & Nelder, 1978). In GLIM, the likelihood equations are solved for β by the method of Fisher scoring. Given an initial estimate of β, β^o, an updated estimate, β^*, can be calculated from the formula

$$\{E(\partial^2 L/(\partial\beta_i \partial\beta_j))\}|_{\beta^o}(\beta^* - \beta^o) = \{\partial L/\partial\beta_i\}|_{\beta^o}$$

and the procedure is repeated till convergence. The parameter ϕ cancels out in the above calculations and the likelihood is not maximised over ϕ. If the model is based on the canonical link function for the probability distribution being used, the expected and observed information are equal and the procedure is equivalent to the Newton-Raphson algorithm.

The procedure is equivalent to a series of weighted least-squares regressions of a modified dependent variable $\tilde{Y}$ on $\mathbf{X}$ (Nelder & Wedderburn, 1972). The dependent variable, $\tilde{Y}$, has components

$$\tilde{Y}_i = \eta_i + (y_i - \mu_i)\{d\eta/d\mu\}_i$$

and the weights, $\mathbf{W}=\text{Diag}\{W_1,\ldots,W_I\}$, are given by

$$W_i = \{1/b''(\theta)\}_i\{d\mu/d\eta\}_i^2.$$

In the first stage of the iteration suitable starting values for μ, such as $\mu=\mathbf{y}$, have to be chosen. At later stages μ is calculated from the current β^*. When the model has Binomial error structure the modified dependent variable and weights are

$$Y_i = \eta_i + (y_i - \mu_i)/[\mu_i(1-\mu_i)]$$

and

$$W_i = N_i\mu_i(1-\mu_i).$$

At the final stage of the iteration the approximate covariance matrix for the estimate of β, $\hat{\beta}$, is given by $(\mathbf{XWX}^T)^{-1}$, where $\mathbf{W}$ is calculated from $\hat{\beta}$. Confidence intervals can be derived by assuming approximate Normality of the estimates.

3.1.4 Selection of Covariates in the Model

This section concerns the selection of covariates in the regression and assumes that the error structure and link function are correctly specified. Two extreme models can be identified. The first contains only one parameter and $\mathbf{X}$ is the 1xI vector of ones. This is called the minimal model and restricts the means μ_i, $1\le i\le I$, to a common value. At the other extreme, if I covariates are included so that $\mathbf{X}$ is a non-singular IxI matrix, the model fits the data perfectly in that $\mu=\mathbf{y}$. This is called the complete model, having the same number of parameters as observations, or distinct covariate combinations. The minimal model is so constrained that no systematic variation is allowed, and any trends in the data are ignored, whereas in the complete model all the variation is accounted for by μ, and there is no stochastic element. In practice the model should include a limited number of covariates which are associated with substantial trends in the data. The remaining variation in the data is consigned to random error.

The fit of a model can be assessed by calculating the deviance, $D(\mathbf{y};\tilde{\mu})$. Assuming that $a_i(\phi)=\phi/w_i$, the deviance is based on the difference between the maximium log likelihood achievable (by the complete model with $\mu=\mathbf{y}$) and the log likelihood of the current model based on p+1 parameters with estimated model mean $\tilde{\mu}$ multiplied by the scale parameter ϕ. That is

$$D(\mathbf{y};\tilde{\mu}) = 2\sum_{i=1}^{I} w_i\{y_i(\theta_i^*-\tilde{\theta}_i)+b(\tilde{\theta}_i)-b(\theta_i^*)\},$$

where $\tilde{\theta}_i$ and θ_i^* are the maximum likelihood estimates of θ under the current and complete models respectively. The formula does not depend on ϕ. The scaled deviance $D(\mathbf{y};\tilde{\mu})/\phi$ is the log likelihood ratio between the current and complete models. In the applications presented here the low frequency of most covariate combinations invalidates comparison to a χ^2 percentile.

To test the significance of a set of covariates, the log likelihood ratio between model 1, with estimated mean $\tilde{\mu}^{(1)}$, and model 2, with estimated mean $\tilde{\mu}^{(2)}$, in which the given set of covariates is not included but other aspects of the model remain the same, is calculated. The log likelihood ratio is equal to the difference of the scaled deviances,

$$D(\mathbf{y};\tilde{\mu}^{(1)})/\phi - D(\mathbf{y};\tilde{\mu}^{(2)})/\phi,$$

and can be compared to a $\chi^2_{p_1-p_2}$ distribution, where p_1+1 and p_2+1 are the numbers of covariates in the first and second model respectively.

The deviance for the Binomial distribution is given by

$$D(\mathbf{y};\tilde{\mu}) = 2\sum_{i=1}^{I} N_i\{y_i \ln(y_i/\tilde{\mu}_i)+(1-y_i)\ln[(1-y_i)/(1-\tilde{\mu}_i)]\}.$$

The value of ϕ in the Binomial distribution is 1 and the deviance and scaled deviance are equivalent.

Some covariates may not have a simple additive effect on the linear component of the model. Possible interactions between covariates are considered by including separate parameters corresponding to each combination of the interacting covariates. Details of the parameterisation of models containing interactions are given in section 3.4.2. The significance of the interaction is tested by comparing a model with additive parameters with a model where the parameters of the interaction are included. A model with interactions contains extra parameters, and by including sufficient higher order interactions the complete model is achieved.

3.1.5 Testing the Link Function in Generalised Linear Models

It may be preferable to change other aspects of the model rather than lose the simplicity of a model containing only additive covariate parameters. The two other properties of the model that can be modified are the error structure and the link function. This section examines the possibility of changing the latter.

Pregibon (1980) has suggested a test of the adequacy of the link function, assuming that the error structure and the linear component of the model are appropriately specified. The hypothesized link function, $g^{o}(\mu)$, is embedded in a parametric family of functions which is assumed to contain the true link function $g^{*}(\eta)$. In a two parameter family, for example, g^{o} is written

$$g^{o}(\mu) = g(\mu;\alpha^{o},\delta^{o})$$

and

$$g^{*}(\mu) = g(\mu;\alpha^{*},\delta^{*})$$

Using a first-order Taylor expansion about the hypothesized link function, we obtain

$$g^{*}(\mu) \doteq g^{o}(\mu)+(\alpha^{*}-\alpha^{o})D_{\alpha}(g^{o})+(\delta^{*}-\delta^{o})D_{\delta}(g^{o}),$$

where

$$D_{\alpha}(g^{o}) = \{\partial/\partial\alpha \; g(\mu;\alpha,\delta)\}_{\alpha=\alpha^{o}\,,\;\delta=\delta^{o}}$$

and similarly for $D_{\delta}(g^{o})$. The correct link function is approximated by

$$g^{o}(\mu) = \mathbf{x}^{T}\beta+\mathbf{z}\gamma,$$

where $\mathbf{z}=\{D_{\alpha}(g^{o}),D_{\delta}(g^{o})\}$ and $\gamma^{T}=-\{\alpha^{*}-\alpha^{o},\delta^{*}-\delta^{o}\}$. The problem of testing the adequacy of the link function is now stated in terms of the hypothesized link function with additional factors in the linear component of the model describing local differences in the link function.

Consider an initial fit of the model with link function g^{o} which yields estimate $\tilde{\beta}$, and fitted values $g^{o}(\tilde{\mu}_i)=\mathbf{x}_i^{T}\tilde{\beta}$, $1\leq i\leq I$. It is then possible to calculate $\tilde{\mathbf{z}}=\mathbf{z}_{\beta=\tilde{\beta}}$, and the model can be refitted using $\tilde{\mathbf{z}}$ as an additional explanatory variable. A significant reduction in the resulting deviance indicates a departure from the hypothesized link function. Pregibon suggests the following family of link functions for use with Binomial errors

$$g(\mu;\alpha,\delta) = \frac{\mu^{\alpha-\delta}-1}{\alpha-\delta}+\frac{(1-\mu)^{\alpha+\delta}-1}{\alpha+\delta},$$

the logit link function is given by

$$g^{o}(\mu) = \lim_{\alpha,\delta\to 0} g(\mu;\alpha,\delta).$$

The estimated derivatives $D_{\alpha}\{g^{o}(\mu_i)\}$ and $D_{\alpha}\{g^{o}(\mu_i)\}$ are

$$\tilde{z}_{i1} =1/2\{\ln^2(\tilde{\mu}_i)-\ln^2(1-\tilde{\mu}_i)\},$$

and

$$\tilde{z}_{i2} = -1/2\{(\ln^2(\tilde{\mu}_i)+\ln^2(1-\tilde{\mu}_i)\}.$$

and can be calculated using the estimated means, $\tilde{\mu}$, from logistic regression. This procedure is not equivalent to full maximum likelihood estimation of α,δ and β, but Pregibon suggests that in most applications the method provides a reasonable approximation.

3.2 Logistic Regression Model for Binary Data

3.2.1 The Model

An important special case of the generalised linear model is the logistic regression model (Cox, 1970), which relates the probability of a binary response for each individual, Z, taking values 1 and 0 to covariate vector, $\mathbf{x}$. The form of the dependence of the response variable Z on $\mathbf{x}$ in the logistic regression model is

$$E(Z;\mathbf{x}) = P(Z=1;\mathbf{x}) = \frac{\exp(\mathbf{x}^T\beta)}{1+\exp(\mathbf{x}^T\beta)},$$

or equivalently the following relationship holds between the log odds and the linear predictor, $\mathbf{x}^T\beta$,

$$\text{logit}(P(Z=1;\mathbf{x})) = \log\left(\frac{P(Z=1;\mathbf{x})}{1-P(Z=1;\mathbf{x})} \right) = \mathbf{x}^T\beta .$$

In examples where the components of $\mathbf{x}$ are dummy variables representing categorical variables and if $P(Z=1)$ is small, then $\exp(\beta_j)$, $1 \le j \le p$, is approximately equal to the relative risk of $Z=1$ for $x_j=1$ compared to $x_j=0$, all other components of $\mathbf{x}$ being unchanged.

In a sample that comprises I distinct covariate combinations and assuming that the response data for individual cases are independent, the proportion of positive values, Y_i, for each covariate combination is taken as the response variable in the generalised linear model and N_iY_i has a Binomial distribution given by $\text{Bin}(N_i, P(Z=1;\mathbf{x}_i))$, $1 \le i \le I$.

3.2.2 Maximum Likelihood under Three Sampling Schemes

The logistic regression model can be fitted under three commonly occurring sampling schemes, prospective studies, joint sampling and retrospective, case-control studies. Suppose, in a prospective study, that N_i individuals are sampled for covariate combination i, and of these n_{i1} are observed to be positive and n_{i0} to be negative. The likelihood, l, given by

$$l = \prod_{i=1}^{I} \{P(Z=1;\mathbf{x}_i)\}^{n_{i1}} \{P(Z=0;\mathbf{x}_i)\}^{n_{i0}}$$

and the formula for logistic regression of Z on $\mathbf{x}$ substituted for

$P(Z;\mathbf{x})$.

If, as is the case in the analyses presented here, data are sampled from the joint distribution of $(Z,\mathbf{x})$, the likelihood is given by

$$l = \prod_{i=1}^{I} \{P(Z=1,\mathbf{x}_i\}^{n_{i1}} \{P(Z=0,\mathbf{x}_i)\}^{n_{i0}} .$$

Conditioning on x gives

$$l = \prod_{i=1}^{I} \{P(Z=1|\mathbf{x}_i)\}^{n_{i1}} \{P(Z=0|\mathbf{x}_i)\}^{n_{i0}} \prod_{i=1}^{I} \{P(\mathbf{x}_i)\}^{N_i} .$$

The logistic model makes no assumpions about the marginal distribution of $\mathbf{x}$, and the first part of this formula contains all the information about the parameters, β. Maximum likelihood estimates are obtained by maximising over the first part, and the term $\prod_{i=1}^{I} \{P(\mathbf{x}_i)\}^{N_i}$ cancels out of likelihood ratios. The model can be fitted as though the data were prospectively sampled.

In retrospective studies fixed numbers of cases with $Z=1$, and controls with $Z=0$, are sampled. Prentice & Pyke (1979) show that under this sampling scheme also, the logistic model for Z given $\mathbf{x}$ can be fitted as though the data were prospectively sampled.

3.2.3 Assessment of Fit of the Logistic Regression Model

The adequacy of the fit of a logistic regression model can be assessed in various ways. Firstly, the estimated model means, $\tilde{\mu}$, can be compared to the observed values $\mathbf{y}$. When the data are proportions and the majority of the covariate combinations contain only one individual, a simple plot of y against $\tilde{\mu}$ is not very useful, because most points will be at one extreme or other of the y scale. Grouping the $\tilde{\mu}$ scale into categories that are well represented, the observed proportion and associated confidence intervals can be plotted within each category. These points should not differ substantially from the line $\mathbf{y}=\tilde{\mu}$ if the model is adequate.

A second method of assessing the fit of a logistic regression is to examine residuals. The simplest definition of residuals for Binomial data is the Pearson residual

$$r_i = N_i\{y_i-\tilde{\mu}_i\}/[N_i\tilde{\mu}_i(1-\tilde{\mu}_i)]^{1/2},$$

which is the difference between the observed and expected counts, divided by the estimated standard error of N_iY_i. If N_i is small or $\tilde{\mu}_i$ near 0 or 1, the adjusted deviance residual, d_i, may be preferable. This is given by

$$d_i = \pm\{2N_iy_i\ln(y_i/\tilde{\mu}_i)+2N_i(1-y_i)\ln[(1-y_i)/(1-\tilde{\mu}_i)]\}^{1/2} \\ +(2\tilde{\mu}_i-1)/\{6[N_i\tilde{\mu}_i(1-\tilde{\mu}_i)]^{1/2}\},$$

where the sign is that of $y_i-\tilde{\mu}_i$. These residuals are related to components of the likelihood and are more nearly Normally distributed than the Pearson residuals. However the approximation may be poor in extreme cases where many of the N_i are 1 (McCullagh & Nelder, 1983).

3.3 Regression Models For Ordered Polytomous Data

3.3.1 Generalised Linear Regression Models for Polytomous Data

A class of linear models for polytomous response data was presented by McCullagh (1980). If an underlying, possibly unobservable, continuous variable is grouped into contiguous categories the model can be interpreted in terms of the underlying variable. The model can also be used in the case of ordered qualitative data, when the results are interpreted in terms of the categories available. In this section the generalised model is described. Section 3.3.2 describes the special cases of logistic and proportional hazards models and suggests when each might be appropriate.

Let Z be a response variable taking qualitative values 1,...,k, with probabilities $\pi_1(\mathbf{x}),\ldots,\pi_k(\mathbf{x})$ for an individual with covariates $\mathbf{x}$. Further define for $1\leq j<k$

$$P(Z\leq j) = \gamma_j(\mathbf{x}) = \pi_1(\mathbf{x})+\ldots+\pi_j(\mathbf{x})$$

and $\gamma_k(\mathbf{x})=1$. The quantities $\{\gamma_j(\mathbf{x})\}$ are related to the linear combination of the covariates, $\mathbf{x}^T\beta$, by a link function

$$g\{\gamma_j(\mathbf{x})\} = \theta_j+\mathbf{x}^T\beta, \qquad \text{for } 1\leq j<k.$$

The function g(.) is monotonic, increasing and maps (0,1) onto $(-\infty,\infty)$, and the same considerations apply as for the link function of the binary models of section 3.1.2. The covariate vector, $\mathbf{x}$, does not include a first component, x_0, set identically equal to one. In earlier sections the parameter β_0 corresponded to the mean of a scalar response in the reference level, $\mathbf{x}=(1,0,0,\ldots,0)$. In the model for polytomous data the parameters $\{\theta_j\}$ replace β_0 and correspond to the reference level of $\{g[\gamma_j(\mathbf{x})]\}$, when $\mathbf{x}=\mathbf{0}$. The model describes a strict stochastic ordering, in that if two cases with covariates $\mathbf{x}_1$ and $\mathbf{x}_2$ are compared, then

$$g\{\gamma_j(\mathbf{x}_1)\}-g\{\gamma_j(\mathbf{x}_2)\} = (\mathbf{x}_1-\mathbf{x}_2)^T\beta = \Delta,$$

a constant, for all j. Since the link is a monotonic function it follows that either

$$\gamma_j(\mathbf{x}_1) < \gamma_j(\mathbf{x}_2) \qquad \text{for all } j,$$

or

$$\gamma_j(\mathbf{x}_1) > \gamma_j(\mathbf{x}_2) \qquad \text{for all } j.$$

A sample of N independent cases which take values according to the above model form a multinomial observation $(n_1,n_2,\ldots,n_k)$, $\sum_{j=1}^{k} n_j=N$, and have likelihood

$$\propto \pi_1^{n_1}\ \pi_2^{n_2}\ \ldots\ \pi_k^{n_k}.$$

Defining cumulative quantities $R_1=n_1$, $R_2=n_1+n_2$, ..., $R_k=N$, and $Y_1=R_1/N$, $Y_2=R_2/N$, ..., $Y_k=R_k/N=1$, we may write the likelihood as a product of k-1 terms involving the cumulative parameters γ_j

$$\prod_{j=1}^{k-1}\left\{\left(\frac{\gamma_j}{\gamma_{j+1}}\right)^{R_j}\left(\frac{\gamma_{j+1}-\gamma_j}{\gamma_{j+1}}\right)^{R_{j+1}-R_j}\right\} .$$

The contribution from each j corresponds to the probability that the first j cells divide in the ratio $R_{j-1}:R_j-R_{j-1}$. If we define

$$\phi_j = \ln\{\gamma_j/(\gamma_{j+1}-\gamma_j)\}$$

and

$$h(\phi_j) = \ln\{\gamma_{j+1}/(\gamma_{j+1}-\gamma_j)\}, \qquad \text{for } 1\leq j<k,$$

the log likelihood from the observation can be written

$$L = N\sum_{j=1}^{k-1}\{Y_j\phi_j-Y_{j+1}h(\phi_j)\}.$$

The log likelihood from I multinomial samples corresponding to the distinct covariate combinations $\{\mathbf{x}_i\}$ of sizes $\{N_i\}$ is a summation of contributions $\{L_i\}$ from the above formula with quantities γ_j, ϕ_j and $h(\phi_j)$ replaced by functions of $\mathbf{x}_i$, $\gamma_j(\mathbf{x}_i)$, ϕ_{ij} and $h(\phi_{ij})$, and Y_j replaced by Y_{ij}, $1\leq i<I$.

These models were fitted using the algorithm suggested in the appendix to McCullagh (1980). A listing of the program is given in Appendix 3.

3.3.2 Logistic and Complementary Log-Log Link Functions

In the general formulation described above any monotonic increasing function mapping (0,1) to $(-\infty,\infty)$ can be used, including the various link functions mentioned in connection with binary regression models. Choosing the logistic link function leads to the model

$$\begin{aligned} g\{\gamma_j(\mathbf{x})\} &= \ln\{\gamma_j(\mathbf{x})/[1-\gamma_j(\mathbf{x})]\} \\ &= \theta_j+\mathbf{x}^T\beta, \qquad 1\leq j<k. \end{aligned}$$

The log odds ratio of $Z\leq j$

$$\ln\{\gamma_j(\mathbf{x}_1)(1-\gamma_j(\mathbf{x}_2))/[(1-\gamma_j(\mathbf{x}_1))\gamma_j(\mathbf{x}_2)]\} = (\mathbf{x}_1-\mathbf{x}_2)^T\beta,$$

does not depend on the level, j, of the response, but only on the difference of covariate values, $\mathbf{x}_1-\mathbf{x}_2$, and the parameter β. Logistic regression for binary data is the special case of k=2.

Choosing the complementary log-log transform leads to the model

$$\begin{aligned} g(\gamma_j(\mathbf{x})\} &= \ln\{-\ln[1-\gamma_j(\mathbf{x})]\} \\ &= \theta_j+\mathbf{x}^T\beta, \qquad 1\leq j<k. \end{aligned}$$

or

$$P(Z>j;\mathbf{x}) = 1-\gamma_j(\mathbf{x}) = \exp\{-\exp(\theta_j)\}^{\exp(x^T\beta)}.$$

The relationship between P(Z>j) and $\mathbf{x}$ in this formula is equivalent to the dependence of the survivor function on $\mathbf{x}$ in a proportional hazards model, and the use of this model with grouped survival data is described in section 3.3.3.

In the discussion of McCullagh's paper, Bartholomew (1980) gave a rationale for choosing between link functions. Suppose a continuous random variable Θ, with distribution function $F(\theta)$, underlies the polytomous response. The cutpoints of the categories on the Θ scale are given by $\theta_1, \theta_2, \ldots, \theta_{k-1}$, and are generally unknown. The distribution function is allowed to vary according to parameters a and b and takes the form

$$F_\Theta(\theta) = \psi\left(\frac{m(\theta)-a}{b}\right), \qquad (-\infty<m(\theta)<\infty,\ -\infty<a<\infty,\ b>0),$$

$m(\theta)$ is a monotonic function of θ, and $\psi(.)$ is an arbitary distribution function. Defining p_j by

$$p_j = \psi\left(\frac{m(\theta_j)-a}{b}\right), \qquad 1\leq j<k,$$

then $m(\theta_j)=a+b\psi^{-1}(p_j)$. Comparing two distributions with parameters (a_1,b_1) and (a_2,b_2), and equating the cutpoints $\{m(\theta_j)\}$, we have

$$0 = a_1-a_2+b_1\psi^{-1}(p_{1j})-b_2\psi^{-1}(p_{2j}), \qquad 1\leq j<k.$$

If the scale parameters b_1 and b_2 are equal, and without loss of generality taken to be 1, then

$$a_1-a_2 = \psi^{-1}(p_{1j})-\psi^{-1}(p_{2j}), \qquad 1\leq j<k.$$

and covariates are introduced by equating $a_i=\mathbf{x}_i^T\beta$. Bartholomew suggested a family of distributions including special cases with the differences above constant over j. The family is given by

$$1-F(\theta) = \{1+(1/u)\exp[(\frac{m(\theta)-a}{b})]\}^{-u}, \qquad (u>0),$$

where the parameters a and b and m(.) have ranges as given before. Within this family

$$\psi^{-1}(p_j) = \ln(u)+\operatorname{logit}\{(1-p_j)^u\},$$

so that

$$a_1-a_2 = \operatorname{logit}\{(1-p_{2j})^{1/u}\}-\operatorname{logit}\{(1-p_{1j})^{1/u}\}.$$

If the underlying variable has distribution characterised by $u=1$, which is the requirement for a symmetric probability density function, then use of the logistic link function in the generalised model produces a constant difference between the $\{a_i\}$ at the cutpoints of the two distributions. While in the limit as $u\to\infty$ use of the complementary log-log transform will produce constant differences between parameters $\{a_i\}$.

This argument depended on the underlying continuous variable having distribution in the above family and further, that the only change in distribution between samples was in the location parameter. If $a_1=a_2$ but $b_1\neq b_2$ a parallel class of models can be obtained for $\ln(b_i)=\mathbf{x}_i^T\beta$.

3.3.3 Proportional Hazards Model

In this section some basic probability results for continuous survival data are given and the proportional hazards model described. The results are taken from Kalbfleisch & Prentice (1980) where a more complete description of the area and related topics can be found. The proportional hazards model was used in the analysis of gestational age which was available for all births, and the description is thus limited to the case where no survival times are censored.

A continuous random variable, T, measuring survival time has range $[0,\infty)$, probability density function $f(t)$, survivor function $S(t)=P(T\geq t)$ and hazard function specifying the instantaneous rate of failure at time t conditional on survival up to time t, defined by

$$\lambda(t) = \lim_{dt \to 0^+} \frac{P(t \le T < t+dt \mid T \ge t)}{dt}$$

$$= \frac{f(t)}{S(t)} = \frac{-d\ln S(t)}{dt}.$$

The distribution of T is completely specified by $\lambda(t)$, and on integrating the last result with the boundary condition of $S(0)=1$, is given by

$$S(t) = \exp\{-\int_0^t \lambda(u)du\},$$

from which it can be seen that $\lambda(t)$ is a non-negative function with $\int_0^\infty \lambda(u)du=\infty$. The probability density function of T expressed as a function of $\lambda(t)$ is

$$f(t) = \lambda(t)\exp\{-\int_0^t \lambda(u)du\}.$$

The proportional hazards model (Cox,1972) is based on the assumption that the covariates have a constant multiplicative effect on the hazard function over time. Under this assumption the hazard function for the covariate vector, $\mathbf{x}$, is given by

$$\lambda(t;\mathbf{x}) = \lambda_o(t)\exp(\mathbf{x}^T\beta)$$

where $\lambda_o(t)$, the reference hazard function for $\mathbf{x}=\mathbf{0}$, is a function of t, and the multiplicative factor $\exp(\mathbf{x}^T\beta)$ depends only on $\mathbf{x}$ and the parameter vector β. The density function for covariates $\mathbf{x}$ is then

$$f(t;\mathbf{x}) = \lambda_o(t)e^{\mathbf{x}^T\beta}\exp\{-e^{\mathbf{x}^T\beta}\int_0^t \lambda_o(u)du\},$$

and the survivor function is

$$S(t;\mathbf{x}) = [S_o(t)]^{\exp(\mathbf{x}^T\beta)}$$

where $S_o(t)$ is the survivor function for $\mathbf{x}=\mathbf{0}$.

Prentice & Gloeckler (1978) adapted the proportional hazards models for use with survival data grouped into intervals $[c_1,c_2),\ldots,[c_j,c_{j+1}),\ldots,[c_k,c_{k+1})$, with $c_1=0$ and $c_{k+1}=\infty$. The probability of observing a failure in the j^{th} interval is

$$p(c_j \le T < c_{j+1}) = [1-\alpha_j^{\exp(\mathbf{x}^T\beta)}] \prod_{s=1}^{j-1} \alpha_s^{\exp(\mathbf{x}^T\beta)},$$

where

$$\alpha_j = \exp\{-\int_{c_j}^{c_{j+1}} \lambda_o(u)du\}$$

is the probability of surviving the j^{th} interval conditional on survival to c_j when $\mathbf{x}=\mathbf{0}$. The probability of surviving to the beginning of the j^{th} interval is

$$S(c_j;\mathbf{x}) = \prod_{s=1}^{j-1} \alpha_s^{\exp(\mathbf{x}^T\beta)}.$$

Prentice & Gloeckler (1978) suggest a reparameterisation to $\theta_j^*=\ln\{-\ln(\alpha_j)\}$ which has range $(-\infty,\infty)$, and the corresponding contribution to the log likelihood is

$$L = \ln[1-\exp\{-\exp(\theta_j^*+\mathbf{x}^T\beta)\}] - \sum_{s=1}^{j-1} \exp(\theta_s^*+\mathbf{x}^T\beta).$$

The probability of surviving to the $j+1^{st}$ interval is

$$\begin{aligned} S(c_{j+1};\mathbf{x}) &= \prod_{s=1}^{j-1} \exp\{-\exp[\theta_s^*+\mathbf{x}^T\beta]\}, \\ &= \exp\{-\sum_{s=1}^{j} \exp(\theta_s^*)\}^{\exp(\mathbf{x}^T\beta)} \end{aligned}$$

and comparison with the form of $P(Z>j;\mathbf{x})$ when employing the complementary log-log link function within the generalised linear models of section 3.3.2 reveals that the two approaches are equivalent. The survivor functions when $\mathbf{x}=\mathbf{0}$ are, in the two parameterisations are given by

$$\exp\{-\exp(\theta_j)\} = \prod_{s=1}^{j} \exp\{-\exp(\theta^*_s)\}.$$

Parameters $\{\theta_j\}$ measure the survivor function, while $\{\theta^*_j\}$ measure the probability of survival within each time interval.

Maximum likelihood estimates were calculated using Newton-Raphson iteration (Appendix 4).

3.4 Parameterisation

3.4.1 Notation

The following sections describe the parameterisation of the linear component in the models described in sections 3.1, 3.2 and 3.3. It is convenient to adopt an alternative notation for covariates A and B with r_A and r_B levels respectively. A model for a scalar response variable including only the main effects of A contains the following r_A parameters

$$m, a_2, \ldots, a_{r_A}.$$

(If A has only two levels $a_{r_A} = a_2$.) The parameters $\{a_j\}$ can be equated to components of the parameter vector β of section 3.1 thus

$$\begin{aligned} m &= \beta_0, \\ a_2 &= \beta_1, \\ &\vdots \\ a_{r_A} &= \beta_{r_A-1}. \end{aligned}$$

The parameter m corresponds to the value of the linear component in the reference level of A. The linear component

$$\eta = m + a_j, \qquad 2 \leq j < r_A,$$

for an individual in the j^{th} level of A. The parameters $\{a_j\}$ thus contrast the linear component between the j^{th} and the reference level of A. Similarly a model containing the main effects of both A and B contains parameters

$$m \backslash a_2, \dots, a_{r_A} \backslash b_2, \dots, b_{r_B} \backslash$$

and in terms of parameters β, $\{b_j\}$ are

$$\begin{aligned} b_2 &= \beta_{r_A} \\ &\vdots \\ b_{r_A} &= \beta_{r_A + r_B - 1} \end{aligned}$$

3.4.2 Covariate Interactions

Interactions are introduced by contrasting the levels of A within each level of B, which is treated as a stratifying variable. This approach gives the two covariates a different status in the model and, in practice, it is necessary to decide which should be the stratification variable. The choice does not affect the values of the linear component or the likelihood, but the selection of one covariate of primary interest may be more convenient for interpretation. When an interaction between A and B is included, with A stratified by B, the model contains the following parameters

$$m \backslash b_2, \dots, b_{r_B} \backslash (ab)_{21}, \dots, (ab)_{r_A 1} \backslash (ab)_{22}, \dots, (ab)_{r_A 2} \backslash \\ \dots \backslash (ab)_{2r_B}, \dots, (ab)_{r_A r_B} \backslash \quad .$$

Parameters $\{(ab)_{i1}\}$ contrast the levels of A within the reference level of B, while the $\{(ab)_{ij}\}$ parameters contrast the levels of A within the j^{th}, $1 \leq j \leq r_B$, level of B. Parameters $\{b_j\}$ fit the baseline levels of B. Altogether the model contains $1+(r_B-1)+r_B(r_A-1)=r_Ar_B$ parameters.

A more complicated model might, for example, contain the parameters

$$m,\{b_j\},\{c_j\},\{d_j\},\{(ad)_{i1}\},\{(ad)_{i2}\},\\ \ldots,\{(ad)_{ir_D}\},$$

representing the contrasts between the levels of B, C and D, and the contrasts between the levels of A within each level of D. This model contains

$$1+(r_B-1)+(r_C-1)+(r_D-1)+r_D(r_A-1)\\ = r_Ar_D+(r_B-1)+(r_C-1)$$

parameters. The parameter vector β is equivalent to the parameters taken sequentially in the above list. Corresponding values of the covariate vector are calculated by setting $x_0=1$, and $x_j=1$, $1\leq j\leq r_B-1$, for cases in the $j+1^{th}$ level of B and $x_j=0$ otherwise. Values for the x_j corresponding to the main effects of C and D are similarly calculated. The x_j corresponding to parameters $\{(ad)_{ij}\}$ equal 1 in the appropriate levels of both A and D, and otherwise are 0.

Interactions between three or more covariates can be included. An interaction between A, B and C is treated as the contrasts for A within all combinations of the levels of stratifying covariates, B and C. Parameters $\{b_j\}$, $\{c_j\}$ and $\{(bc)_{ij}\}$ should also to be included in the model.

3.4.3 Covariate Interactions with a Polytomous Response

In the description of the model for polytomous response data the linear component was constant at all levels of the response. This constraint leads to the property of stochastic ordering in section 3.3.1. The restriction has important consequences for interpretation and should be justified in practice. In this section the effects of covariates are allowed to vary freely over the levels of the response. Provision for such terms was made in the program for the generalised polytomous regression model (Appendix 3).

Suppose that covariate A, with r_A levels, does not have a constant effect at all k levels of the response variable. The association can be modelled by including the following parameters.

$$m_1,\ldots,m_{k-1}\backslash a_{21},\ldots,a_{r_A 1}\backslash\ldots\backslash a_{2j},\ldots,a_{r_A j}\backslash$$
$$\ldots\backslash a_{2,(k-1)},\ldots,a_{r_A,(k-1)}\backslash.$$

The parameters $\{m_j\}$ are equivalent to parameters $\{\theta_j\}$ of section 3.2.1 and give the reference level of the linear component at the j^{th} level of the response. The parameters $\{a_{ij}\}$ contrast the levels of A in the j^{th} category of the response. The model thus contains $k-1+(k-1)(r_A-1)=r_A(k-1)$ parameters and is equivalent to a separate set of $\{m_j\}$ for each level of A. The covariate vector, $\mathbf{x}$, changes as the response variable varies. Component x_s, $1\leq s\leq r_A(k-1)$, corresponding to a parameter in the above list, is set to 1 if the individual is in the appropriate level of A and category of the response, otherwise $x_s=0$.

When several covariates are included in a model the parameters corresponding to one or more may be allowed to vary while the others are constrained to be constant over the categories of the response. If parameters for all covariates vary over the response the model is equivalent to a series of separate binary analyses at each response category. It is possible to allow interactions between two covariates, A and B, to vary over the response. A separate set of parameters contrasting the levels of A within B, would be included for each level of the response. In these cases it may be appropriate to allow only the $\{(ab)_{ls}\}$ parameters to vary, or, possibly, the $\{b_l\}$ parameters should be allowed to vary also.

When the logistic link function is used the log odds vary over categories of the response. If the complementary log-log transform is used, the proportional hazards assumption is modified, and the covariates do not have a constant multiplicative effect on the hazard.

3.4.4 Linear Covariate Dependence in the Hazards Model

The proportional hazards model was used in the analysis of gestational age where response categories represented equal lengths of time (weeks of gestation). It is reasonable to suppose that the hazards, if not constant, might increase or decrease linearly with time. The model for grouped survival data described in section 3.3.3 permits a linear dependence on the time index, but the hazards must remain proportional within time intervals (Prentice & Gloeckler, 1978). If covariate A has a multiplicative effect on the hazard which varies linearly with time the following parameters are included in the model

$$m_1, \ldots, m_{k-1} \backslash \alpha_2, \ldots, \alpha_{r_A} \backslash \beta_2, \ldots, \beta_{r_A} \backslash .$$

Parameters $\{m_j\}$ are the $\{\theta^*_j\}$ of section 3.3.3 and $\{\alpha_j\}$ and $\{\beta_j\}$ correspond to intercept and slope parameters respectively, describing the contrasts between each level and the reference of A increasing or decreasing over the time index. The hazard function in the l^{th} category of A during the j^{th} time category is thus given by

$$\lambda(t;l) = \lambda_o(t)\exp(\alpha_l+j\beta_l), \qquad c_j \leq t < c_{j+1}.$$

If the effects of several covariates are being modelled simultaneously it would be possible to allow some or all to be linearly dependent on time.

Introducing linear time dependence into the model comes between the proportionality assumption and free dependence on time as described in 3.4.2. If survival times are grouped into more than three categories, it is possible to introduce a further parameter allowing for quadratic dependence on time. The hazard in the j^{th} time category and the l^{th} covariate category would be given by

$$\lambda(t;l) = \lambda_o(t)\exp(\alpha_l+j\beta_l+j^2\gamma_l), \qquad c_j \leq t \leq c_{j+1}.$$

Higher order terms could also be considered.

The program in Appendix 4 allows for the inclusion of linear time dependent effects in the survival model.

3.5 Alternative Regression Techniques for Categorical Data

3.5.1 Introduction

In these final sections some methods connected with the fitting of regression models to categorical data that were not used in this study are described. First, several graphical methods for assessing the adequacy of binary logistic regression models are the subject of section 3.5.2. An alternative logistic model for polytomous response data is described in section 3.5.3. Finally, in section 3.5.4, the use of standard log-linear packages to fit grouped survival models and the generalised polytomous model is outlined.

3.5.2 Assessment of Fit of the Logistic Regression Model

Landwehr, Pregibon & Shoemaker (1984) presented three graphical methods for assessing the fit of logistic regression models. The first, a plot of local mean deviances, is based on an analogy with linear regression with replicated observations, where the residual sum of squares is partitioned into pure error within replicates and a term due to lack of fit. In order to achieve this when there are few if any replicates, groups of observations are defined using a clustering algorithm on the space of covariates. The groups need to be large enough to provide sufficient degrees of freedom to estimate the local mean rate, but if they are too large some of the local deviance may reflect heterogeneity within groups. Within each of the resultant K groups a local estimate $\hat{\mu}^L$ of the mean from the model $\text{logit}(\mu_i^L)=\mathbf{x}_i^T\beta+\gamma_k$ is fitted, where the i^{th} observation is in the k^{th} group. As Fienberg & Gong (1984) point out in the discussion to the above paper a likelihood ratio test comparing this model with $\text{logit}(\mu_i)=\mathbf{x}_i^T\beta$ with K-1 degrees of freedom would be appropriate, but in order to obtain a plot the local deviance contributions of each observation $d(y_i;\hat{\mu}_i^L)$ and their sums within the k^{th} group D_k^L are calculated. The groups are then ordered according to a measure of within group homogeneity, which could be the metric of the original clustering algorithm. Running estimates of pure error, within groups,

$$\bar{D}^L(t) = \sum_{k=1}^{t} D_k^L \Big/ \sum_{k=1}^{t}(N_k-1)$$

are calculated over the t tightest groups, where N_k is the number of observations in the k^{th} group. When $\bar{D}^L(t)$ is plotted against its degrees of freedom, $\sum_{k=1}^{t}(N_k-1)$, $1 \leq k \leq K$, for large t, $\bar{D}^L$ should tend to $D(\mathbf{y};\hat{\mu})/(N-m)$, where $\hat{\mu}$ is the fitted mean from the model $\text{logit}(\mu)=\mathbf{x}^T\beta$ with m degrees of freedom. This technique was devised for applications involving continuous covariates. However, data from several categorical covariates may involve many unique covariate combinations and clustering into groups of near replicates (possibly involving a change in only one dichotomous covariates, or to an adjacent level of an ordered polytomous covariate) may prove useful. The disadvantage of the approach is the expense of the clustering algorithm, particularly when there are many observations.

The second graphical method suggested by Landwehr et al is an empirial probability plot of the residual quantities $r_i=y_i-\hat{\mu}_i$ from the model $\text{logit}(\mu)=\mathbf{x}^T\beta$. M data vectors $\mathbf{y}^*$ are simulated from the distribution $y_i^*=\text{Bin}(1,\hat{\mu}_i)$, $1 \leq i \leq I$, and fitted means $\hat{\mu}^*$ are calculated for each simulation. Typical values and upper and lower quantiles (corresponding to a chosen confidence level) are obtained from the M sets of ordered residuals $r^*_{m(i)}$, where (i) refers to the ordered $r^*_{mi}=y^*_{mi}-\hat{\mu}_{mi}$ within the m^{th} simulation, $1 \leq m \leq M$. The observed ordered residuals are plotted against the typical values and the upper and lower confidence intervals from the simulations. When the model fits adequately the points should lie on the diagonal line from (0,0) to (1,1), and lack of fit is detected by a significant departure from this line beyond the confidence intervals.

The final graphical method, suggested by Landwehr et al, partial residual plots, examines the functional form of the dependence of μ on a continuous covariate. Suppose that **y** is generated by the logistic model

$$\text{logit}\{\mu(z)\} = \mathbf{x}^T\beta + g(z),$$

where g is an unknown function of a continuous covariate z, and a linear logistic model

$$\text{logit}\{\mu^{(1)}(z)\} = \mathbf{x}^T\beta + \gamma z$$

is fitted. The quantity

$$G(Z) = \frac{y-\mu^{(1)}}{\mu^{(1)}(1-\mu^{(1)})} + \gamma z$$

should have expectation approximately equal to g(z), if $\mu/\mu^{(1)} \doteq 1$ and $(1-\mu^{(1)})/(1-\mu) \doteq 1$. Substituting estimates for $\mu^{(1)}$ and γ in the formula and plotting the estimates of G(z) against z gives two sets of points, at y=0 and y=1, and these require smoothing before the form of E(G(z)) is revealed. The final values for E(G(z)) will be biased if the true logit is not well approximated by the linear model, and in this situation it is most important to identify g(z). Hastie (1984) in the discussion to Landwehr et al suggested that the functional form of the dependence on p covariates could be investigated in the model

$$\text{logit}(\mu_i) = \sum_{s=1}^{p} \phi_s(x_{is}).$$

The model can be fitted by estimating local linear approximations to $\phi_s(x_{is})$, $\hat{\phi}_s(x_{is}) = \hat{a}_{is} + \hat{b}_{is}x_{is}$, in the neighbourhood defined by the closest k observations to x_{is}. Repeating the procedure over i for each covariate s in turn produces a smooth estimate of the function ϕ_s. This method of determining the functional form of the dependence of μ on continuous covariates is similar to the smoothed estimate

$$\hat{\mu}(z) = \frac{\sum_{i=1}^{N} y_i \psi\{h^{-1}(x-x_i)\}}{\sum_{i=1}^{N} \psi\{h^{-1}(x-x_i)\}}$$

for a single covariate suggested by Copas (1983). The parameter h controls the amount of smoothing and the kernel $\psi(\ )$ is commonly taken to be the standard Normal density function. These methods have the advantage over grouping continuous covariates, in that the functional form within categories can be examined, but individual cases need to be treated as separate observations, and when large amounts of data are involved the models may be prohibitively expensive to fit

Several other approaches to assessing the adequacy of logistic regression models have been investigated. Pregibon (1981) describes coefficients that can be used to identify the effect of each observation on various aspects of the fit of the model. The simplest of these are the contributions of each observation to the deviance and Pearson goodness of fit statistics. Other statistics identify extreme points in the covariate space and points that are influential in determining β. These and the methods of Landwehr et al and Copas were the subject of a comparative study reported by Kay & Little (1986) where a logistic regression involving several covariates was scrutinised by examining various diagnostics.

3.5.3 An Alternative Regression Model for Polytomous Response Data

Anderson (1984) suggested an alternative model for polytomous response data in which it is possible to test for ordering of the response categories and their distinguishability with respect to covariates $\mathbf{x}$. Starting with the logistic model for a qualitative response variable, Z, with k categories, the probability P(Z=j) is given by

$$P(Z=j;\mathbf{x}) = \exp(\beta_{0j}+\mathbf{x}^T\beta_j)/\sum_{s=1}^{k}\exp(\beta_{0s}+\mathbf{x}^T\beta_s), \quad 1\leq j\leq k,$$

where $\beta_{0k}=0$ and $\beta_k=0$. The probabilities of falling in each category, rather than the cumulative probabilities are thus the basis of this approach. This model is equivalent to the following specification of the density function, f, of $\mathbf{x}$ within a level, j, of the response

$$f_j(\mathbf{x})/f_k(\mathbf{x}) = \exp(\beta^*_{0j}+\mathbf{x}^T\beta_j), \quad 1\leq j\leq k,$$

because if the k distributions $\{f_j(\mathbf{x})\}$ are mixed in proportions $\pi_1,\ldots,\pi_k$, and

$$P(Z=j;\mathbf{x}) = \pi_j f_j(\mathbf{x})/\{\sum_{s=1}^{k}\pi_s f_s(\mathbf{x})\},$$

the first formulation results from setting $\beta_{0j}=\beta^*_{0j}-\beta^*_{0k}+\log(\pi_j/\pi_k)$.

The parameters $\{\beta_j\}$ can be restricted to be parallel, that is

$$\beta_j = -\phi_j\beta, \quad 1\leq j\leq k,$$

with the restrictions $\phi_1=1$ and $\phi_k=0$ to ensure identifiability. The model is given by

$$P(Z=j;\mathbf{x}) = \exp(\beta_{0j}-\phi_j\mathbf{x}^T\beta)/\sum_{s=1}^{k}\exp(\beta_{0s}-\phi_s\mathbf{x}^T\beta),$$

with a corresponding form for $\{f_j(\mathbf{x})\}$. Various properties were demonstrated by Anderson for the case where, in addition, the ϕ_i are ordered as follows

$$1 = \phi_1 > \phi_2 > \dots > \phi_k = 0.$$

First, if the distribution of $\mathbf{x}$, conditional on Z=j is assumed to be multivariate Normal with the same variance-covariance matrix for all j, then the distributions within each j are displaced along the line of $\mathbf{x}^T\beta$ in the $\mathbf{x}$-space, and are thus ordered in one dimension. The distribution of Z given $\mathbf{x}$ was also shown to be stochastically ordered in the sense described in section 3.3.1, and hence the model describes an ordered response with respect to $\mathbf{x}$.

The model can be generalised by allowing for greater dimensionality of the relationship between Z and $\mathbf{x}$, for example in a two dimensional relationship, β_j is given by

$$\beta_j = -\phi_j\beta - \psi_j\gamma, \qquad 1\leq j\leq k,$$

so that Z depends on two functions, $\mathbf{x}^T\beta$ and $\mathbf{x}^T\gamma$, of $\mathbf{x}$. Further restrictions have to be placed on the $\{\psi_j\}$ to ensure identifiability. The choice of dimensionality can be made on the basis of the likelihood of models of increasing dimensionality, but comparison to χ^2 percentiles may not be valid under these circumstances.

Two categories of the response variable, s and t, are indistinguishable if $\beta_s=\beta_t$, and this implies that $\phi_s=\phi_t$ in the one dimensional parallel model. In the two dimensional parallel model indistinguishability implies $\phi_s=\phi_t$ and $\psi_s=\psi_t$, and similarly for higher dimensions. Alternatively, models can be fitted in which, for example, $\beta_1=\ldots=\beta_r$, $\beta_{r+1}=\ldots=\beta_k=0$, where the categories are considered in two groups, and it is not possible to distinguish between categories within either group on the basis of $\mathbf{x}$. Models in which r takes all possible values, $1\leq r\leq k$, for the break point between the two groups would be examined. The procedure could be repeated for any number up to k-1 groups of categories.

For several reasons the approach of McCullagh (1980) was preferred in the analysis of birthweight standardised for gestational age. First, the categorical response is based on an underlying continuous variable and so it would seem appropriate to use a model based its existence. Secondly, McCullagh's model is based directly on $P(Z\leq j;\mathbf{x})$ rather than $P(Z=j;\mathbf{x})$ and parameter estimates can be interpreted in terms of birthweight below the 10^{th} or 5^{th} percentiles or above the upper percentiles. It was possible to extend McCullagh's model by relaxing the ordered property for each covariate in turn, and the resultant model gave some idea of the change in shape of the distribution of the underlying continuous variable across the levels of the covariate. In Anderson's approach it would also be feasible to impose the parallel constraint of $\beta_j=-\phi_j\beta$ for the components of β corresponding to several chosen covariates, while others had an unconstrained impact over j. The ability to identify categories that are not distinguishable with respect to each covariate would provide insight into the association between the covariates and birthweight, and, as Anderson points out, it is difficult to formulate distinguishablity in terms of McCullagh's model.

3.5.4 Fitting Grouped Survival Models and Generalised Polytomous Models using Log-Linear Packages

Several authors (Aitkin & Clayton, 1980; Holford, 1980; Laird & Olivier, 1981) have shown that grouped survival models can be fitted using standard packages for log-linear models if survival times are assumed to be exponentially distributed within intervals. If a sample of N independent observations, with survival times t_i, are exponentially distributed with hazard function θ and survivor function $e^{-\theta t}$, the likelihood is

$$l(\theta) = \prod_{i=1}^{N} \theta^{w_i} e^{-\theta t_i}$$

$$= \theta^{d} e^{-\theta T},$$

where $w_i=1$ if the i^{th} individual fails at t_i and $w_i=0$ if it is right censored at t_i, the observed number of deaths, $d= \sum_{i=1}^{N} w_i$ and the exposure, $T= \sum_{i=1}^{N} t_i$. If the number of deaths is assumed to follow a Poisson distribution with mean θT the likelihood $l_p(\theta)$ is

$$l_p(\theta) = (T\theta)^{d} e^{-\theta T}/d!$$

$$= l(\theta).$$

The maximum likelihood estimate of θ under both sampling schemes is d/T, and the two models are interchangeable with respect to likelihood inference about θ.

When the exponential hazard, θ, varies across k time intervals, the time observation is represented by $(t_{1i},\ldots,t_{ji},\ldots,t_{ki})$ where t_{ji} is the length of time spent by individual i in the j^{th} time interval, and w_i becomes $(w_{1i},\ldots,w_{ji},\ldots,w_{ki})$ where $w_{ji}=1$ if the j^{th} individual fails in the j^{th} interval and 0 otherwise. If an individual, i, fails in the j_1^{th} interval then $t_{ji}=0$ for all $j>j_1$. Letting $(\theta_1,\ldots,\theta_k)$ represent the hazards in the k intervals, the likelihood is

$$l(\theta) = \prod_{i=1}^{N} \prod_{j=1}^{k} \theta_j^{w_{ji}} e^{-\theta_j t_{ji}}$$

$$= \prod_{j=1}^{k} \theta_j^{d_j} e^{-\theta_j T_j}$$

where $d_j = \sum_{i=1}^{N} w_{ji}$ is the number of failures and $T_j = \sum_{i=1}^{N} t_{ji}$ the total time spent in the j^{th} interval. If each d_j is assumed to be an independent Poisson with mean, conditional on T_j, equal to $\theta_j T_j$ then the likelihood will, as before, be proportional to that from the exponential model.

Covariates can be introduced assuming a proportional hazards model, by letting $d_{ji(1)\ldots i(p)}$ denote the number of deaths, and $T_{ji(1)\ldots i(p)}$ denote the total exposure among individuals with covariate vector $(i(1),\ldots,i(p))$ in the j^{th} time interval, and letting each $i(.)$ index the levels of one of p covariates. Representing the time subscript, j, by $i(0)$, the likelihood is

$$l(\theta) = \prod_{i(0)\ldots i(p)} \{m_{i(0)\ldots i(p)}\}^{d_{i(0)\ldots i(p)}} e^{\{-m_{i(0)\ldots i(p)}\}},$$

where

$$m_{i(0)\ldots i(p)} = T_{i(0)\ldots i(p)}\theta_{i(0)\ldots i(p)}$$

and

$$\ln[\theta_{i(0)\ldots i(p)}] = u_{0\{i(0)\}} + u_{1\{i(1)\}} + \ldots + u_{p\{i(p)\}}.$$

The terms $u_{0\{i(0)\}}$, $1 \leq i(0) \leq k-1$, represent the reference hazard in the $i(0)^{th}$ interval. This is a log-linear model with additional multiplicative constants $\{T_{i(0)\ldots i(p)}\}$, which can be introduced into the iterative proportional fitting algorithm by setting the initial cell values to the exposure vector, or into GLIM by using offsets. The above formula describes a proportional hazards model in which only the covariate main effects are included, second order interactions can be accomodated by including appropriate u-terms, $u_{sr\{i(s)i(r)\}}$. If interactions involving the first subscript, the time index, are introduced, the model is no longer proportional and covariate effects vary over time.

In this development the survival times were assumed to be exponentially distributed within time intervals, or, equivalently, the hazards are assumed constant during each interval. The approach of Aitkin & Clayton (1980) is more general and allows for a variety of survival distributions within intervals.

Although proportional hazards models can be fitted in standard log-linear packages, there may be drawbacks to this approach. Each covariate and time interval combination has to be treated as a separate observation from the point of view of the log-linear program. In the analysis of preterm gestational ages there were 514 separate covariate combinations for primiparae and 1521 for multiparae, and 10 time intervals, resulting in data sets of the order of 5,000 and 15,000 cases respectively. Many covariate combinations were observed only once, but no substantial reduction in the cross-classification can be achieved since the birth will often fall in the final time period, representing a normal outcome, and contributes to the exposure in earlier intervals.

The generalised linear models for polytomous data can also be fitted in log linear programs, as described by Thompson & Baker (1981). The mean of the frequency of the response category conditional on the covariate combination is equal to the frequency of the combination multiplied by the difference between a distribution function evaluated at the cutpoints defining the response category on an underlying continuous scale. The distribution function is specified by parameters that depend on the covariate combination, and the mean is thus the difference of the inverse link function applied to linear combinations of parameters corresponding to the two cutpoints. In GLIM the expected mean is usually equated to the single inverse link function of the corresponding linear combination of the parameters. An adaptation of the least squares formulation of GLIM described in section 3.1.3, in which each observation can be associated with more than one linear combination of the parameters can be made to cover the case where the mean is a composite function of inverse link functions. Like the preceding approach to fitting proportional hazards models to grouped survival data, this adaptation involves data sets of the order of the number of distinct covariate combinations multiplied by the number response categories and may prove an expensive method of fitting the model.

CHAPTER 4 : REGRESSION ANALYSIS OF BIRTHWEIGHT, GESTATIONAL AGE AND BIRTHWEIGHT STANDARDISED FOR GESTATIONAL AGE

4.1 Summary of the Main Findings

This section comprises a summary of the main findings from, first, a logistic regression analysis of birthweight below 2,500 gms; secondly, a proportional hazards model for the risk of delivery throughout the preterm period; and, thirdly, a polytomous regression model for birthweight standardised for gestational age amongst term infants, focussing on the risk of delivering an SGA or LGA (small- or large-for-gestational-age) infant. A more detailed discussion of the results follows in later sections. Only parameter estimates and associated 95 per cent confidence intervals are presented here (Tables 4.1 and 4.2 for primiparae and multiparae, respectively). Later sections contain a comparison of adjusted and unadjusted risks, an examination of the importance of interactions in the models and additional investigations of the adequacy of the models.

The analyses showed that the most important covariate in predicting low birthweight and birthweight standardised for gestational age was maternal height. Both primiparae and multiparae of height <150 cm experienced risk of delivering a low birthweight infant that was doubled, while for women $\geq$165 cm the risk was reduced by 40 per cent compared to the reference category of women of height 150-164 cm. A similar gradient in risk of an SGA infant was found, and, conversely, a reduction of 60 per cent in the risk of an LGA infant for women <150 cm and an increase of nearly 90 per cent in the risk of an LGA infant for women $\geq$165 cm. Although maternal height was significantly associated with the hazard of preterm birth, the gradient in

TABLE 4.1

SUMMARY OF ADJUSTED RISKS FOR PRIMIPARAE: BIRTHWEIGHT <2500gm PRETERM DELIVERY AND THE BIRTH OF AN SGA OR LGA INFANT

Covariate	< 2,500 gm*	Preterm Delivery**	SGA***	LGA***
Sex of Infant				
Male	1.00	1.00	-	-
Female	1.17(1.11,1.24)	.89(.80,.98)	-	-
Marital Status				
Married	1.00	1.00	1.00	1.00
Single	1.28(1.18,1.39)	1.45(1.26,1.67)	1.19(1.11,1.28)	.79(.74,.86)
Social Class				
I-II	.79(.72,.86)	.84(.71,.99)	.77(.71,.84)	.98(.92,1.04)
III	1.00	1.00	1.00	1.00
IV-V	1.15(1.07,1.24)	1.10(.96,1.27)	1.20(1.12,1.28)	.88(.83,.95)
Unknown	1.11(1.02,1.20)	1.14(1.00,1.31)	1.09(1.03,1.17)	.82(.76,.87)
Maternal Height				
<150 cm	2.21(1.97,2.47)	1.46(1.15,1.84)	2.20(1.99,2.44)	.37(.30,.47)
150-164 cm	1.00	1.00	1.00	1.00
≥165 cm	.61(.57,.66)	.75(.66,.86)	.50(.46,.53)	1.92(1.85,2.04)
Maternal Age				
<18 yrs	1.15(1.02,1.28)	1.47(1.21,1.77)	.81(.73,.90)	1.32(1.16,1.43)
18-24 yrs	1.00	1.00	1.00	1.00
25-34 yrs	1.03(.97,1.10)	1.13(1.00,1.27)	.93(.88,1.98)	1.04(.99,1.10)
≥35 yrs	1.76(1.47,2.10)	1.81(1.30,2.53)	.99(.81,1.20)	1.32(1.12,1.54)
Previous Spontaneous Abortion				
0	1.00	1.00	1.00	1.00
1	1.30(1.18,1.43)	1.21,(1.01,1.46)	1.01(.92,1.11)	1.09(1.00,1.18)
≥2	2.01(1.66,2.43)	1.54(1.05,2.24)	1.22(.99,1.50)	1.32(1.09,1.59)
Previous Induced Abortion				
0	1.00	1.00	1.00	1.00
≥1	1.20(1.08,1.34)	1.28(1.06,1.56)	.98(.88,1.09)	.98(.88,1.09)

* Relative Odds Ratios, 1980-82
** Relative Hazards, 1981
*** Relative Odds Ratios, Term Infants, 1980-82

TABLE 4.2

SUMMARY OF ADJUSTED RISKS FOR MULTIPARAE: BIRTHWEIGHT <2500g PRETERM DELIVERY AND THE BIRTH OF AN SGA OR LGA INFANT

Covariate	< 2,500g*	Preterm Delivery**	SGA***	LGA***
Sex of Infant				
Male	1.00	1.00	-	-
Female	1.26(1.19,1.33)	.85(.77,.94)	-	-
Marital Status				
Married	1.00	1.00	1.00	1.00
Single	1.63(1.40,1.89)	1.65(1.29,2.13)	1.42(1.26,1.60)	.76(.65,.89)
Social Class				
I-II	.65(.59,.72)	.75(.64,.88)	.70(.65,.75)	1.10(1.04,1.16)
III	1.00	1.00	1.00	1.00
IV-V	1.16(1.08,1.25)	1.09(.96,1.24)	1.19(1.12,1.26)	.85(.85,.94)
Unknown	1.21(1.12,1.31)	1.18(1.03,1.34)	1.28(1.21,1.36)	.84(.79,.89)
Maternal Height				
<150 cm	2.05(1.84,2.29)	1.08(.85,1.36)	2.27(2.09,2.47)	.40(.35,.48)
150-164 cm	1.00	1.00	1.00	1.00
≥165 cm	.62(.57,.67)	.91(.80,1.02)	.51(.48,.54)	1.85(1.79,1.92)
Maternal Age				
<18 yrs	2.19(1.44,3.33)	3.61(1.91,6.79)	1.20(.78,1.85)	1.27(.78,2.08)
18-24 yrs	1.00	1.00	1.00	1.00
25-34 yrs	.84(.79,.90)	.83(.74,.93)	.83(.79,.87)	1.27(1.20,1.32)
≥35 yrs	1.07(.95,1.20)	1.20(1.00,1.44)	.92(.84,1.01)	1.45(1.35,1.56)
Previous Spontaneous Abortion				
0	1.00	1.00	1.00	1.00
1	1.24(1.15,1.34)	1.26(1.11,1.43)	1.08(1.02,1.15)	1.04(.99,1.10)
≥2	2.04(1.83,2.27)	1.81(1.30,2.17)	1.48(1.35,1.63)	.91(.82,1.00)
Previous Induced Abortion				
0	1.00	1.00	1.00	1.00
≥1	1.32(1.19,1.47)	1.59(1.34,1.89)	1.06(.97,1.16)	.99(.92,1.09)
Previous Caesarean Section				
0	1.00	1.00	1.00	1.00
≥1	1.15(1.05,1.26)	1.19(1.02,1.39)	.94(.87,1.01)	1.28(1.19,1.37)
Previous Perinatal Death				
0	1.00	1.00	1.00	1.00
≥1	2.31(2.10,2.55)	2.36(2.02,2.76)	1.36(1.24,1.49)	.86(.78,.95)
Previous Livebirth				
0-2	1.00	1.00	1.00	1.00
≥3	1.13(1.03,1.24)	1.26(1.09,1.45)	1.13(1.06,1.21)	1.19(1.12,1.27)

* Relative Odds Ratios, 1980-82

** Relative Hazards, 1981

*** Relative Odds Ratios, term infants 1980-82 hazard associated with maternal height was less extreme. The adjusted hazard of preterm delivery associated with maternal height amongst multiparae was not significant.

A woman's obstetric history was also strongly associated with the three perinatal outcomes. A history of previous perinatal death amongst multiparae was associated with increases of over 130 per cent in both the risk of low birthweight and the hazard of preterm delivery, and also with higher risk of an SGA infant and correspondingly lower risk of an LGA infant. A history of, particularly ≥ 2 spontaneous abortions was associated with a risk of a low birthweight infant that was doubled and also with a high hazard of preterm delivery. Multiparae with ≥ 2 spontaneous abortions experienced increased risk of the birth of an SGA infant but no significant change in risk of an LGA infant, while primiparae with ≥ 2 spontaneous abortions experienced increased risk of the birth of both SGA and LGA infants. In both groups of women a history of induced abortion was associated with increased risk of a low birthweight infant. This was primarily due to their increased risk of preterm delivery, since their association with birthweight standardised for gestational age was marginal. Two other aspects of obstetric history, a previous caesarean section and ≥ 3 livebirths, were associated with only small increases in the risk of low birthweight and preterm delivery. Both were, however, associated with increased risk of the birth of an LGA infant.

Two socio-economic covariates, social class and marital status were included in the study. Decreased risk of both low birthweight and preterm delivery was associated with membership of classes I-II (professional and managerial) compared to the reference category, class III (clerical and skilled manual).

Women in classes IV-V (semi- and unskilled manual) experienced increased risk of low birthweight but no significant increase in the hazard of preterm delivery. Generally, women with unknown social class experienced similar or slightly higher risk than social class IV-V women. Social class I-II women experienced reduced risk of the birth of an SGA infant and social class IV-V women experienced increased risk. A corresponding gradient of decreased risk of an LGA infant associated with lower social class was not found. Single marital status was associated with increased risk of low birthweight, preterm delivery and lower birthweights standardised for gestational age, represented by an increased risk of the birth of an SGA infant and decreased risk of an LGA infant. The association with marital status was in all cases more extreme for multiparae than primiparae.

The pattern of risk of low birthweight and preterm delivery associated with maternal age showed the traditional pattern of high risk associated with either extreme of the age distribution. Amongst primiparae, minimum risk was experienced by women aged 18-24 years, while amongst multiparae, minimum risk was associated with the 25-34 age category. Primiparae experienced particularly high risk in the $\geq$35 age group, while multiparae aged <18 years experienced highest risk. The pattern of risk of an SGA or LGA infant associated with maternal age was not so clear. Primiparae aged <18 years experienced lowest risk of an SGA infant, wheras multiparae aged under 25 years experienced higher risk of an SGA infant than those over 25 years. In both groups of women higher risk of an LGA infant was associated with the $\geq$35 year age category.

Finally, while female infants experienced higher risk of low birthweight than male infants, they had lower risk of preterm delivery. Sex of infant was not examined in the analysis of birthweight standardised for gestational age since the percentiles were additionally standardised for sex. However, The tables in Appendix 2 show that birthweight percentiles were lower for female than male infants.

4.2 Logistic Regression Models for Birthweight

4.2.1 Main Effects Models for Birthweight

The results of this section were obtained from a series of binary logistic regression analyses of birthweight below 2,500 gms, 2,000 gms, 1,500 gms and 1,000 gms. Estimated parameters are expressed as odds ratios comparing the risk for each category of a covariate with the reference category, and 95 per cent confidence intervals of the odds ratios are given. Since the incidence of birthweight below 2,500 gms is small (approximately 6 per cent) and the incidence of birthweight below the other cut-points is lower, the odds ratios can be interpreted as approximate relative risks. Unadjusted and adjusted odds ratios are given as an indication of the extent to which association of any one covariate with low birthweight can be explained by the remaining covariates. Interactions between pairs of covariates were also examined (section 4.2.2), to examine departure from the main effects model. The importance of parameters in the models was assessed by likelihood ratio tests which were generally performed at the 5 per cent level.

Tables 4.3 to 4.6 present the adjusted and unadjusted odds ratios of birthweight below 2,500 gms, 2,000 gms, 1,500 gms and 1,000 gms, respectively, associated with the ten covariates. The likelihood ratio statistics for inclusion of each covariate in

TABLE 4.3
ADJUSTED AND UNADJUSTED ODDS RATIOS OF BIRTHWEIGHT<2500gms PRIMIPARAE AND MULTIPARAE 1980-82

COVARIATE	PRIMIPARAE		MULTIPARAE	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Sex of Infant				
Male	1.00	1.00	1.00	1.00
Female	1.17(1.10,1.23)	1.17(1.11,1.24)	1.26(1.19,1.33)	1.26(1.19,1.33)
Marital Status				
Married	1.00	1.00	1.00	1.00
Single	1.41(1.32,1.52)	1.28(1.18,1.39)	2.02(1.75,2.33)	1.63(1.40,1.89)
Social Class				
I-II	.76(.69,.83)	.79(.72,.86)	.60(.54,.66)	.65(.59,.72)
III	1.00	1.00	1.00	1.00
IV-V	1.19(1.11,1.29)	1.15(1.07,1.24)	1.24(1.15,1.33)	1.16(1.08,1.25)
Unknown	1.22(1.13,1.31)	1.11(1.02,1.20)	1.35(1.26,1.46)	1.21(1.12,1.31)
Maternal Height				
<150 cm	2.30(2.06,2.58)	2.21(1.97,2.47)	2.24(2.01,2.49)	2.05(1.84,2.29)
150-164 cm	1.00	1.00	1.00	1.00
≥165 cm	.60(.55,.64)	.61(.57,.66)	.57(.53,.62)	.62(.57,.67)
Maternal Age				
≤18 years	1.33(1.19,1.48)	1.15(1.02,1.28)	2.48(1.64,3.75)	2.19(1.44,3.33)
18-24 years	1.00	1.00	1.00	1.00
25-34 years	.90(.85,.96)	1.03(.97,1.10)	.77(.72,.82)	.84(.79,.90)
≥35 years	1.57(1.31,1.86)	1.76(1.47,2.10)	1.06(.95,1.17)	1.07(.95,1.20)
Previous Spontaneous Abortion				
0	1.00	1.00	1.00	1.00
1	1.26(1.15,1.39)	1.30(1.18,1.43)	1.25(1.16,1.35)	1.24(1.15,1.34)
≥2	2.00(1.65,2.41)	2.01(1.66,2.43)	2.17(1.95,2.41)	2.04(1.83,2.27)
Previous Induced Abortion				
0	1.00	1.00	1.00	1.00
≥1	1.20(1.07,1.34)	1.20(1.08,1.34)	1.37(1.23,1.53)	1.32(1.19,1.47)
Previous Caesarean Section				
0			1.00	1.00
≥1			1.27(1.16,1.39)	1.15(1.05,1.26)
Previous Perinatal Death				
0			1.00	1.00
≥1			2.59(2.36,2.85)	2.31(2.10,2.55)
Previous Livebirth				
0-2			1.00	1.00
≥3			1.27(1.17,1.38)	1.13(1.03,1.24)

TABLE 4.4
ADJUSTED AND UNADJUSTED ODDS RATIOS OF BIRTHWEIGHT<2000gms PRIMIPARAE AND MULTIPARAE 1980-82

COVARIATE	PRIMIPARAE		MULTIPARAE	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Sex of Infant				
Male	1.00	1.00	1.00	1.00
Female	.95(.86,1.05)	.95(.86,1.06)	1.04(.94,1.16)	1.04(.94,1.16)
Marital Status				
Married	1.00	1.00	1.00	1.00
Single	1.54(1.36,1.75)	1.38(1.19,1.59)	2.23(1.74,2.85)	1.89(1.45,2.45)
Social Class				
I-II	.82(.70,.97)	.84(.71,.99)	.76(.64,.90)	.81(.68,.96)
III	1.00	1.00	1.00	1.00
IV-V	1.29(1.13,1.48)	1.25(1.09,1.43)	1.30(1.14,1.50)	1.22(1.06,1.40)
Unknown	1.33(1.17,1.51)	1.18(1.03,1.35)	1.35(1.18,1.56)	1.19(1.03,1.38)
Maternal Height				
<150 cm	2.19(1.80,2.66)	2.08(1.70,2.53)	1.75(1.41,2.17)	1.56(1.25,1.94)
150-164 cm	1.00	1.00	1.00	1.00
≥165 cm	.69(.60,.78)	.71(.62,.81)	.69(.59,.79)	.74(.64,.85)
Maternal Age				
<18 years	1.63(1.36,1.95)	1.39(1.15,1.68)	3.42(1.80,6.50)	2.93(1.53,5.63)
18-24 years	1.00	1.00	1.00	1.00
25-34 years	.96(.86,1.08)	1.10(.97,1.23)	.78(.69,.88)	.83(.73,.94)
≥35 years	1.82(1.36,2.43)	1.96(1.46,2.63)	1.35(1.12,1.61)	1.30(1.07,1.58)
Previous Spontaneous Abortion				
0	1.00	1.00	1.00	1.00
1	1.41(1.19,1.67)	1.46(1.24,1.74)	1.30(1.12,1.49)	1.27(1.10,1.46)
≥2	2.73(2.06,3.63)	2.76(2.07,3.68)	2.97(2.50,3.53)	2.69(2.25,3.20)
Previous Induced Abortion				
0	1.00	1.00	1.00	1.00
≥1	1.36(1.13,1.64)	1.36(1.13,1.64)	1.57(1.30,1.90)	1.51(1.25,1.82)
Previous Caesarean Section				
0			1.00	1.00
≥1			1.41(1.19,1.65)	1.27(1.08,1.50)
Previous Perinatal Death				
0			1.00	1.00
≥1			3.48(2.99,4.07)	3.03(2.59,3.54)
Previous Livebirth				
0-2			1.00	1.00
≥3			1.17(1.00,1.37)	.98(.82,1.16)

TABLE 4.5
ADJUSTED AND UNADJUSTED ODDS RATIOS OF BIRTHWEIGHT <1500 gms PRIMIPARAE AND MULTIPARAE 1980-82

COVARIATE	PRIMIPARAE		MULTIPARAE	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Sex of Infant				
Male	1.00	1.00	1.00	1.00
Female	.99(.84,1.16)	.99(.84,1.16)	1.02(.86,1.22)	1.02(.85,2.22)
Marital Status				
Married	1.00	1.00	1.00	1.00
Single	1.33(1.09,1.64)	1.18(.93,1.48)	1.77(1.14,2.76)	1.46(.92,2.34)
Social Class				
I-II	.76(.59,.98)	.77(.59,1.01)	.72(.53,.96)	.74(.54,.99)
III	1.00	1.00	1.00	1.00
IV-V	1.28(1.04,1.57)	1.23(1.00,1.52)	1.31(1.04,1.64)	1.23(.98,1.55)
Unknown	1.18(.96,1.44)	1.06(.86,1.31)	1.48(1.18,1.86)	1.35(1.06,1.70)
Maternal Height				
<150 cm	1.46(1.02,2.10)	1.39(.97,1.99)	1.77(1.24,2.53)	1.53(1.06,2.20)
150-164 cm	1.00	1.00	1.00	1.00
≥165 cm	.68(.55,.83)	.71(.58,.87)	.96(.77,1.18)	1.05(.84,1.30)
Maternal Age				
<18 years	1.74(1.32,2.28)	1.67(1.25,2.22)	2.83(.90,8.89)	2.62(.82,8.33)
18-24 years	1.00	1.00	1.00	1.00
25-34 years	.94(.79,1.13)	1.02(.85,1.23)	.82(.67,.99)	.84(.68,1.03)
≥35 years	1.98(1.29,3.06)	2.01(.59,3.12)	1.36(1.01,1.83)	1.24(.90,1.72)
Previous Spontaneous Abortion				
0	1.00	1.00	1.00	1.00
1	1.71(1.34,2.19)	1.78(1.39,2.28)	1.51(1.20,1.89)	1.47(1.16,1.85)
≥2	3.19(2.11,4.82)	3.25(2.14,4.93)	3.96(3.06,5.12)	3.52(2.70,4.59)
Previous Induced Abortion				
0	1.00	1.00	1.00	1.00
≥1	1.47(1.11,1.95)	1.49(1.13,1.97)	1.88(1.41,2.50)	1.80(1.35,2.41)
Previous Caesarean Section				
0			1.00	1.00
≥1			1.45(1.12,1.88)	1.32(1.01,1.72)
Previous Perinatal Death				
0			1.00	1.00
≥1			4.19(3.30,5.31)	3.54(2.78,4.51)
Previous Livebirth				
0-2			1.00	1.00
≥3			1.08(.82,1.41)	.87(.66,1.16)

TABLE 4.6
ADJUSTED AND UNADJUSTED ODDS RATIOS OF BIRTHWEIGHT <1000 gms PRIMIPARAE AND MULTIPARAE
1980-82

COVARIATE	PRIMIPARAE		MULTIPARAE	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Sex of Infant				
Male	1.00	1.00	1.00	1.00
Female	.98(.72,1.34)	.98(.72,1.34)	1.20(.84,1.71)	1.19(.83,1.70)
Marital Status				
Married	1.00	1.00	1.00	1.00
Single	1.27(.85,1,91)	1.15(.73,1.81)	1.06(.34,3.28)	1.07(.33,3.43)
Social Class				
I-II	.81(.50,1.33)	.87(.53,1.43)	.79(.46,1.38)	.84(.48,1.48)
III	1.00	1.00	1.00	1.00
IV-V	1.19(.79,1.78)	1.12(.75,1.69)	1.20(.76,1.89)	1.11(.70,1.76)
Unknown	1.05(.70,1.57)	.93(.61,1.42)	1.13(.70,1.83)	1.06(.65,1.74)
Maternal Height				
<150 cm	1.57(.80,3.10)	1.50(.77,2.95)	2.35(1.21,4.53)	2.08(1.07,4.04)
150-164 cm	1.00	1.00	1.00	1.00
≥165 cm	.61(.40,.93)	.64(.42,.97)	1.07(.70,1.63)	1.14(.75,1.75)
Maternal Age				
<18 years	1.52(.88,2.62)	1.56(.87,2.77)	.10(.00,169.74)	.10(.00,231.98)
18-24 years	1.00	1.00	1.00	1.00
25-34 years	.83(.58,1.17)	.85(.58,1.22)	.92(.62,1.36)	.85(.57,1.28)
≥35 years	1.98(.87,4.54)	1.86(.81,4.28)	.66(.30,1.45)	.51(.22,1.19)
Previous Spontaneous Abortion				
0	1.00	1.00	1.00	1.00
1	1.83(1.15,2.94)	1.92(1.19,3.08)	2.20(1.44,3.35)	2.16(1.41,3.30)
≥2	4.18(2.04,8.56)	4.35(2.11,8.98)	4.20(2.47,7.13)	3.86(2.25,6.64)
Previous Induced Abortion				
0	1.00	1.00	1.00	1.00
≥1	1.72(1.03,2.89)	1.74(1.04,2.92)	1.12(.55,2.29)	1.09(.53,2.24)
Previous Caesarean Section				
0			1.00	1.00
≥1			1.09(.60,1.96)	.95(.52,1.74)
Previous Perinatal Death				
0			1.00	1.00
≥1			4.80(3.04,7.57)	4.17(2.62,6.63)
Previous Livebirth				
0-2			1.00	1.00
≥3			1.02(.58,1.77)	.95(.53,1.70)

the null model and exclusion from the model containing all main effects are given in Table 4.7. In general it can be seen from Tables 4.3 to 4.6 that, as would be expected, the covariates had more explanatory power in the unadjusted model, with the adjusted relative odds ratios being less extreme. However, the same patterns of risk were observed in the adjusted and unadjusted models. Looking firstly at the risks associated with the sex of the infant, we see that female infants of both primiparous and multiparous mothers have higher risk of birthweight below 2,500 gms (increased risks of 17 per cent and 26 per cent respectively), and these risks were virtually unaltered when the other covariates were included in the model. Below the more extreme birthweight cutpoints, sex of infant did not have a significant impact, and in most cases, the difference in risk between male and female infants was negligable.

The marital status of the mother had a significant impact on the risk of birthweight below 2,500 gms and 2,000 gms. Considering birthweight below 2,500 gms, single primiparous mothers experienced a 28 per cent increase in risk and single, multiparous mothers experienced a 63 per cent increase in risk compared to their married counterparts. The risk of birthweight below 1,500 gms amongst single women was not significantly different after adjusting for the other covariates, and, in the extreme case of birthweight below 1,000 gms even the unadjusted association was not significant. In all cases odds ratios associated with single marital status were lower after adjusting for the other covariates. The second socio-economic covariate, social class, was associated with the risk of birthweight below 2,500 gms, 2,000 gms and 1,500 gms. For example, social class I-

TABLE 4.7

LIKELIHOOD RATIO STATISTICS FROM LOGISTIC REGRESSION ANALYSES OF BIRTHWEIGHT: 1980-82

COVARIATE	Degrees of freedom	<1000gm		<1500gm		< 2000gm		< 2500gm	
		Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
Primiparae									
Sex of Infant	1	.02	.01	.03	.02	.88	.80	27.98***	29.19***
Marital Status	1	1.27	.35	7.10***	1.86	41.84***	18.72***	81.88***	33.97***
Social Class	3	2.09	1.06	17.05***	11.01*	46.27***	23.95***	122.36***	63.11***
Maternal Height	2	8.19*	6.71*	20.76***	16.01***	95.39***	80.02***	423.61***	374.16***
Maternal Age	3	6.37	5.31	23.80***	18.66***	41.92***	27.19***	69.52***	38.40***
Prev. Spontaneous Abortion	2	15.19***	15.97***	36.54***	38.24***	49.66***	52.03***	62.21***	65.40***
Prev. Induced Abortion	1	3.67	3.86*	6.62*	6.98**	9.85**	9.82**	9.91**	10.19**
Multiparae									
Sex of Infant	1	.97	.91	.06	.03	.61	.56	60.14***	60.58***
Marital Status	1	.01	.00	5.29*	2.30	32.40***	19.55***	79.24***	36.60***
Social Class	3	2.18	.87	27.62***	17.05***	54.12***	24.81***	298.14***	156.96***
Maternal Height	2	5.21	4.03	8.68*	4.59	58.49***	36.17***	443.60***	330.23***
Maternal Age	3	1.59	3.15	15.58**	10.34*	54.45***	36.42***	104.47***	51.56***
Prev. Spontaneous Abortion	2	28.96***	25.55***	85.83***	70.55***	125.22***	100.83***	190.78***	156.94***
Prev. Induced Abortion	1	.09	.06	15.60***	13.68***	19.78***	16.51***	30.68***	23.60***
Prev. Caesarean Section	1	.07	.03	6.88**	3.73	15.50***	7.62**	25.71***	8.30**
Prev. Perinatal Death	1	32.60***	27.29***	101.61***	80.29***	188.72***	150.13***	311.68***	240.47***
Prev. Livebirth	1	.00	.03	.30	.88	3.70	.05	30.45***	7.17*

* significant at 5% level

** significant at 1% level

*** significant at .1% level

II, primparous mothers experienced a 21 per cent reduction in the risk of birthweight below 2,500 gms, and social class IV-V, primiparous mothers experienced a 15 per cent increase in risk compared to their social class III counterparts. Similarly, multiparous, social class I-II women experienced a 35 per cent reduction in risk, while social class IV-V multiparae experienced a 16 per cent increase in risk. Women with unknown social class generally experienced similar risk to social class IV-V women. The gradient in risk associated with social class was apparent at the other three cutpoints, but was not significant at 1,000 gms.

Maternal height was one of the most important covariates in the models for birthweight below 2,500 gms in terms of reducing the likelihood ratio statistic. Amongst both primiparae and multiparae, women of height less than 150 cm experienced a risk of birthweight below 2,500 gms that was doubled, while women of height greater than 165 cm experienced an approximate 40 per cent reduction in risk compared to women of height 150-164 cm. Amongst primiparae a similar though less extreme gradient in risk was observed at 2,000 gms, 1,500 gms and 1,000 gms. For example, primiparae under 150 cm in height experienced a 50 per cent increase in the risk of birthweight below 1,000 gms, while the risk for primiparae over 165 cm was decreased by 36 per cent. Amongst multiparae, although a similar pattern of risk was observed below all the birthweight cutpoints, it was only significant in the analysis of birthweight below 2,500 gms and 2,000 gms.

Both primiparae and multiparae in the most frequently occurring age category experienced the lowest risk of birthweight below 2,500 gms, 2,000 gms and 1,500 gms. For primiparae this was the reference category, 18-24 years, for multiparae the age group

25-34 years had the highest frequency. The most important association with maternal age was the increased risk experienced by women aged under 18 years and women aged over 35 years. Primiparae aged over 35 years experienced highest risk of birthweight below 2,500 gms (76 per cent increase), below 2,000 gms (96 per cent increase) and below 1,500 gms (101 per cent increase). Multiparae experienced particularly high risk in the less than 18 age group (119 per cent increase in the risk of birthweight below 2,500 gms, 193 per cent increase below 2,000 gms, and a 162 per cent increase below 1,500 gms).

A history of spontaneous abortion was a highly significant factor in the analyses of birthweight below all four cutpoints. The pattern of risk associated with spontaneous abortion was increasingly extreme at the lower cutpoints. Amongst primiparae a history of one spontaneous abortion was associated with 192 per cent increase in the risk of birthweight below 1,000 gms. A 335 per cent increase in the risk of birthweight below 1,000 gms was observed amongst primiparae with two or more previous spontaneous abortions. Similarly, multiparae with one previous spontaneous abortion experienced a 116 per cent increase in risk of birthweight below 1,000 gms and those with two or more previous spontaneous abortions experienced a 286 per cent increase. A history of induced abortions amongst primiparae was associated with increased risk of birthweight below the four cutpoints (20 per cent increase below 2,500 gms, 36 per cent increase below 2,000 gms, 49 per cent increase below 1,500 gms and 74 per cent below 1,000 gms). Amongst multiparae, a history of induced abortion was associated with increased risk of birthweight below 2,500 gms, 2,000 gms and 1,500 gms (32 per cent, 51 per cent and

80 per cent respectively). The association of previous induced abortion with birthweight below 1,000 gms was negligible for multiparae.

The final three covariates were aspects of obstetric history relevent only to multiparae. A history of caesarean section was associated with the risk of birthweight below 2,500 gms, 2,000 gms and 1,500 gms, but the association was greatly diminished after adjusting for the other covariates. A history of caesarean section did not have a significant impact on the risk of birthweight below 1,000 gms. A previous perinatal death was the most important single risk factor for multiparae, associated with risk increased by 131 per cent below 2,500 gms, 203 per cent below 2,000 gms, 254 per cent below 1,500 gms and 317 per cent below 1,000 gms. A history of three or more livebirths played a significant role only in the analysis of birthweight below 2,500 gms, and its associated increased risk (27 per cent) was halved (13 per cent) after adjusting for the other covariates.

4.2.2 Interactions in Birthweight Models

Interactions between each pair of covariates were examined in the analysis of birthweight. Tables 4.8 to 4.11 give the likelihood ratio statistics for their inclusion in the main effects models of birthweight below 2,500 gms, 2,000 gms, 1,500 gms and 1,000 gms, the appropriate degrees of freedom are also given in the tables. In the case of the interaction between previous spontaneous abortion and maternal age for multiparae, there were no women aged less than eighteen with two or more abortions and the degrees of freedom for the interaction is five not six.

LIKELIHOOD RATIO STATISTICS FOR INTERACTIONS IN THE ADJUSTED LOGISTIC REGRESSION MODELS OF BIRTHWEIGHT < 2500 gms

PRIMIPARAE	Social Class	Maternal Age	Sex of Infant	Maternal Height	Previous Induced Abortion	Previous Spontaneous Abortion
Marital Status	6.88 (3)	12.00* (3)	.06 (1)	3.67 (2)	1.89 (2)	2.38 (2)
Previous Spontaneous Abortion	3.32 (6)	3.63 (6)	2.29 (2)	2.27 (4)	1.82 (2)	
Previous Induced Abortion	2.48 (3)	2.68 (3)	1.52 (1)	1.12 (2)		
Maternal Height	1.52 (6)	13.10* (6)	1.50 (2)			
Sex of Infant	2.37 (3)	4.83 (3)				
Maternal Age	18.08* (9)					

MULTIPARAE	Social Class	Maternal Age	Sex of Infant	Maternal Height	Previous Livebirth	Previous Perinatal Death	Previous Caesarean Section	Previous Induced Abortion	Previous Spontaneous Abortion
Marital Status	4.82 (3)	5.88 (3)	.14 (1)	7.92* (2)	.01 (1)	6.33* (1)	6.15* (1)	2.54 (1)	.38 (2)
Previous Spontaneous Abortion	7.62 (6)	20.03* (5)	1.32 (2)	4.72 (4)	3.08 (2)	1.10 (2)	.17 (2)	.93 (2)	
Previous Induced Abortion	.85 (3)	5.01 (3)	3.67 (1)	2.09 (2)	0.0 (1)	.06 (1)	.13 (1)		
Previous Caesarean Section	6.35 (3)	1.42 (3)	.09 (1)	.88 (2)	.15 (1)	1.51 (1)			
Previous Perinatal Death	1.68 (3)	7.19 (3)	.69 (1)	5.81 (2)	9.41* (1)				
Previous Livebirth	.89 (3)	6.19 (3)	.86 (1)	.80 (2)					
Maternal Height	3.60 (4)	6.46 (6)	.68 (2)						
Sex of Infant	3.93 (3)	3.84 (3)							
Maternal Age	8.60 (9)								

*significant at 5 per cent level

TABLE 4.8

LIKELIHOOD RATIO STATISTICS FOR INTERACTIONS IN THE ADJUSTED LOGISTIC REGRESSION MODELS OF BIRTHWEIGHT ≤2000 gms

PRIMIPARAE	Social Class	Maternal Age	Sex of Infant	Maternal Height	Previous Induced Abortion	Previous Spontaneous Abortion
Marital Status	3.69 (3)	5.41 (3)	0.48 (1)	0.52 (2)	3.91 (1)	1.10 (2)
Previous Spontaneous Abortion	2.22 (6)	4.47 (6)	3.19 (2)	4.63 (4)	2.46 (2)	
Previous Induced Abortion	0.92 (3)	2.50 (3)	0.92 (1)	5.40 (2)		
Maternal Height	2.21 (6)	8.48 (6)	0.29 (2)			
Sex of Infant	1.43 (3)	6.53 (3)				
Maternal Age	20.57* (9)					

MULTIPARAE	Social Class	Maternal Age	Sex of Infant	Maternal Height	Previous Livebirth	Previous Perinatal Death	Previous Caesarean Section	Previous Induced Abortion	Previous Spontaneous Abortion
Marital Status	2.21 (3)	0.37 (3)	1.80 (1)	7.88* (2)	0.12 (1)	0.75 (1)	6.17* (1)	2.77 (1)	1.06 (2)
Previous Spontaneous Abortion	7.08 (6)	8.39 (5)	1.75 (2)	1.11 (4)	2.13 (2)	0.34 (2)	1.28 (2)	0.60 (2)	
Previous Induced Abortion	0.69 (3)	3.62 (3)	2.77 (1)	3.11 (2)	0.00 (1)	1.73 (1)	1.34 (1)		
Previous Caesarean Section	5.89 (3)	2.71 (3)	0.03 (1)	0.74 (2)	0.70 (1)	2.60 (1)			
Previous Perinatal Death	1.01 (3)	8.56 (3)	0.58 (1)	4.48 (2)	9.96* (1)				
Previous Livebirth	1.42 (3)	0.81 (3)	0.06 (1)	1.18 (2)					
Maternal Height	4.46 (6)	9.22 (6)	3.09 (2)						
Sex of Infant	0.27 (3)	4.03 (3)							
Maternal Age	8.05 (9)								

*significant at the 5 per cent level

TABLE 4.9

LIKELIHOOD RATIO STATISTICS FOR INTERACTIONS IN THE ADJUSTED LOGISTIC REGRESSION MODELS OF BIRTHWEIGHT <1500 gms

PRIMIPARAE	Social Class	Maternal Age	Sex of Infant	Maternal Height	Previous Induced Abortion	Previous Spontaneous Abortion
Marital Status	3.90 (3)	1.53 (3)	0.92 (1)	2.90 (2)	0.01 (1)	4.06 (2)
Previous Spontaneous Abortion	3.83 (6)	2.61 (6)	2.87 (2)	3.49 (4)	5.31 (2)	
Previous Induced Abortion	1.20 (3)	6.97 (3)	0.48 (1)	3.61 (2)		
Maternal Height	9.71 (6)	9.93 (6)	0.04 (2)			
Sex of Infant	4.17 (3)	4.42 (3)				
Maternal Age	13.89 (9)					

MULTIPARAE	Social Class	Maternal Age	Sex of Infant	Maternal Height	Previous Livebirth	Previous Perinatal Death	Previous Caesarean Section	Previous Induced Abortion	Previous Spontaneous Abortion
Marital Status	1.40 (3)	4.33 (3)	0.20 (1)	0.68 (2)	0.01 (1)	0.66 (1)	4.83* (1)	0.95 (1)	0.25 (2)
Previous Spontaneous Abortion	6.91 (6)	5.01 (5)	4.13 (2)	1.54 (4)	0.65 (2)	1.05 (2)	0.96 (2)	0.29 (2)	
Previous Induced Abortion	1.17 (3)	0.94 (3)	0.58 (1)	9.03 (2)	0.03 (1)	0.12 (1)	0.66 (1)		
Previous Caesarean Section	2.84 (3)	0.66 (3)	0.96 (1)	2.54 (2)	1.48 (1)	0.12 (1)			
Previous Perinatal Death	4.05 (3)	2.22 (3)	0.02 (1)	7.33* (2)	0.77 (1)				
Previous Livebirth	1.55 (3)	1.01 (3)	2.32 (1)	1.71 (2)					
Maternal Height	3.59 (6)	4.42 (6)	0.47 (2)						
Sex of Infant	1.15 (3)	0.69 (3)							
Maternal Age	16.51 (9)								

* significant at the 5 per cent level

TABLE 4.10

LIKELIHOOD RATIO STATISTICS FOR INTERACTIONS IN THE ADJUSTED LOGISTIC REGRESSION MODELS OF BIRTHWEIGHT <1,000 gms

PRIMIPARAE	Social Class	Maternal Age	Sex of Infant	Maternal Height	Previous Induced Abortion	Previous Spontaneous Abortion
Marital Status	3.21 (3)	6.25 (3)	1.08 (1)	2.81 (2)	0.03 (1)	1.80 (2)
Previous Spontaneous Abortion	10.82 (6)	5.99 (6)	1.19 (2)	2.83 (4)	3.90 (2)	
Previous Induced Abortion	0.94 (3)	3.56 (3)	0.61 (1)	3.10 (2)		
Maternal Height	3.42 (6)	2.50 (6)	3.04 (2)			
Sex of Infant	0.91 (3)	0.94 (3)				
Maternal Age	6.10 (9)					

MULTIPARAE	Social Class	Maternal Age	Sex of Infant	Maternal Height	Previous Livebirth	Previous Perinatal Death	Previous Caesarean Section	Previous Induced Abortion	Previous Spontaneous Abortion
Marital Status	3.67 (3)	1.87 (3)	0.51 (1)	1.11 (2)	0.30 (1)	0.54 (1)	0.52 (1)	0.96 (1)	2.73 (1)
Previous Spontaneous Abortion	9.11 (6)	3.41 (5)	2.93 (2)	1.58 (4)	1.89 (2)	0.44 (2)	0.68 (2)	0.06 (2)	
Previous Induced Abortion	3.06 (3)	3.40 (3)	1.80 (1)	6.21 (2)	0.00 (1)	1.51 (1)	0.13 (1)		
Previous Caesarean Section	3.39 (3)	0.22 (3)	0.05 (1)	1.09 (2)	3.72 (1)	1.39 (1)			
Previous Perinatal Death	2.02 (3)	5.65 (3)	0.06 (1)	11.53* (2)	0.02 (1)				
Previous Livebirth	2.48 (3)	2.54 (3)	3.82 (1)	0.40 (2)					
Maternal Height	4.74 (4)	3.87 (6)	1.30 (2)						
Sex of Infant	7.38 (3)	0.34 (3)							
Maternal Age	6.28 (9)								

* significant at the 5 per cent level.

TABLE 4.11

Three pairs of covariates interacted significantly at the 5 per cent level in the analysis of birthweight below 2,500 gms for primiparae. Estimated odds ratios describing these interactions and their 95 per cent confidence intervals are given in Table 4.12. First, social class I-II primiparae experienced lowest risk of birthweight below 2,500 gms in the 25-34 age group, social class III primiparae experienced lowest risk in the 18-24 age group, while social class IV-V primiparae experienced lowest risk when aged under eighteen. These differentials in risk can be compared with the findings of Resseguie (1977), described in section 1.2.9, that the age of minimum risk of stillbirth in first pregnancy increased with increasing educational status. The second significant interaction in the analysis of primiparae was that between maternal age and marital status. Primiparae aged less than eighteen had lower risk of birthweight below 2,500 gms if they were single than if they were married. In all other age groups single marital status was associated with increased risks of birthweight below 2,500 gms. Finally, a significant interaction was found between maternal height and age. Short women, <150 cm, had lowest risk when aged less than eighteen, women of height 150-164 cm had lowest risk if they were in the 18-24 age group, while the tall women had lowest risk in the 25-34 age group. The patterns of risk associated with age within height groups parallel the risk associated with age within social class groups, and may be partly explained by the correlation between social class and maternal height.

In the analysis of birthweight below 2,500 gms for multiparae five pairs of covariates were associated with significant interaction at the 5 per cent level (Table 4.13). First, a history of two or more spontaneous abortions was

TABLE 4.12

INTERACTIONS FOR BIRTHWEIGHT < 2500 gms: ADJUSTED PRIMIPARAE: 1980–82

a) Social Class – Maternal Age

	I–II	III	IV–V	Unknown
< 18 years	1.17 (.51,2.69)	1.49 (1.20,1.84)	.90 (.69,1.17)	1.11 (.95,1.29)
18–24 years	1.00	1.00	1.00	1.00
25–34 years	.94 (.79,1.12)	1.04 (.94,1.14)	1.16 (1.01,1.32)	.96 (.83,1.11)
≥ 35 years	1.29 (.87,1.90)	1.83 (1.37,2.45)	1.87 (1.25,2.78)	2.09 (1.43,3.03)

b) Marital Status – Maternal Age

	< 18 years	18–24 years	25–34 years	≥ 35 years
Married	1.00	1.00	1.00	1.00
Single	.93 (.76,1.14)	1.32 (1.21,1.45)	1.53 (1.22,1.92)	1.41 (.65,3.04)

c) Maternal Height – Maternal Age

	< 150 cm	150–164 cm	≥ 165 cm
< 18 years	.78 (.52,1.16)	1.17 (1.03,1.33)	1.30 (.96,1.75)
18–24 years	1.00	1.00	1.00
25–34 years	1.12 (.87,1.44)	1.05 (.97,1.13)	.94 (.81,1.08)
≥ 35 years	3.26 (1.58,6.70)	1.82 (1.49,2.24)	1.29 (.85,1.95)

TABLE 4.13

INTERACTIONS FOR BIRTHWEIGHT <2500 gms: ADJUSTED MULTIPARAE: 1980-82

a) Maternal Age - Spontaneous Abortion

	< 18 years	18-24 years	25-34 years	≥ 35 years
0	1.00	1.00	1.00	1.00
1	.05 (.00,194.58)	1.41 (1.23,1.63)	1.16 (1.04,1.28)	1.24 (1.00,1.53)
≥2		2.13 (1.66,2.72)	2.23 (1.96,2.55)	1.26 (.93,1.70)

b) Maternal Height - Marital Status

	< 150 cm	150-164 cm	≥165 cm
Married	1.00	1.00	1.00
Single	1.21 (.73,1.99)	1.55 (1.31,1.84)	2.64 (1.84,3.80)

c) Previous Livebirths - Previous Perinatal Death

	0-2 Livebirths	≥3 Livebirths
0	1.00	1.00
≥1	2.47 (2.22,2.74)	1.60 (1.23,2.09)

d) Marital Status - Previous Perinatal Death

	Married	Single
0	1.00	1.00
≥1	2.26 (2.04,2.49)	4.25 (2.67,6.78)

e) Marital Status - Previous Caesarean Section

	Married	Single
0	1.00	1.00
≥1	1.17 (1.07,1.29)	.57 (.32, 1.04)

associated with much higher risk in the two age groups 18-24 and 25-34 years (relative odds ratios of 2.13 and 2.23 respectively) than for women aged over 35 years (relative odds ratio 1.26). The risk associated with single marital status increased with increasing maternal height from an odds ratio of 1.21 for multiparae of height <150 cm to 2.64 for multiparae of height $\geq$165 cm. Multiparae with three or more livebirths experienced an increased risk of 60 per cent if they had additionally had one or more perinatal deaths, while for women with less than three livebirths the increased risk associated with a previous perinatal death was 147 per cent. Finally, single women had much higher risk of birthweight below 2,500 gms associated with a previous perinatal death than married women, but lower risk associated with a previous caesarean section

The improvement in the likelihood of the models for birthweight below 2,000 gms and 1,500 gms due to including the eight interactions that were significant in the model of birthweight below 2,500 gms was reduced, and none of these eight interactions was significant at 1,000 gms. In the models for birthweight below 1,500 gms and 1,000 gms one extra interaction, that between maternal height and a history of perinatal death for multiparae, was significant. The estimated odds ratios of this interaction in the model for birthweight below 1,000 gms for multiparae are given in Table 4.14. Multiparae of height <150 cm experienced an odds ratio of 7.24 for birthweight below 1,000 gms associated with previous perinatal death, for multiparae of height 150-164 cm the odds ratio was 5.36, while for multiparae of height $\geq$165 cm the odds ratio was 0.02. The confidence intervals around these estimates were wide.

TABLE 4.14

MATERNAL HEIGHT - PREVIOUS PERINATAL DEATH INTERACTION IN THE ANALYSIS OF BIRTHWEIGHT $<$1000 gm: MULTIPARAE

Perinatal Deaths	$<$ 150 cm	150-164 cm	$\geq$ 165 cm
0	1.00	1.00	1.00
$\geq$ 1	7.24(2.02,25.99)	5.36(3.18,9.02)	0.02(0.00,33.86)

4.2.3 Assessment of Fit of the Main Effects Model

The main focus of the study of low birthweight was to estimate the risk associated with individual covariates, and to examine the extent to which these risks could be explained by adjusting for the other covariates. It can be seen from the previous section that, although several of the interactions between pairs of covariates were associated with significant improvements in the fit of the model for birthweight below 2,500 gms, these terms were not nearly so important as the main effects in the model. In this section additional diagnostic tests of the fit of the main effects logistic models for primiparae and multiparae are given.

Extreme residuals from the main effects models of birthweight below each of the four cutpoints were examined. Tables 4.15 gives the number of extreme Pearson residuals (section 3.2.2) above 1.96 and below -1.96. If the residuals are Normally distributed then approximately 5 per cent (33 in the data for primiparae and 105 in the data for multiparae) should lie outside these limits. In the analysis of birthweight below 2,500 gms and 2,000 gms, greater numbers of extreme positive residuals were observed. However, as suggested in section 3.2.2, the Pearson residuals may not be normally distributed even when the model is adequate. To investigate this possibility Binomial deviates were simulated corresponding to the number of individuals observed for each distinct combination of covariates. In order to simplify the simulation, the Binomial deviates were assumed to have a constant success probability which was taken to be the average of the fitted probabilities over the cells in each of the analyses. The simulated Pearson residuals were calculated by replacing the fitted probability with the known mean of the

TABLE 4.15

COMPARISON OF OBSERVED NUMBERS OF PEARSON RESIDUALS GREATER THAN 1.96 AND LESS THAN -1.96 WITH SIMULATED MEANS AND STANDARD DEVIATIONS

	PRIMIPARAE		MULTIPARAE	
	Observed	Simulated Mean (s.d.)	Observed	Simulated Mean (s.d.)
Birthweight < 2500 gms	Average observed probability = .10929		Average observed probability = .09777	
Number > 1.96	35	34.40 (5.90)	128	107.20 (9.12)
Number < -1.96	3	3.50 (1.80)	4	5.03 (2.37)
Birthweight < 2000 gms	Average observed probability = .04116		Average observed probability = .03496	
Number > 1.96	42	35.26 (5.91)	131	121.96 (10.22)
Number < -1.96	0	1.90 (1.26)	0	2.27 (1.39)
Birthweight < 1500 gms	Average observed probability = .01672		Average observed probability = .01496	
Number > 1.96	45	29.24 (5.68)	90	90.80 (8.96)
Number < -1.96	0	1.05 (1.06)	0	1.06 (1.00)
Birthweight < 1000 gms	Average observed probability = .00502		Average observed probability = .00309	
Number > 1.96	25	22.08 (4.26)	46	51.71 (6.97)
Number < -1.96	0	0.40 (0.68)	0	0.27 (.55)

Binomial deviate. A hundred repetitions of each analysis were simulated and the average number (and standard deviation) of extreme residuals were obtained and are given in Table 4.15. In all cases the observed numbers of extreme residuals were close to their simulated means, suggesting that the high numbers of extreme residuals may not be indicative of inadequacies in the model.

Table 4.16 presents the results of a similar comparison of observed numbers of extreme deviance residuals (section 3.1.4) with mean numbers obtained from a hundred simulations. In the analysis of birthweight below 2,500 gms and 2,000 gms neither the observed nor simulated residuals appear to be skewed and further, both show fewer than expected outside the nominal 5 per cent limits. However, in the case of birthweight below 1,000 gms a high proportion of the deviance residuals (16 per cent for primiparae and 48 per cent for multiparae) were below -1.96. This was caused by the adjustment to reduce bias, the term $(2\mu-1)/\{6[n\mu(1-\mu)]^{1/2}\}$, which, when μ is very small becomes large and negative. A repetition of the analysis calculating deviance residuals without the adjustment for bias, Table 4.17, avoided this effect, however the residuals were again distributed with mean greater than the nominal level of zero.

Figures 4.1 and 4.2 show plots of observed against fitted values of the risk of birthweight below 2,500 gms, 2,000 gms, 1,500 gms and 1,000 gms from the main effects models of the data for primiparae and multiparae respectively. The fitted scales in the plots were grouped into the categories shown in Table 4.18. These categorisations were chosen so that the categories were adequately represented and covered the range of values of fitted

TABLE 4.16

COMPARISON OF OBSERVED NUMBERS OF DEVIANCE RESIDUALS GREATER THAN 1.96 AND LESS THAN -1.96 WITH SIMULATED MEANS AND STANDARD DEVIATIONS

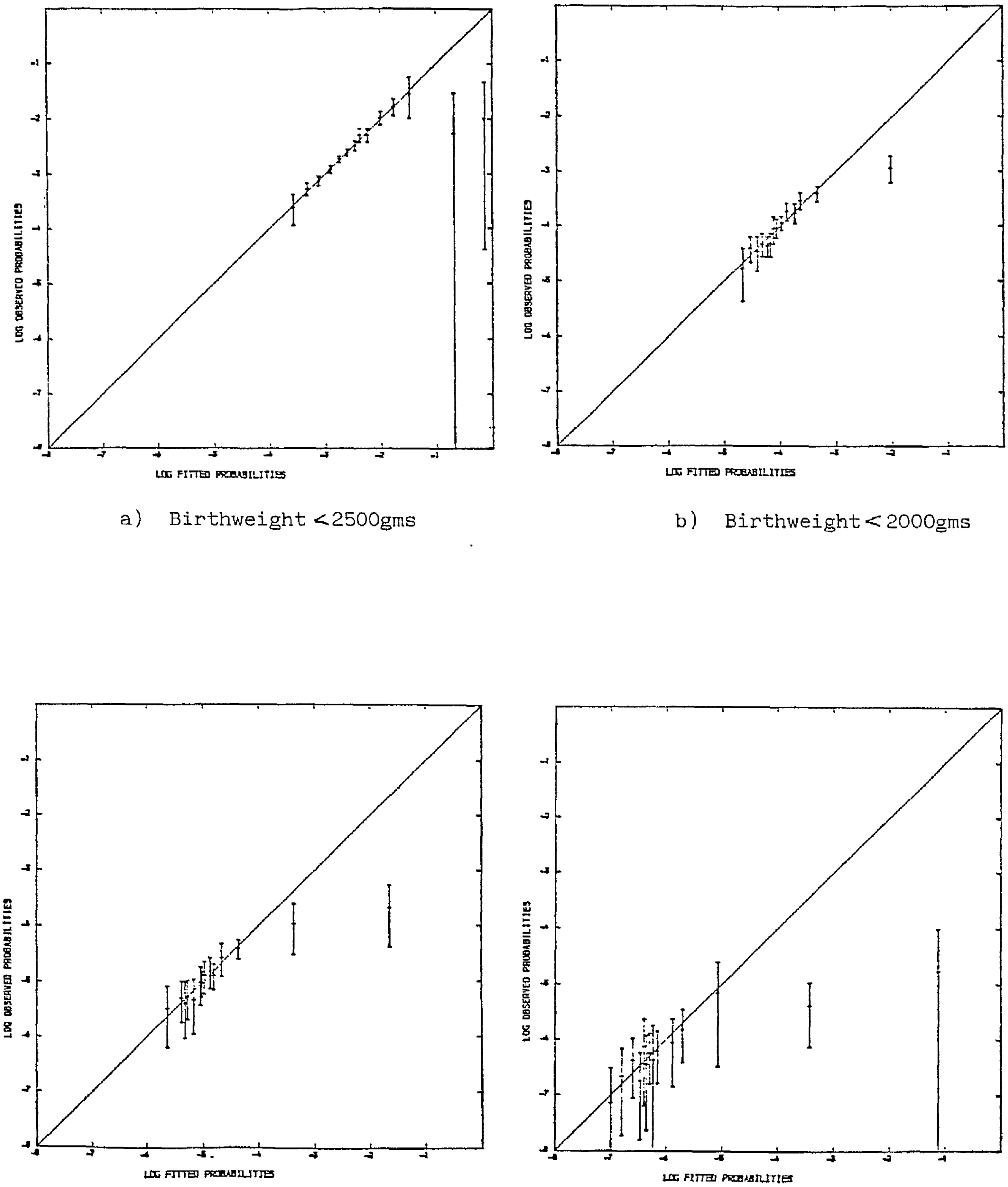
	PRIMIPARAE		MULTIPARAE	
	Observed	Simulated Mean (s.d.)	Observed	Simulated Mean (s.d.)
Birthweight <2500 gms	Average observed probability = .10929		Average observed probability = .09777	
Number > 1.96	12	9.63 (3.40)	27	22.01 (4.74)
Number < -1.96	10	11.98 (3.40)	30	24.79 (5.25)
Birthweight <2000 gms	Average observed probability = .04116		Average observed probability = .03496	
Number > 1.96	7	7.05 (3.20)	15	15.30 (3.86)
Number < -1.96	7	9.64 (2.74)	13	15.88 (3.70)
Birthweight <1500 gms	Average observed probability = .01672		Average observed probability = .01496	
Number > 1.96	3	5.06 (2.14)	6	9.27 (3.05)
Number < -1.96	5	5.47 (2.06)	126	10.09 (3.12)
Birthweight <1000 gms	Average observed probability = .00502		Average observed probability = .00309	
Number > 1.96	1	2.64 (1.66)	2	3.59 (2.75)
Number < -1.96	103	138.05 (1.83)	1013	861.38 (2.75)

TABLE 4.17

COMPARISON OF OBSERVED NUMBERS OF DEVIANCE RESIDUALS UNADJUSTED FOR BIAS GREATER THAN 1.96 AND LESS THAN -1.96 WITH SIMULATED MEANS AND STANDARD DEVIATIONS

	PRIMIPARAE		MULTIPARAE	
	Observed	Simulated Mean (s.d.)	Observed	Simulated Mean (s.d.)
Birthweight <2500 gms	Average observed probability = .10929		Average observed probability = .09777	
Number >1.96	20	25.84 (4.79)	74	91.84 (9.85)
Number <-1.96	7	9.21 (2.92)	19	20.47 (4.21)
Birthweight <2000 gms	Average observed probability = .04116		Average observed probability = .03496	
Number >1.96	20	14.84 (3.31)	58	59.68 (6.43)
Number <-1.96	3	7.90 (2.58)	11	11.03 (3.12)
Birthweight <1500 gms	Average observed probability = .01672		Average observed probability = .01496	
Number >1.96	17	12.88 (3.49)	46	45.64 (6.59)
Number <-1.96	3	4.94 (2.33)	2	7.11 (2.36)
Birthweight <1000 gms	Average observed probability = .00502		Average observed probability = .00309	
Number >1.96	13	11.44 (3.06)	24	25.63 (4.81)
Number <-1.96	0	1.61 (1.51)	1	1.14 (1.13)

FIGURE 4.1 LOG OBSERVED AGAINST FITTED PROBABILITIES FROM MAIN EFFECTS MODEL FOR BIRTHWEIGHT: PRIMIPARAE



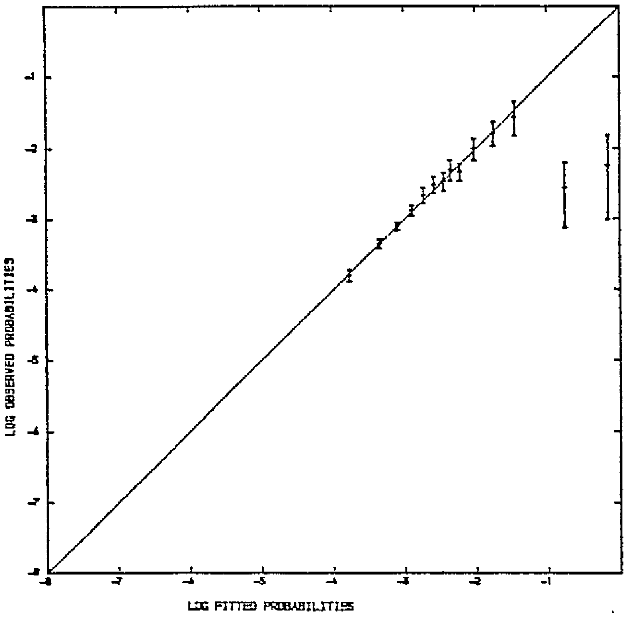
a) Birthweight <2500gms

b) Birthweight <2000gms

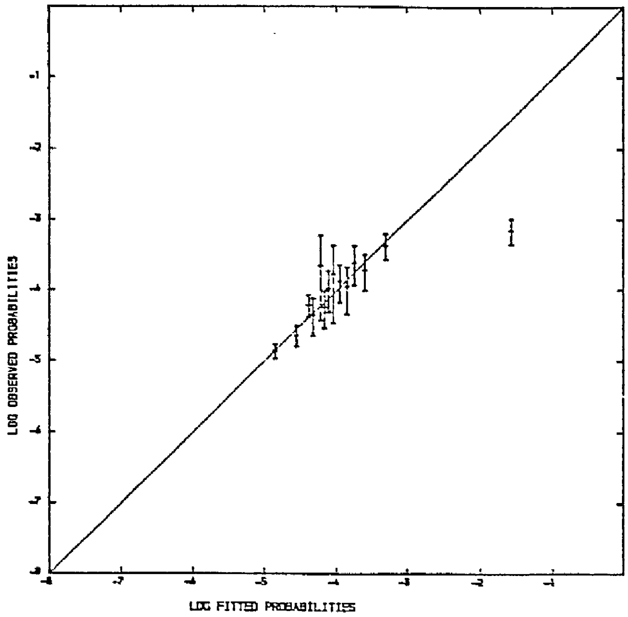
c) Birthweight <1500gms

d) Birthweight <1000gms

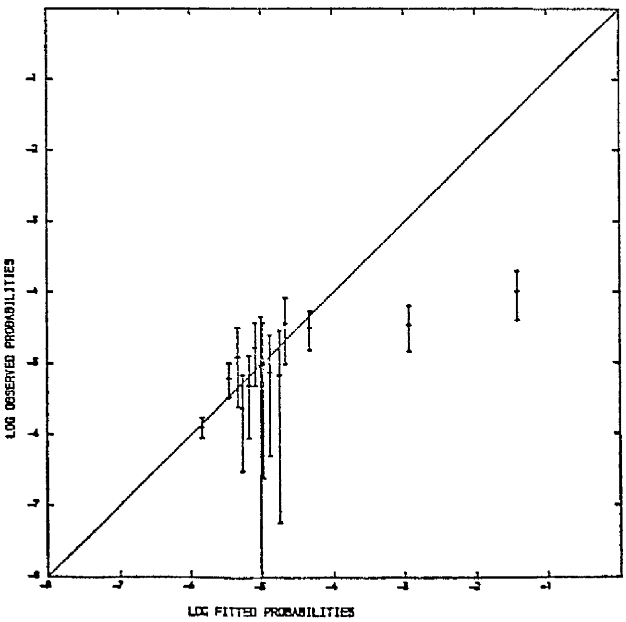
FIGURE 4.2 LOG OBSERVED AGAINST FITTED PROBABILITIES FROM MAIN EFFECTS MODEL FOR BIRTHWEIGHT: MULTIPARAE



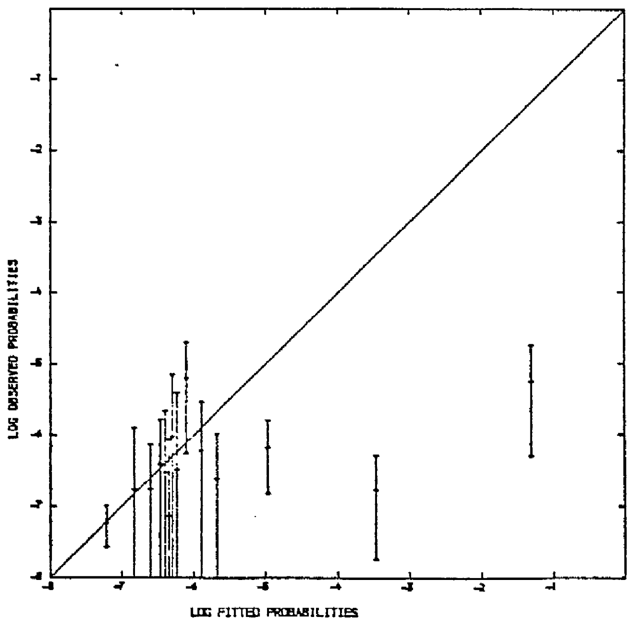
a) Birthweight < 2500gms



b) Birthweight < 2000gms



c) Birthweight < 1500gms



d) Birthweight <1000 gms

RANGE OF FITTED PROBABILITIES, NUMBER OF CASES AND OBSERVED PROBABILITIES IN OBSERVED vs FITTED BIRTHWEIGHT PLOTS

<1,000 gms					<1,500 gms					< 2,000 gms					< 2,500 gms				
Range of fitted probs.	Primiparae		Multiparae		Range of fitted probs.	Primiparae		Multiparae		Range of fitted probs.	Primiparae		Multiparae		Range of fitted probs.	Primiparae		Multiparae	
	Freq.	Obs. Rate	Freq.	Obs. Rate		Freq.	Obs. Rate	Freq.	Obs. Rate		Freq.	Obs. Rate	Freq.	Obs. Rate		Freq.	Obs. Rate	Freq.	Obs. Rate
.0-.001	6323	.0008	64087	.0007	.0-.004	3643	.0041	68174	.0028	.0-.01	2258	.0084	48760	.0077	.0-.03	1868	.0268	26051	.0223
.001-.00125	7054	.0013	1736	.0012	.004-.00475	6016	.0050	12302	.0055	.01-.012	5942	.0123	19656	.0097	.03-.04	7575	.0378	24029	.0350
.00125-.0015	9476	.0017	4324	.0012	.00475-.005	3733	.0046	2021	.0074	.012-.013	3457	.0116	10943	.0148	.04-.05	11585	.0438	26414	.0447
.0015-.0016	7684	.0012	3053	.0016	.005-.0055	6447	.0051	3040	.0036	.013-.014	6807	.0132	4401	.0130	.05-.06	14959	.0539	12645	.0559
.0016-.0017	4168	.0022	1372	.0015	.0055-.006	3732	.0048	2829	.0049	.014-.015	8922	.0128	504	.0258	.06-.07	18819	.0651	4651	.0694
.0017-.0018	5102	.0016	1265	.0008	.006-.0065	5315	.0066	2601	.0085	.015-.016	6255	.0133	3964	.0144	.07-.08	12871	.0729	3447	.0806
.0018-.0019	10894	.0019	824	.0024	.0065-.00675	11849	.0066	604	.0066	.016-.017	3583	.0176	2513	.0187	.08-.09	5208	.0831	2572	.0844
.0019-.002	2954	.0017	658	.0015	.00675-.007	6209	.0076	900	.0067	.017-.018	7043	.0179	654	.0229	.09-.10	2200	.1009	1674	.0998
.002-.0025	9961	.0020	1637	.0055	.007-.008	6371	.0082	1340	.0060	.018-.02	12139	.0195	2695	.0208	.10-.12	2411	.1008	2487	.0965
.0025-.003	5576	.0023	2008	.0020	.008-.009	9665	.0076	875	.0057	.02-.0225	6172	.0241	1931	.0191	.12-.15	1777	.1379	1095	.1342
.003-.004	6455	.0029	3719	.0013	.009-.01	4546	.0103	1688	.0118	.0225-.025	5179	.0234	1859	.0269	.15-.20	832	.1695	668	.1677
.004-.01	1215	.0058	8184	.0021	.01-.02	11050	.0121	4691	.0111	.025-.03	5389	.0293	2480	.0242	.20-.30	113	.2124	267	.2097
.01-.10	3060	.0046	8792	.0011	.02-.10	1105	.0190	3463	.0116	.03-.05	5929	.0334	3326	.0343	.30-.70	29	.1034	248	.0766
.10-1.0	351	.0085	1703	.0053	.10-1.00	595	.0252	1834	.0185	.05-1.0	1201	.0533	2676	.0426	.70-1.0	29	.1379	114	.1053

TABLE 4.18

probabilities in the analysis of primiparae. The same categorisations were used in the plots for multiparae, and this resulted in the categories representing the lowest risks of birthweight below each cutpoint having higher frequencies. Observed values were plotted against the weighted average of the fitted probabilities in each category. Both axes were plotted on the scale of log probabilities and the same range was used in all four plots.

The plots of observed versus fitted probabilities from the models of birthweight below 2,500 gms for both groups of women showed that observed were close to fitted probabilities for all but the two categories representing highest fitted risk. Even the highest points in the confidence intervals about the observed probabilities in these two categories were considerably lower than the corresponding averaged fitted values. However, these categories accounted for only 0.07 per cent of primiparae and 0.34 per cent of multiparae. The categories covering the large majority of both groups of women showed close agreement between observed and fitted values. The plots from the analysis of birthweight below 2,000 gms, 1,500 gms and 1,000 gms showed a similar pattern. The final categories in which the observed rate lay significantly below the line of equality between observed and fitted values in these plots, represented a larger proportion of the study population. This was particularly the case in the analysis of birthweight below 1,000 gms for multiparae when 21 per cent of the study population was represented by the final four categories in which the observed rate was significantly lower than the fitted rate.

Section 4.2.2 described how several interactions were significant, particularly in the model of birthweight below 2,500 gms. The plots, however, showed the two models of birthweight below 2,500 gms to be the best fitting in terms of matching observed and fitted probabilities. To examine whether the inclusion of the single significant interaction (that between maternal height and previous perinatal death) in the model for birthweight below 1,000 gms for multiparae, the least adequately fitting model revealed by the plots, would substantially improve the fit of the main effects model, observed probabilities were plotted against fitted with the interaction included (Figure 4.3). The first of the four confidence intervals below the line of equality was slightly higher when the model included the interaction, just reaching the line of equality. Overall, the fit of the main effects model was not greatly changed.

The plots of observed against fitted values showed that the main effects logistic regression models tended to over-estimate the probability of low birthweight amongst categories of infants with high fitted probabilities. This was particularly the case in the models of birthweight below the lower cutpoints, which were based on fewest numbers of births. Inclusion of an interaction in the model for birthweight below 1,000 gms for multiparae, suggests that the interactions would not greatly improve the prediction of high risk infants.

In section 3.1.5 a method of testing the adequacy of the logistic link function within a two parameter family of functions was described. This involved calculating two extra variables which were functions of the mean vector from the main effects model, and testing for their significance in the presence of the main effects. The results of these tests are presented in Table

FIGURE 4.3

LOG OBSERVED AGAINST FITTED PROBABILITIES
FOR MAIN EFFECTS + ·MATERNAL HEIGHT x
PERINATAL DEATH INTERACTION:
MULTIPARAE < 1,000 gms

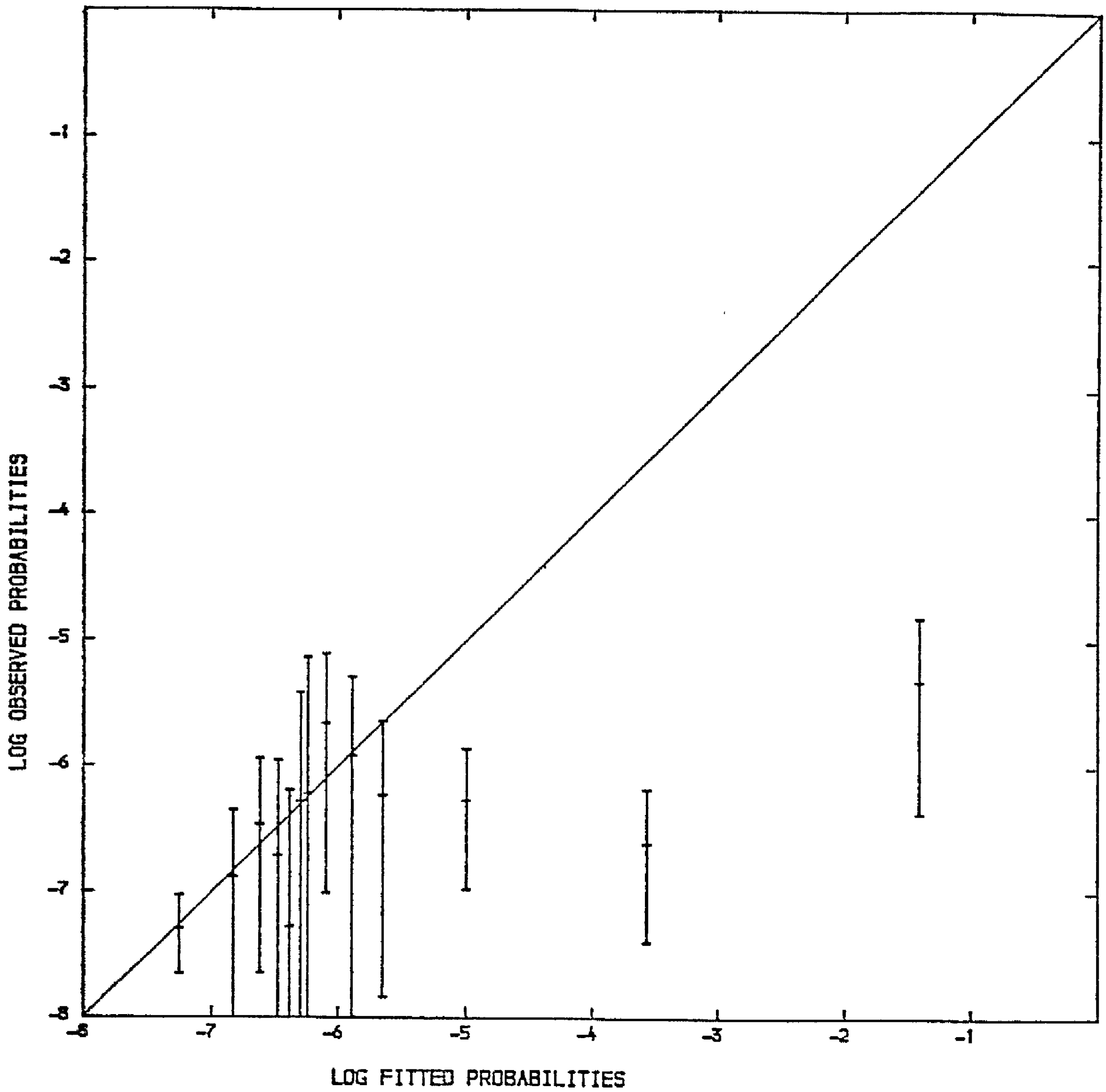


TABLE 4.19

LIKELIHOOD RATIO STATISTICS FOR LOCAL CHANGES IN THE LINK FUNCTION

Birthweight	Primiparae	Multiparae
$<$ 2500 gms	.677 (2 d.f.)	1.798 (2 d.f.)
$<$ 2000 gms	.155 (1 d.f.)	0.000 (1 d.f.)
$<$ 1500 gms	1.867 (1 d.f.)	.401 (1 d.f.)
$<$ 1000 gms	.369 (1 d.f.)	.128 (1 d.f.)

4.19. In the case of birthweight below 2,500 gms the two estimated parameters were very similar in magnitude, and in the remaining analyses the parameters were aliased. In no case did their inclusion result in a significant improvement in the likelihood. This suggests that no function in the family would do substantially better in modelling the data than the logistic, however, if the logistic model is not sufficiently close to the true process that generated the data, the method cannot be expected to indicate any possible improvements to the model, or the lack of them. The tests give only qualified support to the use of the logistic regression model for this data.

4.3 Proportional Hazards Model for Gestational Age

4.3.1 Introduction

A strong association between several of the covariates and birthweight was demonstrated in the previous sections. The analysis now turns to gestational age, which may explain the association with birthweight. If a covariate is associated with low birthweight but not shortened gestation, its influence must lie through birthweight standardised for gestational age, the subject of the analyses of section 4.5.

The relationship between the covariates and gestational age was investigated within the framework of a proportional hazards model. The unborn foetus within a subgroup of the population is assumed to experience a hazard, or instantaneous risk of delivery increased or decreased by a constant multiplicative factor throughout the study period of 28 to 36 weeks of gestation relative to a reference group. An initial examination of the data showed that, while differences in the observed weekly hazard across the categories of most covariates persisted up to the 36th

week of gestation, the hazards of birth converged in later weeks and were mostly very similar around term. The proportionality assumption was thus unlikely to be reasonable for the whole period and the analysis was restricted to the preterm period before the 37^{th} week of gestation. Later births were regarded as a single group and the analysis thus had ten time periods, weekly intervals from the 28^{th} week, and births of gestation 37 completed weeks or more. In terms of the notation of section 3.3.3 the time intervals, $\{[c_j, c_{j+1})\}$, are completed weeks of gestation, the first interval, the 28^{th} week, for example, corresponds to days 196-202 with c_1 equal to midnight on day 195 and c_2 midnight on day 202, the 36^{th} week to days 252-258, and the final time period from day 259 to infinity.

4.3.2 Main Effects Proportional Hazards Model

Tables 4.20 and 4.21 present the unadjusted and adjusted relative hazards of delivery from the proportional hazards models for primiparae and multiparae respectively. Likelihood ratio statistics to test the significance of the hazards are given in Table 4.22. First, female infants experienced a lower risk of preterm delivery than male infants, the adjusted relative hazards for female infants being 0.89 in the analysis of primiparae and 0.85 amongst multiparae. This finding contrasts with the result of increased risk of birthweight below 2,500 gms experienced by female infants.

Single marital status was associated with increased hazards of preterm delivery which were more extreme for multiparae (65 per cent increase) than primiparae (45 per cent increase). Infants in social class I-II experienced a relative hazard of preterm delivery that was decreased by 16 and 25 per cent

TABLE 4.20

PRIMIPARAE: ESTIMATED ADJUSTED AND UNADJUSTED RELATIVE HAZARDS OF DELIVERY (< 37 WEEKS)

95% CONFIDENCE INTERVALS IN BRACKETS

Covariate	Unadjusted Relative Hazard	Adjusted Relative Hazard
Sex		
Male	1.00	1.00
Female	0.88 (0.79, 0.97)	0.89 (0.80, 0.98)
Marital Status		
Married	1.00	1.00
Single	1.62 (1.43, 1.84)	1.45 (1.26, 1.67)
Social Class		
I-II	0.83 (0.71, 0.98)	0.84 (0.71, 0.99)
III	1.00	1.00
IV-V	1.14 (0.99, 1.31)	1.10 (0.96, 1.27)
Unknown	1.31 (1.15, 1.49)	1.14 (1.00, 1.31)
Mother's Height		
<150 cm	1.52 (1.21, 1.93)	1.46 (1.15, 1.84)
150-164 cm	1.00	1.00
≥165 cm	0.73 (0.64, 0.83)	0.75 (0.66, 0.86)
Mother's Age		
<18 yrs	1.76 (1.47, 2.11)	1.47 (1.21, 1.77)
18-24 yrs	1.00	1.00
25-34 yrs	0.99 (0.89, 1.11)	1.13 (1.00, 1.27)
≥35 yrs	1.60 (1.15, 2.22)	1.81 (1.30, 2.53)
Spontaneous Abortion		
0	1.00	1.00
1	1.16 (0.97, 1.39)	1.21 (1.01, 1.46)
2+	1.48 (1.02, 2.16)	1.54 (1.05, 2.24)
Therapeutic Abortion		
0	1.00	1.00
1+	1.28 (1.05, 1.55)	1.28 (1.06, 1.56)

TABLE 4.21

MULTIPARAE: ESTIMATED ADJUSTED AND UNADJUSTED RELATIVE HAZARDS OF DELIVERY (37 WEEKS)

95% CONFIDENCE INTERVALS IN BRACKETS

Covariate	Unadjusted Relative Hazard	Adjusted Relative Hazard
Sex		
Male	1.00	1.00
Female	0.86 (0.78, 0.95)	0.85 (0.77, 0.94)
Marital Status		
Married	1.00	1.00
Single	1.98 (1.56, 2.52)	1.65 (1.29, 2.13)
Social Class		
I-II	0.72 (0.62, 0.85)	0.75 (0.64, 0.88)
III	1.00	1.00
IV-V	1.15 (1.01, 1.30)	1.09 (0.96, 1.24)
Unknown	1.30 (1.15, 1.48)	1.18 (1.03, 1.34)
Mother's Height		
< 150 cm	1.21 (0.95, 1.52)	1.08 (0.85, 1.36)
150-164 cm	1.00	1.00
≥ 165 cm	0.84 (0.75, 0.95)	0.91 (0.80, 1.02)
Mother's Age		
< 18 yrs	3.88 (2.08, 7.25)	3.61 (1.91, 6.79)
18-24 yrs	1.00	1.00
25-34 yrs	0.81 (0.73, 0.90)	0.83 (0.74, 0.93)
≥ 35 yrs	1.31 (1.11, 1.55)	1.20 (1.00, 1.44)
Spontaneous Abortion		
0	1.00	1.00
1	1.29 (1.14, 1.47)	1.26 (1.11, 1.43)
2+	1.95 (1.63, 2.34)	1.81 (1.50, 2.17)
Therapeutic Abortion		
0	1.00	1.00
1+	1.63 (1.38, 1.93)	1.59 (1.34, 1.89)
Previous Caesarean Section		
0	1.00	1.00
1+	1.24 (1.06, 1.45)	1.19 (1.02, 1.39)
Previous Peri-natal Death		
0	1.00	1.00
1+	2.57 (2.21, 3.01)	2.36 (2.02, 2.76)
Previous Live Birth		
0-2	1.00	1.00
3+	1.40 (1.23, 1.60)	1.26 (1.09, 1.45)

TABLE 4.22

LIKELIHOOD RATIO STATISTICS OF UNADJUSTED HAZARDS, ADJUSTED HAZARDS, LINEAR TIME-DEPENDENCE AND FULL TIME-DEPENDENCE FOR EACH CO-VARIATE

Covariate	Primiparae					Multiparae			
	Degrees of freedom	Unadjusted hazard	Adjusted hazard	Linear time dependence	Full * time-dependence	Unadjusted hazard	Adjusted hazard	Linear time-dependence	Full * time-dependence
Sex of Infant	1	5.78	5.15	1.88	6.45 (8)	9.00	10.02	0.03	4.13 (8)
Marital Status	1	50.46	24.77	0.01	11.43 (8)	26.04	13.62	0.46	5.81 (8)
Social Class	3	32.46	10.11	3.08	30.59 (24)	53.39	27.80	1.19	18.24 (24)
Maternal Height	2	38.54	30.88	1.14	11.45 (16)	11.27	3.21	1.56	17.17 (16)
Maternal Age	3	41.05	25.12	2.17	35.70 (21)	53.44	37.63	0.58	15.00 (22)
Previous Spontaneous Abortion	2	6.00	7.94	4.28	15.13 (15)	53.27	41.06	4.40	18.85 (16)
Previous Induced Abortion	1	5.75	5.91	0.03	6.70 (8)	28.42	25.57	0.25	5.61 (8)
Previous Caesarean Section	1	-	-	-	-	6.96	4.59	0.43	11.67 (8)
Previous Per-natal Death	1	-	-	-	-	111.74	92.78	3.70	14.50 (8)
Previous Live Birth	1	-	-	-	-	22.26	9.15	2.16	13.68 (8)

* degrees of freedom in brackets

compared with social class III infants, but although social class IV-V infants experienced increased hazards, the increase was not so great as the corresponding increased risk of birthweight below 2,500 gms. Both confidence intervals for the relative hazard associated with social class IV-V contained the value unity.

The gradient of hazard associated with maternal height amongst primiparae was not as strong as the gradient in risk of low birthweight, with relative hazard of 1.46 for women of height <150 cm, and 0.75 for women of height $\geq$165 cm. This covariate was not responsible for the largest increase in the likelihood of the model for delivery in the preterm period as it was for birthweight below 2,500 gms. Amongst multiparae, even though a slight gradient in hazard of preterm delivery across height categories was observed, maternal height was not significantly associated with preterm delivery. As with low birthweight, both primiparae and multiparae at either end of the age range experienced substantially increased hazard of preterm delivery and the lowest hazard was associated with the most frequently occurring age category. Infants of primiparae aged over 35 experienced increased relative hazard of 1.81, while the infants of very young multiparae, aged under eighteen experienced a relative hazard of 3.61 compared to the 18-24 age category and 4.35 compared to the 25-34 age category.

A history of two or more spontaneous abortions was associated with increased hazard for primiparae (54 per cent increase) and particularly for multiparae (81 per cent increase), while a history of one spontaneous abortion was associated with a hazard increased by over 20 per cent in both groups of women. A previous induced abortion was associated with increased hazard of 28 and 59 per cent for primiparae and multiparae respectively.

The final three aspects of obstetric history for multiparae were all associated with increased hazard. Of the three, a history of perinatal death was the most important, associated with a hazard increased by 136 per cent. In the model for multiparae, a history of perinatal death, followed by spontaneous and induced abortion, were the three factors which had the major impact on the likelihood.

4.3.3 Covariate Interactions in a Binary Logistic Model for Preterm Delivery

Interactions between pairs of covariates were examined in a related model, a binary logistic regression of preterm delivery. Fitting this model was cheaper in computer time and odds ratios from the logistic model for preterm delivery (Table 4.23) were very similar to the estimated hazard of birth during the same period. This similarity is investigated further in section 4.5. By restricting the investigation of interactions to a binary regression, however, it was not possible to test whether the interactions had proportional effects on the hazards.

Table 4.24 gives the likelihood ratio statistics for inclusion of the interactions with the main effects in a logistic regression for preterm delivery. One interaction, that between maternal height and age for primiparae, reached significance at the 5 per cent level. The estimated odds ratios for maternal age within height groups are given in Table 4.25, and show a similar pattern to the corresponding interactions in the model of low birthweight. Infants of primiparae of height <150 cm experienced lowest risk in the under eighteen age category, infants of primiparae of height 150-164 cm experienced lowest risk in the 18-24 age category, while the infants of the tallest primiparae

TABLE 4.23

ADJUSTED ODDS RATIOS OF PRETERM DELIVERY: 1981

Covariate	Primiparae	Multiparae
Sex of Infant		
Male	1.00	1.00
Female	.88 (.79,.98)	.85 (.77,.94)
Marital Status		
Married	1.00	1.00
Single	1.47(1.27,1.70)	1.68(1.29,2.18)
Social Class		
I-II	.83(.70,.99)	.75(.63,.88)
III	1.00	1.00
IV-V	1.11(.96,1.28)	1.09(.96,1.24)
Unknown	1.15(1.00,1.32)	1.18(1.03,1.35)
Maternal Height		
<150 cm	1.48(1.15,1.89)	1.08(.85,1.38)
150-164 cm	1.00	1.00
≥165 cm	.75(.65,.85)	.90(.80,1.02)
Maternal Age		
<18 yrs	1.49(1.22,1.82)	3.77(1.88,7.56)
18-24 yrs	1.00	1.00
25-34 yrs	1.13(1.00,1.28)	.82(.73,.93)
≥35 yrs	1.84(1.30,2.61)	1.21(1.00,1.46)
Previous Spontaneous Abortion		
0	1.00	1.00
1	1.21(1.00,1.47)	1.27(1.11,1.45)
≥2	1.55(1.04,2.29)	1.84(1.52,2.23)
Previous Induced Abortion		
0	1.00	1.00
≥1	1.29(1.06,1.58)	1.62(1.35,1.93)
Previous Caesarean Section		
0	-	1.00
≥1	-	1.20(1.02,1.41)
Previous Perinatal Death		
0	-	1.00
≥1	-	2.43(2.06,2.86)
Previous Livebirth		
0-2	-	1.00
≥3	-	1.27(1.09,1.47)

TABLE 4.24

LIKELIHOOD RATIO STATISTICS FOR INTERACTIONS IN ADJUSTED LOGISTIC REGRESSION OF PRETERM DELIVERY

PRIMIPARAE	Social Class	Maternal Age	Sex of Infant	Maternal Height	Previous Induced Abortion	Previous Spontaneous Abortion
Marital Status	2.68 (3)	3.01 (3)	1.27 (1)	.98 (2)	3.10 (1)	1.10 (2)
Previous Spontaneous Abortion	3.80 (6)	9.41 (6)	5.30 (2)	3.84 (4)	3.03 (2)	
Previous Induced Abortion	1.16 (3)	1.96 (3)	2.59 (1)	.59 (2)		
Maternal Height	2.91 (6)	14.41* (6)	1.74 (2)			
Sex of Infant	4.14 (3)	1.66 (3)				
Maternal Age	5.38 (9)					

MULTIPARAE	Social Class	Maternal Age	Sex of Infant	Maternal Height	Previous Livebirth	Perinatal Death	Previous Caesarean Section	Previous Induced Abortion	Previous Spontaneous Abortion
Marital Status	3.76 (3)	2.43 (3)	.17 (1)	3.64 (2)	1.21 (1)	.52 (1)	.19 (1)	.71 (1)	.98 (2)
Previous Spontaneous Abortion	6.31 (6)	2.00 (5)	9.87* (2)	1.53 (4)	1.70 (2)	3.61 (2)	4.71 (2)	.57 (2)	
Previous Induced Abortion	1.01 (3)	.06 (2)	.00 (1)	.54 (2)	.00 (1)	.09 (1)	.41 (1)		
Previous Caesarean Section	.84 (3)	1.60 (3)	2.16 (1)	2.94 (2)	.23 (1)	.86 (1)			
Previous Perinatal Death	1.36 (3)	2.27 (3)	.30 (1)	.04 (2)	.06 (1)				
Previous Livebirth	1.30 (3)	3.83 (2)	.36 (1)	.09 (2)					
Maternal Height	8.88 (6)	1.32 (6)	6.49* (2)						
Sex of Infant	1.16 (3)	1.88 (3)							
Maternal Age	10.66 (9)								

* significant at the 5 per cent level

TABLE 4.25

INTERACTIONS SIGNIFICANT AT THE 5 PER CENT LEVEL IN LOGISTIC REGRESSION OF PRETERM DELIVERY: 1981

Primiparae a) Maternal Height - Maternal Age

	< 150 cm	150-164 cm	≥ 165 cm
< 18 yrs	.74(.28,1.93)	1.57(1.26,1.95)	1.50(.92,2.44)
18-24 yrs	1.00	1.00	1.00
25-34 yrs	1.10(.63,1.93)	1.21(1.05,1.39)	.90(.69,1.16)
≥ 35 yrs	14.90(3.19,69.54)	1.75(1.14,2.69)	1.45(.72,2.90)

Multiparae a) Previous Spontaneous Abortion - Sex of Infant

	0	1	≥ 2
Male	1.00	1.00	1.00
Female	.93(.83,1.05)	.62(.49,.79)	.71(.49,1.03)

b) Maternal Height - Sex of Infant

	< 150 cm	150-164 cm	≥ 165 cm
Male	1.00	1.00	1.00
Female	.53(.32,.86)	.91(.81,1.02)	.74(.59,.92)

experienced lowest risk in the 25-34 age category. Amongst multiparae, two interactions, those between previous spontaneous abortion, maternal height and sex of infant, were significant at the 5 per cent level. Male infants of multiparous women with a history of spontaneous abortion experienced a greater differential in risk of preterm delivery, than male infants of multiparae without such a history. The changes in risk associated with sex of infant within height categories are not easily interpretable.

4.3.4 Non-Proportional Models

The justification for the proportionality assumption between the 28^{th} and 36^{th} week was examined in several ways. First, logarithms of the empirical weekly hazards were plotted for primiparae and multiparae, Figures 4.4 and 4.5 respectively. The plots for each covariate include weeks for which no births were observed and the value of the logarithm of the weekly hazard was minus infinity, below the lower frame. The plots show that the hazards fluctuated most at lower gestational ages, where there were fewest births and hence greater variability; and for categories that occurred infrequently in the population. The most interesting feature is the difference in the weekly hazard between covariate categories. The proportional hazards model is based on the assumption that these differences are constant on the logarithmic scale, and examination of the plots suggests that this assumption was not unreasonable.

A proportional hazards model for each covariate was compared with two more general models in which time-dependent covariate effects were introduced to provide a test of the validity of the proportionality assumption. In the first model sufficient parameters were included to allow the relative hazards to vary

FIGURE 4.4

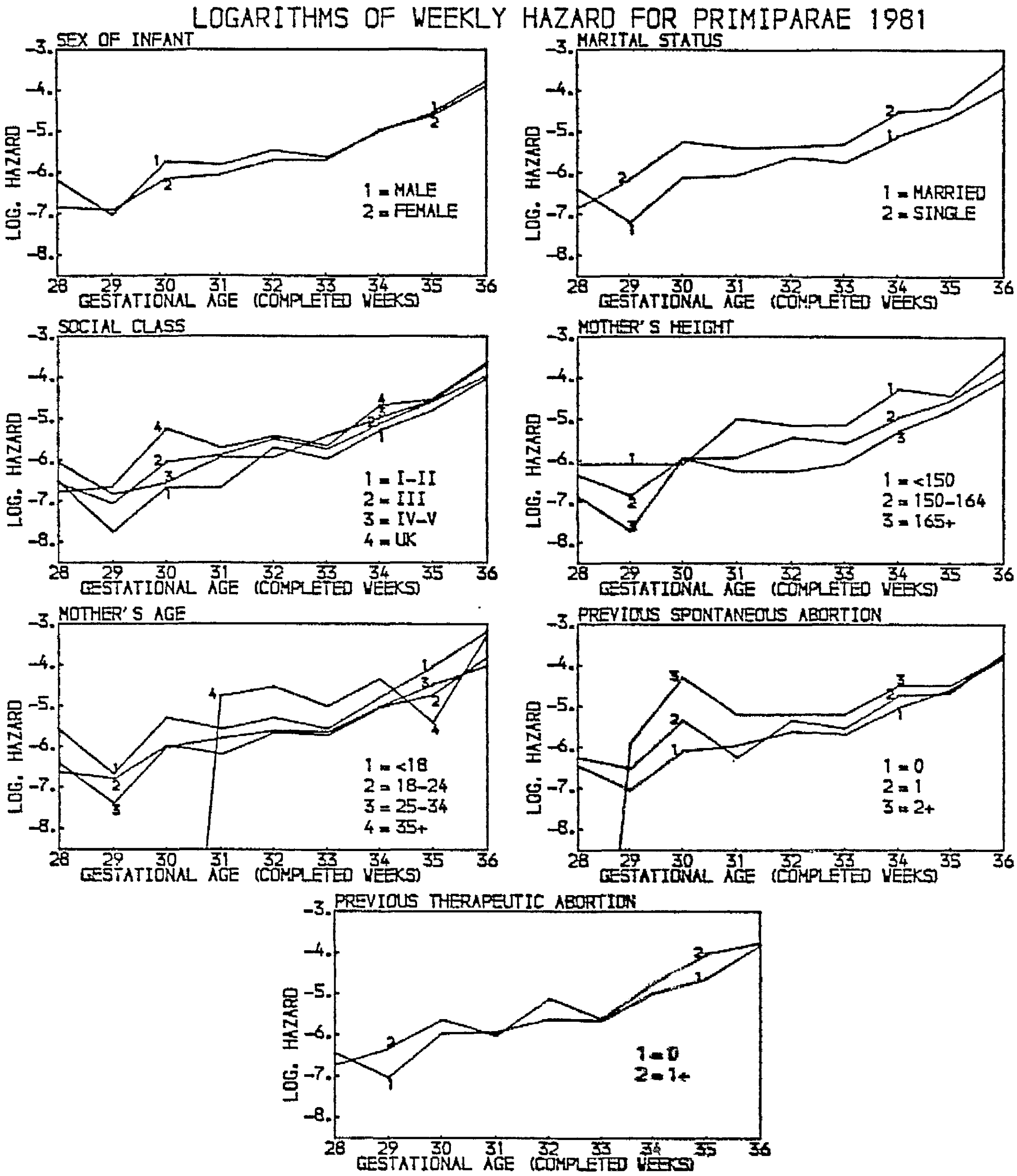
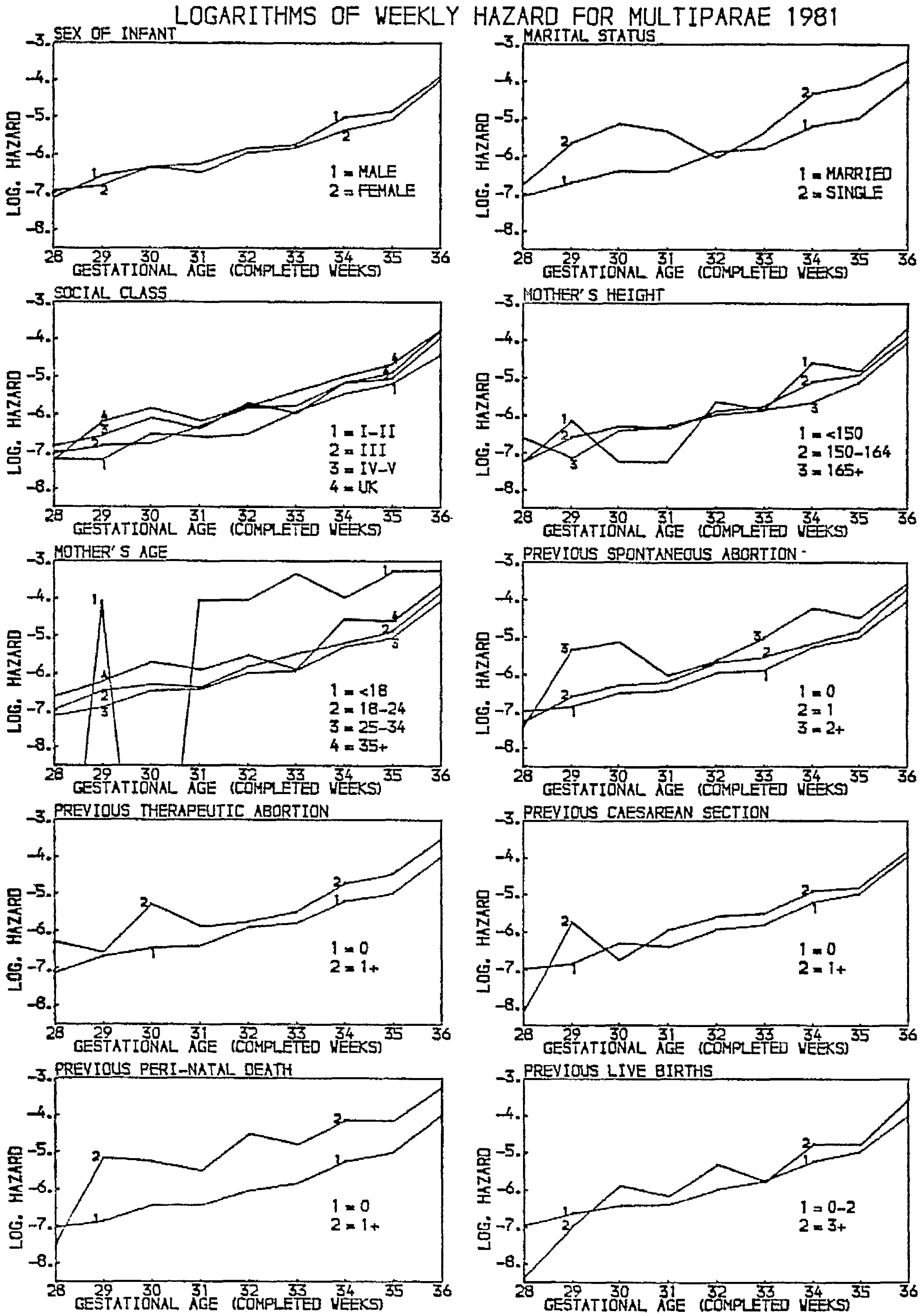

LOGARITHMS OF WEEKLY HAZARD FOR PRIMIPARAE 1981
SEX OF INFANT
1 = MALE
2 = FEMALE
MARITAL STATUS
1 = MARRIED
2 = SINGLE
SOCIAL CLASS
1 = I-II
2 = III
3 = IV-V
4 = UK
MOTHER'S HEIGHT
1 = <150
2 = 150-164
3 = 165+
MOTHER'S AGE
1 = <18
2 = 18-24
3 = 25-34
4 = 35+
PREVIOUS SPONTANEOUS ABORTION
1 = 0
2 = 1
3 = 2+
PREVIOUS THERAPEUTIC ABORTION
1 = 0
2 = 1+
LOG. HAZARD
GESTATIONAL AGE (COMPLETED WEEKS)

FIGURE 4.5


LOGARITHMS OF WEEKLY HAZARD FOR MULTIPARAE 1981
SEX OF INFANT
1 = MALE
2 = FEMALE
MARITAL STATUS
1 = MARRIED
2 = SINGLE
SOCIAL CLASS
1 = I-II
2 = III
3 = IV-V
4 = UK
MOTHER'S HEIGHT
1 = <150
2 = 150-164
3 = 165+
MOTHER'S AGE
1 = <18
2 = 18-24
3 = 25-34
4 = 35+
PREVIOUS SPONTANEOUS ABORTION
1 = 0
2 = 1
3 = 2+
PREVIOUS THERAPEUTIC ABORTION
1 = 0
2 = 1+
PREVIOUS CAESAREAN SECTION
1 = 0
2 = 1+
PREVIOUS PERI-NATAL DEATH
1 = 0
2 = 1+
PREVIOUS LIVE BIRTHS
1 = 0-2
2 = 3+
LOG. HAZARD
GESTATIONAL AGE (COMPLETED WEEKS)

without restriction across time intervals and the hazards were fully dependent on time; in the second model it was assumed that the effect of each covariate on the logarithm of the hazard was a linear function of the time index. In the latter model the relative hazard associated with each covariate increased or decreased with time, providing a plausible model for time trends. Details of the parameterisation of these models is given in section 3.4.4. Both sets of time dependent models were fitted without adjusting for the other covariates. Likelihood ratio tests were performed to assess the significance of the time-dependent terms (Table 4.22). The tests for full time-dependency revealed one significant result, that of maternal age for primiparae. No covariate showed a significant linear time trend.

Table 4.26 gives the estimated parameters from the linear time-dependent models. The intercept parameters measure the difference between the hazard in each category and the reference hazard at 28 weeks, and the slope parameters measure the constant weekly increment in the difference from the reference hazard. The parameters are presented on the hazard scale rather than the log hazard scale, slope parameters less than unity thus represent decreasing relative hazards over time, when greater than unity they represent relative hazards increasing over time. None of the confidence intervals for the slope parameters exclude unity. Some of the intercept parameters are substantially different from the unadjusted relative hazards under the proportionality assumption in Tables 4.20 and 4.21. This can be explained by noting that the estimates of the proportional hazards are based on all weeks and the final weeks are influential in the analysis as they have the highest rates of delivery. The intercept parameter measures the

TABLE 4.26

PARAMETER ESTIMATES FROM LINEAR TIME DEPENDENT MODELS FOR PRETERM DELIVERY: 1981

Covariates	PRIMIPARAE		MULTIPARAE	
	Intercept at 28 weeks	Slope	Intercept at 28 weeks	Slope
<u>Sex of Infant</u>				
Female	.69(.49,.99)	1.03(.99,1.08)	.88(.63,1.25)	1.00(.95,1.04)
<u>Marital Status</u>				
Single	1.65(1.07,2.55)	1.00(.94,1.06)	2.59(1.18,5.67)	.96(.87,1.07)
<u>Social Class</u>				
I-II	.56(.32,1.04)	1.05(.97,1.14)	.90(.52,1.55)	.97(.90,1.04)
IV-V	.87(.53,1.41)	1.04(.97,1.11)	1.15(.73,1.81)	.99(.94,1.06)
Unknown	1.34(.87,2.07)	1.00(.94,1.06)	1.55(.99,2.43)	.98(.92,1.04)
<u>Maternal Height</u>				
$<$150 cm	1.63(.75,3.55)	.99(.89,1.10)	.82(.33,2.04)	1.05(.94,1.18)
$>$165 cm	.58(.36,.92)	1.03(.97,1.10)	.99(.65,1.50)	.98(.93,1.04)
<u>Maternal Age</u>				
$<$18 yrs	1,65(.88,3.11)	1.01(.93,1.10)	10.29(1.89,55.96)	.87(.68,1.11)
25-34 yrs	1.27(.87,1.86)	.97(.92(1.02)	.80(.54,1.18)	1.00(.95,1.05)
$\geq$ 35 yrs	1.42(.44,4.60)	1.02(.87,1.19)	1.53(.86,2.72)	.98(.91,1.06
<u>Previous Spontaneous Abortion</u>				
1	1.58(.88,2.85)	.96(.88,1.04)	1.06(.67,1.70)	1.03(.97,1.09)
$\geq$ 2	3.93(1.43,10.82)	.87(.75,1.00)	3.26(1.84,5.77)	.93(.86,1.01)
<u>Previous Induced Abortion</u>				
$\geq$1	1.35(.70,2.58)	.99(.91,1.08)	1.89(1.06,3.63)	.98(.91,1.06)
<u>Previous Caesarean Section</u>				
$\geq$1	-	-	1.47(.87,2.50)	.98(.91,1.05)
<u>Previous Perinatal Death</u>				
$\geq$1	-	-	4.16(2.53,6.84)	.93(.87,1.00)
<u>Previous Livebirth</u>				
$\geq$ 3	-	-	.98(.59,1.63)	1.05(.98,1.12)

difference at 28 weeks and, after adjusting for the passage of eight time periods, the discrepancy is not large.

Figure 4.6 demonstrates the effect of relaxing the proportionality assumption for the covariate previous spontaneous abortion in the unadjusted model for primiparae. Figure 4.6.a shows the observed weekly hazards and corresponds to the model of full time-dependence in the absence of the other covariates. In Figure 4.6.b the model is restricted to the four parameters measuring linear time-dependence in the relative hazards. Figure 4.6.c shows the estimated weekly hazards under the proportionality assumption. The decreasing relative hazards as gestational age progresses from the 28^{th} to the 36^{th} week in Figure 4.6.b provide a better summary of the observed hazards than the proportional hazards of Figure 4.6.c, but the less restricted model did not result in a significant improvement in the likelihood compared to the proportional hazards model.

4.4 Polytomous Logistic Regression Model for Birthweight Standardised for Gestational Age

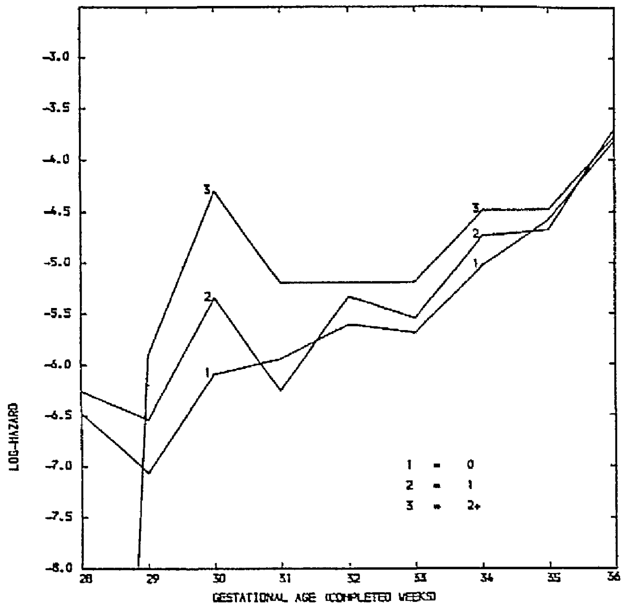
4.4.1 Introduction

Sections 4.2 and 4.3 demonstrated that several covariates were associated with similar patterns of risk of low birthweight and hazard of delivery during the preterm period. The risks of the two perinatal outcomes differed for other covariates, either in direction (sex of infant), or magnitude (maternal height, for example). This section examines the association between the covariates and birthweight standardised for gestational age. If association with birthweight cannot be explained by an increased hazard of preterm delivery we must look to the risk of birthweight standardised for gestational age for an explanation.

FIGURE 4.6

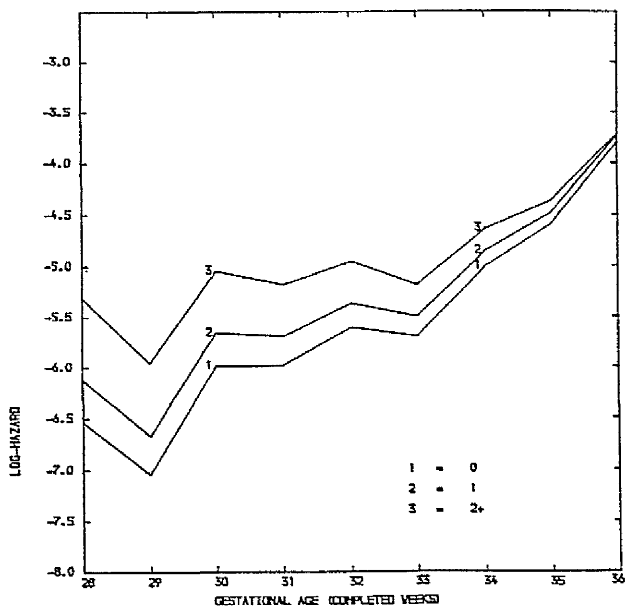
Varying the Proportionality Assumption for Previous Spontaneous Abortion: Primiparae 1981

a) Full-time dependence hazards

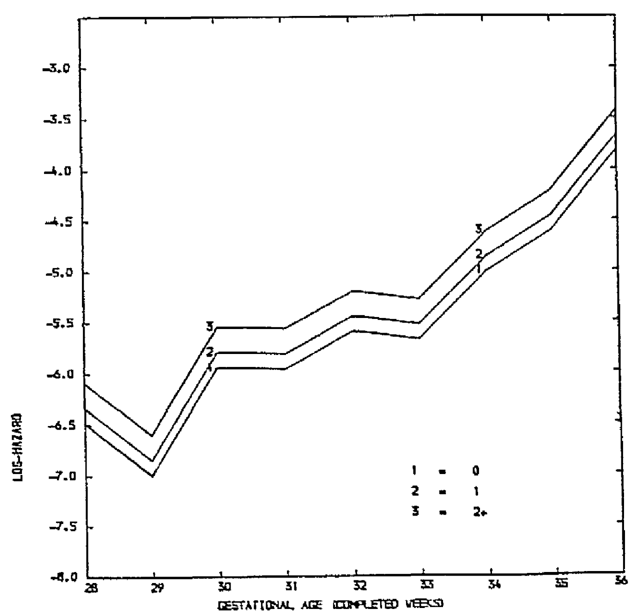


TIME DEPENDENCE

b) Linear time dependence hazards



c) Proportional hazards



The response examined in this section is an ordered polytomous variable with eight categories representing the position of the infant on the birthweight distribution for its gestational age in terms of seven percentiles (the 5^{th}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} and 95^{th}). The models presented in section 4.4 constrained the odds ratios of birthweight below each percentile to be equal. In section 4.4.3 the constraints are lifted to allow the odds ratios to vary over percentiles.

The polytomous models with odds ratios constrained to be equal at each cutpoint on the scale of birthweight standardised for gestational age describe a stochastic ordering of the response variable with respect to the categories of the covariates. Under certain circumstances the constrained model can result from a simple change in location of an underlying continuous variable. In section 3.3.4 by assuming that an underlying continuous variable varied according to a logistic distribution and that the distribution changed between subgroups of the population in terms of location parameter only, it was shown that use of the logistic link function in a generalised model would result in the constrained model having equal odds ratios at each cutpoint. In the following paragraph some evidence for the first assumption that the underlying distribution may be close to a logistic distribution is presented.

In Table 4.27 the skewness and kurtosis of the birthweight distribution at each gestational age are presented, along with approximate 95 percentiles of the distribution of these statistics assuming they come from a Normal distribution. On the whole the observed values showed positive skewness and slightly higher values for kurtosis than would be expected from a Normal sample of the appropriate size. However, considering the large

TABLE 4.27

SKEWNESS AND KURTOSIS OF BIRTHWEIGHT DISTRIBUTION AT EACH GESTATIONAL AGE*

Gestational Age	PRIMIPARAE					MULTIPARAE				
	Total	Skewness	$1.96\sqrt{\frac{6}{n}}$	Kurtosis	$1.96\sqrt{\frac{24}{n}}$	Total	Skewness	$1.96\sqrt{\frac{6}{n}}$	Kurtosis	$1.96\sqrt{\frac{24}{n}}$
28	127	1.506	0.800	5.253	0.852	102	3.730	0.475	19.960	0.951
29	96	3.648	0.490	17.244	0.980	108	2.239	0.462	12.230	0.924
30	169	1.221	0.369	3.784	0.739	151	2.200	0.391	6.800	0.781
31	157	1.184	0.383	4.042	0.766	183	1.981	0.355	6.564	0.710
32	310	0.551	0.273	1.049	0.545	267	0.769	0.294	0.373	0.588
33	316	0.159	0.270	0.521	0.540	318	0.229	0.269	1.504	0.538
34	614	0.299	0.194	0.873	0.388	621	0.272	0.193	1.414	0.385
35	848	0.235	0.165	0.637	0.330	877	0.244	0.162	0.648	0.324
36	1,982	0.076	0.108	0.731	0.216	2,088	0.371	0.105	1.104	0.210
37	3,322	0.023	0.083	0.830	0.167	4,167	0.158	0.074	0.757	0.149
38	8,345	-0.006	0.053	0.598	0.105	11,890	0.149	0.044	0.775	0.088
39	14,970	0.095	0.039	0.532	0.078	21,901	0.116	0.032	0.479	0.065
40	32,901	0.135	0.026	0.836	0.053	45,273	0.123	0.023	0.489	0.045
41	16,244	0.241	0.038	1.163	0.075	19,105	0.161	0.035	0.611	0.069
42	2,985	0.075	0.088	0.355	0.176	3,074	0.171	0.087	0.396	0.173
43	901	-0.540	0.160	1.631	0.320	1,686	-0.197	0.117	0.732	0.234

* All liveborn singleton deliveries 1980-82

sizes of many of the samples, birthweight was close to Normally distributed within each week of gestation. The observed values of kurtosis being postive was evidence of flatter tails to the underlying distribution. The logistic and Normal distributions are similar in appearance both being symmetrical but the logistic has flatter tails with kurtosis of 1.2. In the description of the results when the constrained model adequately describes the association with a covariate, the association is described as a change in location of the distribution. While this provides an easy interpretation of the model, it is only partly justified by the results of Table 4.27, and the assumed lack of change in scale remains untested.

Some preliminary results in this section were based on an analysis of birthweight standardised for gestational age for term and preterm deliveries combined. Table 4.28 gives the estimated odds ratios from the generalised linear models for primiparae and multiparae in which the odds ratios at each of the seven percentiles were constrained to be equal. The covariate, sex of infant, was not included in this, or any of the following analyses of birthweight standardised for gestational age because sex was controlled in the standardisation procedure. In Table 4.29 results from a similar series of analyses are given, in which each covariate in turn was allowed to have different odds ratios amongst term and preterm infants. The likelihoods of these models were compared with the likelihood of the model including the main effects of the covariates and an additional term describing the status, preterm or otherwise of the infant. Likelihood ratio statistics and their appropriate degrees of freedom are also given in Table 4.29. The constrained odds ratios

TABLE 4.28

ADJUSTED ODDS RATIOS IN THE GENERALISED MODEL, WITH ODDS RATIOS CONSTRAINED TO BE EQUAL AT ALL PERCENTILES: TERM AND PRETERM INFANTS COMBINED

Covariate	Primiparae	Multiparae
Marital Status		
Married	1.00	1.00
Single	1.16(1.11,1.20)	1.24(1.16,1.34)
Social Class		
I-II	.93(.90,.96)	.85(.82,.88)
III	1.00	1.00
IV-V	1.13(1.09,1.17)	1.13(1.10,1.16)
Unknown	1.15(1.11,1.19)	1.19(1.16,1.28)
Maternal Height		
<150 cm	2.12(1.98,2.27)	2.11(2.00,2.24)
150-164 cm	1.00	1.00
≥165 cm	.55(.54,.57)	.57(.55,.58)
Maternal Age		
<18 yrs	.80(.76,.85)	1.01(.80,1.28)
18-24 yrs	1.00	1.00
25-34 yrs	.94(.91,.96)	.82(.80,.84)
≥ 35 yrs	.93(.85,1.02)	.80(.76,.84)
Previous Spontaneous Abortion		
0	1.00	1.00
1	.99(.95,1.04)	1.04(1.01,1.08)
≥ 2	1.08(.97,1.20)	1.22(1.16,1.28)
Previous Induced Abortion		
0	1.00	1.00
≥ 1	1.02(.97,1.08)	1.09(1.04,1.14)
Previous Caesarean Section		
0	-	1.00
≥ 1	-	.86(.82,.90)
Previous Perinatal Death		
0	-	1.00
≥ 1	-	1.24(1.18,1.30)
Previous Livebirth		
0-2	-	1.00
≥ 3	-	.96(.93,.99)

ADJUSTED ODDS RATIOS IN THE GENERALISED MODEL AND THEIR LIKELIHOOD RATIO STATISTICS OF COVARIATE INTERACTIONS WITH PRETERM/OTHER DELIVERY

COVARIATE	PRIMIPARAE			MULTIPARAE		
	Term	Preterm	Likelihood Ratio	Term	Preterm	Likelihood Ratio
Marital Status						
Married	1.00	1.00	40.26	1.00	1.00	10.74
Single	1.19(1.14,1.24)	.77(.68,.88)	(1)	1.28(1.19,1.38)	.84(.66,1.07)	(1)
Social Class						
I-II	.93(.89,.96)	1.06(.90,1.26)	14.37	.84(.82,.87)	1.07(.90,1.26)	21.11
III	1.00	1.00	(3)	1.00	1.00	(3)
IV-V	1.13(1.09,1.17)	1.08(.94,1.24)		1.13(1.10,1.16)	1.12(.98,1.29)	
Unknown	1.16(1.12,1.20)	.97(.85,1.11)		1.20(1.16,1.24)	1.00(.87,1.14)	
Maternal Height						
< 150 cm	2.17(2.01,2.33)	1.59(1.26,2.00)	23.02	2.17(2.05,2.30)	1.38(1.09,1.74)	29.57
150-164 cm	1.00	1.00	(2)	1.00	1.00	(2)
⩾ 165 cm	.55(.53,.56)	.70(.62,.80)		.56(.55,.58)	.71(.62,.81)	
Maternal Age						
< 18 yrs	.82(.78,.87)	.61(.51,.73)	18.30	.98(.76,1.27)	1.25(.64,2.42)	25.14
18-24 yrs	1.00	1.00	(3)	1.00	1.00	(3)
25-34 yrs	.93(.90,.96)	1.02(.91,1.15)		.81(.79,.84)	1.03(.92,1.15)	
⩾ 35 yrs	.90(.82,1.00)	1.16(.84,1.59)		.78(.75,.82)	1.70(1.00,1.44)	
Previous Spontaneous Abortion						
0	1.00	1.00	7.00	1.00	1.00	6.52
1	.97(.92,1.02)	1.18(.99,1.42)	(2)	1.04(1.01,1.07)	1.13(.99,1.29)	(2)
⩾ 2	1.03(.92,1.15)	1.43(1.01,2.02)		1.19(1.13,1.26)	1.53(1.26,1.86)	
Previous Induced Abortion						
0	1.00	1.00	1.21	1.00	1.00	.09
⩾ 1	1.01(.96,1.06)	1.13(.93,1.36)	(1)	1.08(1.03,1.13)	1.11(.93,1.34)	(1)
Previous Caesarean Section						
0	-	-	-	1.00	1.00	3.94
⩾ 1	-	-	-	.86(.82,.89)	1.03(.87,1.21)	(1)
Previous Perinatal Death						
0	-	-	-	1.00	1.00	0.14
⩾ 1	-	-	-	1.22(1.16,1.29)	1.27(1.07,1.50)	(1)
Previous Livebirth						
0-2	-	-	-	1.00	1.00	0.56
⩾ 3	-	-	-	.96(.93,1.00)	.91(.78,1.05)	(1)

TABLE 4.29

of birthweight below each percentile associated with several of the covariates differed substantially depending on whether the infant was born term or preterm, and the majority of the likelihood ratios statistics for the interactions were significant at the 5 per cent level.

The main analysis of birthweight standardised for gestational age was performed on term infants alone, and a comparison with preterm infants is given in section 4.4.3. The change in percentile values between gestational ages 37 and 42 is only 80 gms (approximately) amongst both primiparae and multiparae, and male and female infants (see Appendix 2). Thus the analyses, although standardised for gestationl age, are likely to be similar to unstandardised results based on the birthweight distribution amongst term infants.

4.4.2 Main Effects of the Covariates, Term Infants

Table 4.30 gives the unadjusted and adjusted odds ratios of each covariate in the polytomous logistic model for term primiparae and multiparae. The odds ratio were constrained to be equal over the seven percentiles in these analyses. In Table 4.31 likelihood ratio statistics are given of the significance of the odds ratio for each covariate. First, single marital status was associated with increased constrained odds ratios (1.18 relative odds for primiparae and 1.27 relative odds for multiparae compared to married women). Marital status was important in terms of explanatory power in the analysis of birthweight standardised for gestational age, particularly for primiparae. The second socio-economic covariate, social class, was associated with a gradient in risk of birthweight below each percentile as measured by the constrained odds ratios. Social class I-II women experienced decreased odds ratios (0.93 for primiparae and 0.85

TABLE 4.30
UNADJUSTED AND ADJUSTED ODDS RATIOS IN THE CONSTRAINED MODEL WITH ODDS RATIOS EQUAL AT ALL PERCENTILES

Covariate	Term Primiparae		Term Multiparae	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Marital Status				
Married	1.00	1.00	1.00	1.00
Single	1.30(1.26,1.35)	1.18(1.14,1.23)	1.63(1.52,1.76)	1.27(1.18,1.37)
Social Class				
I-II	.86(.83,.89)	.93(1.09,.96)	.76(.74,.79)	.85(.82,.87)
III	1.00	1.00	1.00	1.00
IV-V	1.17(1.13,1.21)	1.13(1.09,1.17)	1.18(1.15,1.21)	1.13(1.10,1.16)
Unknown	1.20(1.16,1.24)	1.15(1.11,1.19)	1.27(1.23,1.30)	1.20(1.16,1.23)
Maternal Height				
<150 cm	2.22(2.06,2.38)	2.16(2.01,1.06)	2.21(2.08,2.34)	2.17(2.05,2.30)
150-164 cm	1.00	1.00	1.00	1.00
≥165 cm	.54(.52,.55)	.55(.53,.56)	.54(.53,.56)	.56(.55,.58)
Maternal Age				
<18 yrs	.95(.90,1.01)	.81(.77,.86)	1.07(.87,1.38)	.97(.75,1.26)
18-24 yrs	1.00	1.00	1.00	1.00
25-34 yrs	.83(.81,.85)	.94(.91,.96)	.74(.72,.75)	.82(.80,.84)
≥35 yrs	.80(.72,.87)	.91(.83,1.00)	.70(.67,.73)	.79(.75,.82)
Previous Spontaneous Abortion				
0	1.00	1.00	1.00	1.00
1	.96(.91,1.00)	.98(.93,1.03)	1.02(.99,1.05)	1.04(1.01,1.07)
≥ 2	1.02(.91,1.14)	1.03(.92,1.16)	1.16(1.10,1.22)	1.20(1.13,1.26)
Previous Induced Abortion				
0	1.00	1.00	1.00	1.00
≥ 1	1.01(.95,1.06)	1.01(.96,1.06)	1.11(1.06,1.16)	1.08(1.03,1.13)
Previous Caesarean Section				
0	-	-	1.00	1.00
≥1	-	-	.93(.90,.97)	.86(.82,.89)
Previous Perinatal Death				
0	-	-	1.00	1.00
≥1	-	-	1.26(1.20,1.33)	1.23(1.16,1.29)
Previous Livebirth				
0-2	-	-	1.00	1.00
≥3	-	-	.98(.95,1.02)	.97(.93,1.00)

TABLE 4.31

LIKELIHOOD RATIO STATISTICS FOR ADJUSTED ODDS RATIOS IN CONSTRAINED MODEL AND FOR DIFFERENCE OVER PERCENTILES (degrees of freedom in brackets)+

Covariate	TERM PRIMIPARAE			TERM MULTIPARAE		
	Unadjusted in Generalised Model	Adjusted in Generalised Model	Difference over Percentage Points	Unadjusted in Generalised Model	Adjusted in Generalised Model	Difference over Percentage Points
Marital Status	206.31(1)***	67.88(1)***	9.40(6)	172.16(1)****	39.06(1)***	9.23(6)
Social Class	357.01(3)***	141.83(3)***	66.66(18)***	958.21(3)***	411.01(3)***	90.81(18)***
Maternal Height	2426.52(2)***	2234.24(2)***	58.74(12)***	3177.70(2)***	2817.25(2)***	62.03(12)***
Maternal Age	197.12(3)***	67.44(3)***	19.84(18)	670.64(3)***	256.78(3)***	50.65(18)***
Previous Spontaneous Abortion	3.46(2)	1.25(2)	24.62(12)*	28.87(2)***	44.64(2)***	51.38(12)***
Previous Therapeutic Abortion	0.03(1)	.11(1)	7.59(6)	19.07(1)***	10.99(1)***	10.58(6)
Previous Caesarean Section	-	-	-	12.47(1)***	64.61(1)***	26.37(6)***
Previous Perinatal Death	-	-	-	74.91(1)***	55.94(1)***	16.98(6)**
Previous Livebirth	-	-	-	0.71(1)	3.58(1)	76.49(6)***

* significant at 2.5 per cent level

*** significant at .1 per cent level

** significant at 1 per cent level

+ these values should be compared to the χ^2 distribution on the appropriate degrees of freedom.

for multiparae), while social class IV-V women experienced increased odds ratios (1.13 for both primiparae and multiparae) compared with the reference group of social class III women.

Women of height <150 cm experienced constrained odds ratios of birthweight below each percentile that were doubled, while for women of height $\geq$165 cm, the odds ratios were halved compared with the height category 150-164 cm. The odds ratios were similar for primiparae and multiparae. Maternal height was more important in terms of increasing the likelihood than the other covariates combined, the likelihood ratio tests for the exclusion of maternal height from the main effects model were 2234 for primiparae and 2817 for multiparae on 2 degrees of freedom. The constrained odds ratios associated with maternal age differed substantially from the risks of both low birthweight and the hazard of preterm delivery. Primiparae aged under eighteen experienced an odds ratio that was significantly reduced by 19 per cent, while for multiparae the odds ratio for this group was marginally reduced by 3 per cent compared with women aged 18-24. Amongst primiparae aged over 25 the odds ratios were reduced by approximately 10 per cent and were only marginally significant, while for multiparae the odds ratios were reduced by approximately 20 per cent. These results are surprising. They suggest that for both groups, women at the extremes of the age distribution were more likely to have heavier infants, while women aged 18-24 years were at higher risk of having an SGA infant, or, more generally, an infant with birthweight below each of the percentiles.

Amongst primiparae a history of either spontaneous or induced abortion was unrelated to the constrained risk of birthweight below each percentile; the estimated odds ratios were very close to 1.00 and the improvement in the likelihood due to their inclusion was negligable. Amongst multiparae a history of two or more spontaneous abortions was associated with a 20 per cent increase in the constrained odds ratio, and a history of induced abortion was associated with an 8 per cent increase. As with low birthweight and preterm delivery a history of perinatal death was an important factor associated with a 23 per cent increase in the constrained odds ratio of birthweight below each percentile. A history of one or more caesarean section was associated with a decrease of 14 per cent in the constrained odds ratio and this decrease was greater after adjusting for the other covariates. A history of three or more previous livebirths was not significantly associated with constrained risk.

4.4.3 Unconstrained Odds Ratios; Term Infants

The results of section 4.4.2 were obtained from constrained analyses in which it was assumed that the odds ratios associated with a covariate were the same at each of the seven percentiles. In this section results from a model in which the constraint was relaxed and the odds ratios of each covariate in turn were allowed to vary over percentiles, while the odds ratios of the other covariates were constrained to be equal, are presented. The parameter estimates are given in Tables 4.32 and 4.33 for primiparae and multiparae respectively. In the first columns the constrained odds ratios are given for reference. The constrained odds ratios can be considered as an average of the specific odds ratios at each percentile, and, as would be expected, they tend to be weighted towards the odds ratio at the 50^{th} percentile (see, for example, the odds ratios for single marital status amongst multiparae). In Tables 4.34 and 4.35 results from a

ADJUSTED ODDS RATIOS FOR TERM PRIMIPARAE: FROM CONSTRAINED MODEL AND WITH ODDS RATIOS VARYING OVER PERCENTILES

Covariate	Constrained Generalised Model	5%	10%	25%	50%	75%	90%	95%
Marital Status								
Married	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Single	1.18(1.14,1.23)	1.18(1.07,1.30)	1.19(1.11,1.28)	1.19(1.13,1.25)	1.15(1.10,1.21)	1.21(1.15,1.28)	1.26(1.16,1.36)	1.32(1.18,1.47)
Social Class								
I-II	.93(.89,.96)	.77(.68,.80)	.77(.71,.84)	.85(.80,.89)	.93(.90,.98)	.96(.92,1.01)	1.02(.96,1.09)	1.02(.94,1.18)
III	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
IV-V	1.13(1.09,1.17)	1.25(1.14,1.37)	1.20(1.12,1.28)	1.15(1.10,1.20)	1.12(1.07,1.16)	1.10(1.06,1.16)	1.13(1.05,1.20)	1.10(1.01,1.20)
Unknown	1.15(1.11,1.19)	1.11(1.02,1.22)	1.09(1.03,1.17)	1.11(1.06,1.16)	1.15(1.10,1.19)	1.19(1.14,1.24)	1.22(1.15,1.31)	1.26(1.16,1.38)
Maternal Height								
<150 cm	2.16(2.01,2.32)	2.31(2.02,2.65)	2.20(1.99,2.44)	2.09(1.92,2.27)	2.11(1.93,2.30)	2.32(2.04,2.64)	2.68(2.14,3.36)	3.44(2.38,4.98)
150-164 cm	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
≥165 cm	.55(.53,.56)	.50(.45,.55)	.50(.46,.53)	.52(.50,.55)	.56(.54,.58)	.56(.54,.58)	.52(.49,.54)	.48(.45,.51)
Maternal Age								
<18 yrs	.81(.77,.86)	.78(.67,.91)	.81(.73,.90)	.81(.75,.87)	.82(.77,.88)	.81(.76,.87)	.76(.70,.86)	.81(.70,.93)
18-24 yrs.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
25-34 yrs	.94(.91,.96)	.97(.90,1.05)	.93(.88,1.98)	.91(.88,.95)	.93(.90,.96)	.95(.92,.99)	.96(.91,1.01)	.96(.89,1.03)
≥35 yrs	.91(.83,1.00)	1.07(.82,1.39)	.99(.81,1.20)	.99(.87,1.13)	.92(.83,1.03)	.90(.80,1.01)	.76(.65,.89)	.73(.60,.91)
Prev. Spontaneous Abortion								
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	.98(.93,1.03)	1.05(.92,1.20)	1.01(.92,1.11)	.99(.92,1.05)	.99(.93,1.04)	.97(.91,1.03)	.92(.85,1.00)	.91(.81,1.02)
≥2	1.03(.92,1.16)	1.48(1.13,1.93)	1.22(.99,1.50)	1.20(1.03,1.39)	1.05(.92,1.19)	.93(.81,1.08)	.76(.63,.92)	.70(.55,.90)
Prev. Therapeutic Abortion								
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
≥1	1.01(.92,1.16)	.85(.72,.99)	.98(.88,1.09)	.99(.92,1.06)	1.01,(.95,1.07)	1.04(.97,1.11)	1.02(.92,1.13)	1.03(.88,1.15)

TABLE 4.32

ADJUSTED ODDS RATIOS FOR TERM MULTIPARAE FROM CONSTRAINED MODEL AND WITH ODDS RATIOS VARYING OVER PERCENTILES

Covariate	Constrained Generalised Model	5%	10%	25%	50%	75%	90%	95%
Marital Status								
Married	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Single	1.27(1.18,1.37)	1.45(1.23,1.71)	1.42(1.26,1.60)	1.26(1.15,1.39)	1.22(1.11,1.32)	1.32(1.18,1.47)	1.32(1.12,1.55)	1.30(1.09,1.75)
Social Class								
I-II	.85(.82,.87)	.64(.57,.71)	.70(.65,.75)	.79(.75,.82)	.85(.82,.88)	.87(.83,.90)	.91(.86,.96)	.87(.82,.94)
III	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
IV-V	1.13(1.10,1.16)	1.19(1.10,1.29)	1.19(1.12,1.26)	1.16(1.11,1.20)	1.13(1.09,1.17)	1.09(1.05,1.13)	1.17(1.06,1.18)	1.10(1.02,1.18)
Unknown	1.20(1.16,1.23)	1.31(1.21,1.42)	1.28(1.21,1.36)	1.21(1.16,1.26)	1.19(1.15,1.23)	1.18(1.13,1.23)	1.19(1.12,1.26)	1.14(1.06,1.24)
Maternal Height								
<150 cm	2.17(2.05,2.30)	2.38(2.13,2.66)	2.27(2.09,2.47)	2.14(2.00,2.29)	2.09(1.94,2.34)	2.21(2.00,2.45)	2.44(2.07,2.87)	2.41(1.91,3.04)
150-164 cm	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
≥165 cm	.56(.55,.58)	.44(.40,.49)	.51(.48,.54)	.54(.52,.56)	.57(.56,.59)	.58(.56,.60)	.54(.52,.56)	.52(.49,.55)
Maternal Age								
<18 yrs	.97(.75,1.26)	.83(.41,1.67)	1.20(.78,1.85)	1.01(.73,1.40)	.99(.74,1.33)	.89(.63,1.26)	.79(.48,1.28)	.68(.36,1.28)
18-24 yrs	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
25-34 yrs	.82(.80,.84)	.84(.79,.90)	.83(79,.87)	.82(79,.85)	.81(.79,.84)	.82(.79,.84)	.79(.76,.83)	.76(.71,.81)
≥35 yrs	.79(.75,.82)	.98(.87,1.11)	.92(.84,1.01)	.85(.80,.91)	.80(.76,.84)	.74(.70,.78)	.69(.64,.74)	.62(.56,.68)
Prev. Spontaneous Abortion								
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1.04(1.01,1.07)	1.15(1.05,1.25)	1.08(1.02,1.15)	1.09(1.04,1.13)	1.04(1.01,1.08)	1.01(.97,1.05)	.96(.91,1.01)	.98(.91,1.05)
≥2	1.20(1.13,1.26)	1.66(1.46,1.88)	1.48(1.34,1.63)	1.28(1.19,1.38)	1.15(1.09,1.23)	1.12(1.03,1.21)	1.11(1.00,1.22)	1.09(.95,1.24)
Prev. Therapuetic Abortion								
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
≥1	1.08(1.03,1.13)	1.02(.90,1.16)	1.06(.97,1.16)	1.06(1.00,1.13)	1.11(1.05,1.17)	1.09(1.03,1.16)	1.01(.92,1.09)	1.08(.96,1.22)
Prev. Caesarean Section								
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
≥1	.86(.82,.89)	1.00(.90,1.11)	.94(.87,1.01)	.90(.86,.95)	.87(.83,.91)	.82(.78,.86)	.78(.73,.84)	.73(.67,.80)
Prev. Perinatal Death								
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
≥1	1.23(1.16,1.29)	1.55(1.36,1.75)	1.36(1.24,1.49)	1.22(1.14,1.31)	1.22(1.15,1.30)	1.20(1.11,1.28)	1.16(1.05,1.29)	1.11(.07,1.28)
Prev. Livebirth								
0-2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
≥3	.97(.93,1.00)	1.19(1.09,1.31)	1.13(1.06,1.21)	1.07(1.02,1.19)	.99(.95,1.03)	.88(.84,.92)	.84(.79,.89)	.80(.74,.87)

TABLE 4.33

ADJUSTED ODDS RATIOS FOR TERM PRIMIPARAE: FROM CONSTRAINED MODEL AND FROM BINARY LOGISTIC MODELS AT EACH PERCENTILE

Covariate	Constrained Generalised Model	5%	10%	25%	50%	75%	90%	95%
Marital Status								
Married	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Single	1.18(1.14,1.23)	1.21(1.10,1.34)	1.20(1.12,1.30)	1.20(1.14,1.26)	1.15(1.10,1.21)	1.22(1.15,1.29)	1.27(1.17,1.38)	1.31(1.17,1.47)
Social Class								
I-II	.93(.89,.96)	.75(.67,.84)	.77(.71,.83)	.85(.80,.89)	.93(.89,.97)	.96(.91,1.00)	1.02(.96,1.10)	1.04(.95,1.13)
III	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
IV-V	1.13(1.09,1.17)	1.26(1.15,1.38)	1.20(1.12,1.28)	1.14(1.09,1.20)	1.11(1.07,1.16)	1.10(1.06,1.16)	1.12(1.05,1.20)	1.10(1.00,1.20)
Unknown	1.15(1.11,1.19)	1.12(1.02,1.24)	1.09(1.02,1.17)	1.10(1.05,1.15)	1.15(1.10,1.19)	1.18(1.13,1.24)	1.22(1.14,1.30)	1.24(1.13,1.37)
Maternal Height								
$<$150 cm	2.16(2.01,2.32)	2.29(2.00,2.62)	2.18(1.97,2.42)	2.08(1.91,2.26)	2.11(1.93,2.30)	2.33(2.05,2.65)	2.70(2.16,3.38)	3.46(2.39,5.01)
150-164 cm	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\geqslant$165 cm	.55(.53,.56)	.50(.45,.55)	.50(.47,.54)	.52(.50,.55)	.56(.54,.58)	.56(.54,.58)	.51(.49,.54)	.48(.45,.57)
Maternal Age								
$<$18 yrs	.81(.77,.86)	.77(.65,.90)	.81(.72,.90)	.81(.75,.87)	.83(.78,.89)	.80(.75,.87)	.74(.67,.83)	.76(.65,.88)
18-24 yrs	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
25-34 yrs	.94(.91,.96)	1.02(.94,1.11)	.97(.91,1.02)	.93(.89,.96)	.92(.89,.95)	.95(.91,.98)	.96(.91,1.01)	.97(.90,.93)
$\geqslant$35 yrs	.91(.83,1.00)	1.11(.85,1.45)	1.02(.84,1.24)	.99(.87,1.14)	.91(.81,1.01)	.90(.80,1.01)	.77(.65,.90)	.75(.61,.93)
Prev. Spontaneous Abortion								
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	.98(.93,1.03)	1.05(.92,1.20)	1.01(.92,1.11)	.99(.92,1.06)	.99(.93,1.04)	.97(.91,1.03)	.92(.84,1.00)	.91(.81,1.03)
$\geqslant$2	1.03(.92,1.16)	1.46(1.12,1.91)	1.22(1.00,1.51)	1.21(1.04,1.40)	1.05(.93,1.20)	.93(.81,1.08)	.76(.63,.92)	.70(.54,.89)
Prev. Therapeutic Abortion								
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\geqslant$1	1.01(.96,1.06)	.84 (.72, .98)	.97 (.88,1.08)	.98(.91,1.06)	1.01(.95,1.07)	1.04(.97,1.11)	1.02(.92,1.13)	1.01(.88,1.15)

TABLE 4.34

ADJUSTED ODDS RATIOS FOR TERM MULTIPARAE FROM CONSTRAINED MODEL AND FROM BINARY LOGISTIC MODELS AT EACH PERCENTILE

Covariate	Constrained Generalised Model	5%	10%	25%	50%	75%	90%	95%
Marital Status								
Married	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Single	1.27(1.18,1.37)	1.42(1.20,1.69)	1.39(1.22,1.57)	1.27(1.15,1.39)	1.23(1.12,1.34)	1.33(1.19,1.49)	1.32(1.12,1.56)	1.37(1.08,1.74)
Social Class								
I–II	.85(.82,.87)	.64(.57,.71)	.70(.65,.75)	.78(.75,.82)	.85(.82,.88)	.86(.83,.90)	.92(.87,.97)	.90(.83,.96)
III	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
IV–V	1.13(1.10,1.16)	1.18(1.09,1.28)	1.19(1.12,1.25)	1.15(1.11,1.20)	1.13(1.09,1.17)	1.09(1.05,1.13)	1.11(1.06,1.18)	1.09(1.01,1.18)
Unknown	1.20(1.16,1.23)	1.30(1.20,1.41)	1.27(1.20,1.35)	1.21(1.16,1.26)	1.19(1.15,1.23)	1.18(1.13,1.23)	1.19(1.12,1.26)	1.14(1.05,1.23)
Maternal Height								
<150 cm	2.17(2.05,2.30)	2.29(2.04,2.56)	2.21(2.03,2.41)	2.11(1.97,2.26)	2.08(1.94,2.23)	2.23(2.02,2.47)	2.50(2.12,2.94)	2.49(1.97,3.15)
150–164 cm	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
≥165 cm	.56(.55,.58)	.46(.41,.50)	.52(.49,.55)	.55(.52,.57)	.57(.56,.59)	.57(.55,.59)	.53(.51,.56)	.51(.49,.54)
Maternal Age								
< 18 yrs	.97(.75,1.26)	.78(.38,1.60)	1.16(.75,1.79)	1.00(.73,1.39)	1.00(.75,1.34)	.89(.63,1.27)	.79(.48,1.29)	.68(.36,1.29)
18–24 yrs	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
25–34 yrs	.82(.80,.84)	.87(.81,.93)	.85(.81,.89)	.82(.79,.85)	.81(.78,.83)	.81(.79,.84)	.80(.76,.84)	.76(.71,.81)
≥ 35 yrs	.79(.75,.82)	.96(.85,1.09)	.91(.83,1.00)	.83(.78,.89)	.79(.75,.83)	.75(.71,.80)	.71(.66,.77)	.65(.58,.73)
Prev. Spontaneous Abortion								
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1.04(1.01,1.07)	1.12(1.03,1.22)	1.07(1.00,1.13)	1.08(1.03,1.12)	1.04(1.01,1.08)	1.02(.98,1.06)	.97(.92,1.03)	.98(.93,1.07)
≥ 2	1.20(1.13,1.26)	1.57(1.38,1.79)	1.43(1.30,1.58)	1.26(1.17,1.35)	1.15(1.08,1.22)	1.14(1.07,1.23)	1.14(1.03,1.26)	1.14(.99,1.30)
Prev. Therapeutic Abortion								
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
≥1	1.08(1.03,1.13)	1.01(.88,1.15)	1.05(.96,1.15)	1.06(.99,1.13)	1.11(1.05,1.17)	1.09(1.03,1.16)	1.01(.93,1.10)	1.09(.96,1.22)
Prev. Caesarean Section								
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
≥ 1	.86(.82,.89)	.98(.88,1.15)	.93(.86,1.01)	.91(.86,.95)	.87(.84,.91)	.81(.78,.85)	.77(.73,.83)	.72(.66,.79)
Prev. Perinatal Death								
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
≥ 1	1.23(1.16,1.29)	1.46(1.29,1.66)	1.31(1.19,1.44)	1.20(1.12,1.29)	1.22(1.15,1.30)	1.21(1.13,1.30)	1.19(1.08,1.32)	1.15(1.00,1.32)
Prev. Live Birth								
0–2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
≥ 3	.97(.93,1.00)	1.10(1.00,1.21)	1.08(1.01,1.16)	1.05(1.00,1.10)	.99(.95,1.03)	.89(.85,.93)	.86(.80,.91)	.83(.77,.91)

TABLE 4.35

series of binary logistic regression analyses at each percentile are given. In these tables the odds ratios for all the covariates were allowed to vary over percentiles (in Tables 4.31 and 4.32 only the odds ratios shown varied, the others were constrained). It can be seen that the two sets of results are very similar, generally differing in only the second significant figure. The following description of the results is based on the polytomous models in which only the odds ratio of the covariate being considered were allowed to vary over percentiles.

The likelihood ratio statistic for the variation in odds ratios over percentiles associated with each covariate are given in Table 4.31. The statistics are the difference between the likelihood of the model in which all except the odds ratios for each covariate were constrained to be equal over percentiles with the model in which all odds ratios were constrained. In the case of maternal height, marital status, and to a lesser extent, social class, maternal age, previous perinatal death and caesarean section, the improvement in the likelihood due to including constrained odds ratios in the model was much greater than the improvement due to allowing the odds ratios to vary over percentiles. For these covariates the difference over percentiles may be significant, but the constrained odds ratios provide a useful description of the association. The changes in the birthweight distribution at each gestational age have been summarised by considering first, covariates associated with changes in the location of the distribution, secondly, covariates associated with increased risks of an SGA or LGA infant, and finally, covariates that do not fall in any of these categories.

Some covariates were mainly associated with changes in the location of birthweight standardised for gestational age. Foremost in this group was maternal height. The odds ratios of birthweight below each percentile for both primiparae and multiparae of height <150 cm were doubled, representing a considerable shift to lower birthweight compared to women of height 150-164 cm. Conversely, a substantial increase in the distribution was associated with taller women over 165 cm. A similar pattern of lesser magnitude was observed for marital status, with single women having lower birthweight infants than married women. Maternal age was also primarily associated with shifts in the distribution of birthweight. For primiparae the lowest birthweight distribution was observed for women of age 18-24 years, women aged 25-34 years had higher birthweight infants as did women aged under eighteen. The birthweight distribution for multiparae aged 18-24 years was lower than the distribution for women aged over 25 years. Primiparous and, particularly, multiparous women aged over 35 years experienced increased risk of an LGA infant.

Other covariates were primarily associated with increased risk of an SGA infant. The risk of birthweight below the 10^{th} percentile was reduced by 23 per cent for primiparous and 30 per cent for multiparous women in social class I-II, while for social class IV-V this risk was increased by 20 per cent in both groups, compared to social class III. Amongst multiparae with a history of two or more spontaneous abortions there was a 48 per cent increase in birthweight below the 10^{th} percentile. A history of perinatal death was associated with a 36 per cent increase in the risk of birthweight below the 10^{th} percentile. The pattern of odds ratios associated with these covariates were, in some cases, apparent at all percentiles but were more marked at the lower extremes of the birthweight distribution. A history of caesarean section was associated with increased risk of an LGA infant but had no significant association at the 10^{th} and 5^{th} percentiles. The risk of birthweight above the 90^{th} percentile was increased by 28 per cent following caesarean section.

Finally the association of some covariates with birthweight standardised for gestational age fell into none of these categories. Slightly lower birthweights were observed following induced abortion for multiparae but generally the odds ratios were close to 1.0. The only significant association amongst primiparae with a history of induced abortion was a 15 per cent reduction in risk below the 5^{th} percentile. Amongst primiparae with a history of two or more spontaneous abortions there was a 48 per cent increase in the risk of birthweight below the 5^{th} percentile, but also a 43 per cent increase in risk of birthweight above the 95^{th} percentile. Similarly, a history of three or more livebirths was associated with a 20 per cent increase in the risk of birthweight above the 90^{th} percentile and a 13 per cent increase in the risk of birthweight below the 10^{th} percentile. The likelihood ratio statistic for variation in the odds ratios associated with three or more previous livebirths was highly significant (76.49 on 6 degrees of freedom), while the constrained odds ratio was not significant and was close to 1.00.

4.4.4 Covariate Interactions; Term Infants

Interactions between pairs of covariates were investigated in binary logistic regression models of birthweight below the 10^{th}, 50^{th} and 90^{th} percentiles of the distribution of birthweight standardised for gestational age. The likelihood ratio statistics of the interactions that were significant at the 5 per cent level are given in Table 4.36. There were more significant results amongst multiparae, even after taking into account the increased number of pairs (36 for multiparae and 15 for primiparae). This could be due to the fact that there were 34 per cent more births to multiparae than primiparae and hence possible interactions were more likely to reach significance. Fitting polytomous models with interactions was expensive in computer time and was only done in the following example.

One interaction, that between marital status and maternal age was significant at all but one of the percentiles examined. This interaction was investigated in three polytomous models. First, a model was fitted which allowed the odds ratios for single marital status within age groups to vary over percentiles, while the age effects were constrained to be equal over the percentiles. Models in which the age effects also differed over percentiles were not considered as they comprised too many parameters (57 for primiparae and 60 for multiparae). In the second model the interaction was constrained to be equal over percentiles but the age effects were allowed to vary. The final model constrained the interaction and the age effects to be equal over percentiles. The three models were adjusted for the other covariates. The estimated odds ratios from these analyses are given in Table 4.37 and 4.38 for primiparae and multiparae respectively. The likelihood ratio statistics for differences in the interaction over percentiles (32.64 for primiparae and 24.92 for multiparae on 24 degrees of freedom) indicated that there was no significant benefit in freeing the terms to vary over percentiles. On the basis of the odds ratios from the second and third models which were similar to each other, birthweight

TABLE 4.36

LIKELIHOOD RATIO STATISTICS FOR INTERACTIONS AT 10th, 50th AND 90th PERCENTILES

Percentile	Term Primiparae			Term Multiparae		
	Interaction	2 X LLR	d.f.	Interaction	2 X LLR	d.f.
10th	Marital Status X Age	22.61	3	Marital Status X Age	9.33	3
	Marital Status X Height	19.85	2	Spontaneous Abortion X Age	18.81	5
50th	Marital Status X Age	24.40	3	Marital Status X Age	16.53	3
	Spontaneous Abortion X Age	12.74	6	Spontaneous Abortion X Age	16.69	5
	Induced Abortion X Age	9.13	3	Induced Abortion X Age	8.08	3
				Livebirths X Age	17.00	3
				Social Class X Livebirths	20.76	3
				Spontaneous Abortion X Height	13.84	4
				Perinatal Death X Caesarean Section	7.01	1
				Spontaneous Abortion X Caesarean Section	13.24	1
				Spontaneous Abortion X Marital Status	7.15	2
90th	Marital Status X Age	9.15	3	Social Class X Livebirths	11.09	3
				Livebirths X Perinatal Death	4.46	1
				Induced Abortion X Perinatal Death	11.82	1
				Spontaneous Abortion X Caesarean Section	7.39	2

TABLE 4.37

AGE SPECIFIC ODDS RATIOS FOR MARITAL STATUS FROM VARIOUS MODELS FOR TERM PRIMIPARAE

		< 18 yrs	18–24 yrs	25–34 yrs	≥ 35 yrs
Married		1.00	1.00	1.00	1.00
SINGLE*	5th	.88(.71,1.10)	1.21(1.09,1.35)	1.35(1.01,1.81)	2.56(1.01,6.49)
	10th	.92(.78,1.08)	1.19(1.10,1.29)	1.66(1.37,2.02)	2.29(1.11,4.73)
	25th	.96(.85,1.08)	1.18(1.11,1.25)	1.58(1.37,1.83)	2.71(1.56,4.71)
	50th	1.00(.89,1.12)	1.13(1.07,1.19)	1.48(1.29,1.70)	1.99(1.14,3.46)
	75th	1.01(.90,1.14)	1.21(1.13,1.28)	1.51(1.28,1.80)	2.40(1.12,5.17)
	90th	.95(.81(1.10)	1.33(1.21,1.46)	1.50(1.16,1.93)	1.10(.46,2.61)
	95th	1.02(.83,1.25)	1.38(1.21,1.58)	1.53(1.07,2.18	1.10(.34,3.57)
Single**		.98(.89,1.09)	1.17(1.12,1.23)	1.54(1.36,1.73)	2.18(1.35,3.52)
Single***		.98(.89,1.09)	1.17(1.12,1.23)	1.53(1.36,1.72)	2.27(1.41,3.67)

* Estimated odds ratios from model in which marital status effects within age groups were allowed to vary over percentiles but the age effects were constrained to be equal at all percentiles. The likelihood ratio statistic for difference of marital status over percentiles was 32.64 on 24 d.f.

** Estimated odds ratios for marital status within age groups, with marital status constrained, but age effects free to vary over percentiles. The likelihood ratio statistic for the marital status age interaction was 36.84 on 3 d.f.

*** Estimated odds ratios for marital status within age groups and age effects constrained to be equal at all percentiles. The likelihood ratio for the interaction was 37.01 on 3 d.f.

TABLE 4.38

AGE SPECIFIC ODDS RATIOS FOR MARITAL STATUS FROM VARIOUS MODELS FOR TERM MULTIPARAE

		< 18 yrs	18–24 yrs	25–34 yrs	≥ 35 yrs
Married		1.00	1.00	1.00	1.00
SINGLE*	5th	.61(.14,2.59)	1.25(1.02,1.53)	1.99(1.50,2.64)	2.93(1.16,7.43)
	10th	1.06(.45,2.48)	1.24(1.06,1.44)	1.88(1.52,2.32)	2.61(1.25,5.44)
	25th	.75(.38,1.51)	1.12(1.00,1.26)	1.63(1.38,1.93)	2.53(1.40,4.57)
	50th	.63(.34,1.17)	1.12(1.01,1.24)	1.50(1.28,1.76)	1.91(1.04,3.51)
	75th	.68(.34,1.37)	1.27(1.11,1.45)	1.48(1.21,1.81)	1.73(.80,3.76)
	90th	.72(.28,1.90)	1.26(1.03,1.55)	1.58(1.16,2.15)	.82(.34,1.97)
	95th	.44(.15,1.27)	1.23(.93,1.63)	2.10(1.28,3.45)	1.28(.31,5.29)
Single**		.70(.40,1.23)	1.16(1.06,1.27)	1.59(1.39,1.83)	2.05(1.23,3.42)
Single***		.69(.39,1.21)	1.16(1.06,1.27)	1.59(1.39,1.83)	2.15(1.29,3.60)

* Estimated odds ratios from a model in which marital status effects with age groups were allowed to vary over percentiles but the age effects were constrained to be equal at all percentiles. The likelihood ratio statistic for difference of marital status over percentiles was 24.92 on 24 d.f.

** Estimated odds ratios for marital status within age groups, with marital status constrained, but age effect free to vary over percentiles. The likelihood ratio statistic for the marital status age interaction was 21.30 on 3 d.f.

*** Estimated odds ratios for marital status within age groups and age constrained to be equal at all percentiles. The likelihood ratio statistic for the interaction was 22.08 on 3 d.f.

standardised for gestational age amongst single women was greater than amongst married women aged under 18 years. For women aged over 18 years, single marital status was associated with lower birthweight particularly in the two older age groups.

4.4.5 Comparison between Term and Other Infants

Results from a constrained polytomous logistic model fitted to preterm infants are given in Table 4.39. The patterns of association were different from the results for term infants in several instances. For most covariates the constrained odds ratios were not significant due to the small sample size. Amongst primiparae, however, single marital status was significantly associated with increased odds ratio of birthweight below each percentile. This follows a pattern decribed by Wilcox (1981), reviewed in section 1.1.4 here. He suggested that if the causal mechanism underlying association with perinatal outcome resulted in lower birthweights and gestational ages, then it is possible that mean birthweight for preterm infants be increased but decreased for term infants.

4.5 Changing the Link Function in Polytomous Regression Models

4.5.1 Comparison of the Logit and Complementary Log-Log Link Functions in Regression Models for Gestational Age and Birthweight Standardised for Gestational Age Data for Primiparae

This section presents the results of a comparison of the logistic and complementary log-log link functions in the analyses of gestational age and birthweight standardised for gestational age. The comparison was carried out on the data for primiparae only, because fitting polytomous models for multiparae, was substantially more expensive in computer time owing to the extra covariates and cases involved. First, marginal data sets, broken

TABLE 4.39

ADJUSTED ODDS RATIOS FOR PRETERM PRIMIPARAE AND MULTIPARAE FROM CONSTRAINED MODEL

Covariate	Preterm Primiparae	Preterm Multiparae
Marital Status		
Married	1.00	1.00
Single	.83 (.72, .96)	.97 (.75, 1.26)
Social Class		
I-II	1.02 (.86, 1.21)	.99 (.83, 1.17)
III	1.00	1.00
IV-V	1.11 (.96, 1.28)	1.17 (1.02, 1.34)
Unknown	1.10 (.95, 1.26)	1.07 (.93, 1.26)
Maternal Height		
<150 cm	1.69 (1.34, 2.13)	1.40 (1.11, 1.77)
150-164 cm	1.00	1.00
≥165 cm	.68 (.59, .77)	.69 (.60, .79)
Maternal Age		
<18 yrs	.70 (.58, .84)	1.41 (.72, 2.74)
18-24 yrs	1.00	1.00
25-34 yrs	.91 (.80, 1.03)	.96 (.85, 1.08)
≥35 yrs	1.03 (.74, 1.42)	1.12 (.92, 1.36)
Previous Spontaneous Abortion		
0	1.00	1.00
1	1.13 (.94, 1.36)	1.09 (.95, 1.24)
≥2	1.34 (.94, 1.90)	1.45 (1.19, 1.76)
Previous Induced Abortion		
0	1.00	1.00
≥1	1.11 (.91, 1.34)	1.13 (.95, 1.36)
Previous Caesarean Section		
0		1.00
≥1		1.02 (.86, 1.20)
Previous Perinatal Death		
0		1.00
≥1		1.27 (1.07, 1.50)
Previous Livebirth		
0-2		1.00
≥3		.84 (.72, .98)

down only by each covariate in turn, were prepared and unadjusted parameter estimates and corresponding likelihood ratio statistics were calculated from models based on each of the two link functions. In the analysis of the marginal data sets the likelihood could be used as the basis of a goodness-of-fit test for the constrained model under each link function because each category of the response variable was adequately represented within all levels of the covariates. Adjusted parameter estimates and the likelihood ratio statistics for their exclusion from the models including all main effects were also examined.

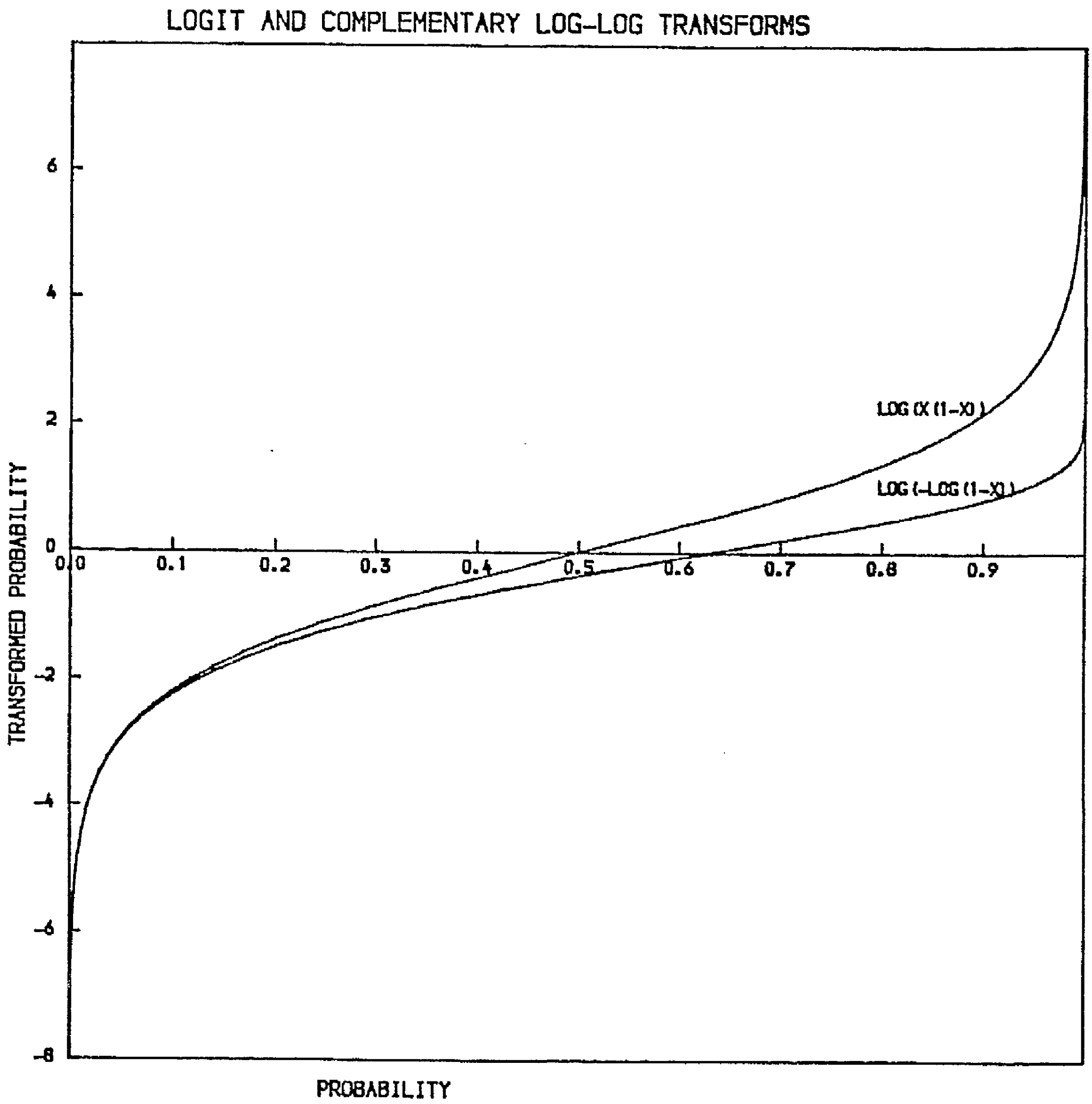
Table 4.40 presents the results of these analyses for the gestational age data for primiparae. Comparing the estimated odds ratios and relative hazards from both the marginal and adjusted data it can be seen that the results were almost identical. There was also very little difference between the sets of likelihood ratio and goodness-of-fit statistics. The deviance from the adjusted model based on the logistic link function at 1786.65 was marginally the lower of the two (the deviance for the proportional hazards model was 1786.91). The proportional hazards model had a natural interpretation in that the regression parameters measured association with the hazard of delivery during the study period, and on these grounds was preferred. The very close similarity between the two models can be explained by noting that for very low probabilities the logistic and complementary log-log transforms are virtually identical (Figure 4.7), it is only for higher probabilities that the two curves diverge, and greater differences might be expected between the models in the analyses of birthweight standardised for gestational age.

GESTATIONAL AGE DATA, PRIMIPARAE 1981 — COMPARISON OF LOGISTIC AND COMPLEMENTARY LOG-LOG LINK FUNCTIONS

COVARIATE	LOGISTIC LINK FUNCTION					COMPARISON LOG-LOG LINK FUNCTION				
	Marginal Data			Cross Classified Data		Marginal Data			Cross Classified Data	
	Odds Ratio	Likelihood Ratio Test	Goodness-of-fit	Adjusted Odds Ratio	Likelihood Ratio Test	Relative Hazards	Likelihood Ratio Test	Goodness-of-fit	Adjusted Relative Hazard	Likelihood Ratio Test
Male	1.00	5.83	6.40	1.00	5.27	1.00	5.77	6.45	1.00	5.15
Female	.88(.79,.98)	(1)	(9)	.88(.79,.98)	(1)	.88(.79,.98)	(1)	(9)	.89(.80,.98)	(1)
Married	1.00	50.42	11.48	1.00	24.86	1.00	50.45	11.44	1.00	24.76
Single	1.65(1.44,1.88)	(1)	(9)	1.47(1.27,1.71)	(1)	1.62(1.43,1.84)	(1)	(9)	1.45(1.26,1.67)	(1)
Social Class										
I-II	.83(.70,.98)	32.55	30.50	.83(.70,.99)	13.11	.83(.71,.98)	32.46	30.59	.84(.71,.99)	13.11
III	1.00	(3)	(27)	1.00	(3)	1.00	(3)	(27)	1.00	(3)
IV-V	1.14(.99,1.31)			1.10(.96,1.28)		1.14(.99,1.31)			1.10(.96,1.27)	
Unknown	1.32(1.15,1.51)			1.15(1.00,1.32)		1.31(1.15,1.49)			1.14(1.00,1.31)	
Maternal Height										
<150 cm	1.55(1.21,1.98)	38.71	11.28	1.48(1.15,1.89)	31.01	1.52(1.20,1.93)	38.53	11.45	1.46(1.15,1.84)	25.12
150-164 cm	1.00	(2)	(18)	1.00	(2)	1.00	(2)	(18)	1.00	(2)
≥ 165 cm	.72(.63,.83)			.75(.65,.85)		.73(.64,.83)			.75(.66,.86)	
Maternal Age										
<18 yrs	1.80(1.49,2.17)	40.71	36.04	1.48(1.22,1.81)	24.89	1.76(1.47,2.11)	41.04	35.70	1.47(1.21,1.77)	30.88
18-24 yrs	1.00	(3)	(27)	1.00	(3)	1.00	(3)	(27)	1.00	(3)
25-34 yrs	.99(.89,1.12)			1.14(1.00,1.28)		.99(.89,1.11)			1.13(1.00,1.27)	
≥ 35 yrs	1.62(1.15,2.28)			1.84(1.30,2.61)		1.60(1.15,2.22)			1.81(1.30,2.53)	
Previous Spontaneous Abortion										
0	1.00	6.18	14.94	1.00	8.07	1.00	6.00	15.12	1.00	7.94
1	1.17(.97,1.41)	(2)	(18)	1.22(1.01,1.48)	(2)	1.16(.97,1.39)	(2)	(18)	1.21(1.01,1.46)	(2)
≥ 2	1.52(1.03,2.23)			1.57(1.06,2.31)		1.48(1.02,2.16)			1.54(1.06,2.34)	
Previous Induced Abortion										
0	1.00	5.80	6.64	1.00	5.97	1.00	5.75	6.69	1.00	5.91
≥ 1	1.29(1.06,1.57)	(1)	(9)	1.29(1.06,1.58)	(1)	1.28(1.05,1.55)	(1)	(9)	1.28(1.06,1.56)	(1)

TABLE 4.40

FIGURE 4.7



LOGIT AND COMPLEMENTARY LOG-LOG TRANSFORMS
6
4
2
0
-2
-4
-6
-8
TRANSFORMED PROBABILITY
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
LOG (X (1-X))
LOG (-LOG (1-X))
PROBABILITY

Table 4.41 gives results of the two analyses of birthweight standardised for gestational age for primiparae. The likelihood ratio statistics for each of the first four covariates alone were substantially greater, and the unadjusted parameters for these covariates were more extreme in the logistic than in the proportional hazards model. The goodness-of-fit statistics in the logistic model were significant at the 5 per cent level for two covariates, social class and maternal height. In the proportional hazards model four covariates had significant goodness-of-fit statistics. After adjusting, marital status, social class, maternal height, maternal age and a history of induced abortion had greater likelihood ratio statistics in the logistic model. The deviance was also substantially lower in the logistic model including all covariates (2483 in the logistic model, 2873 in the proportional hazards model) largely due to just one covariate, maternal height.

4.5.2 Constant Location but Varying Scale Models

The results of section 4.4.3 showed that two covariates, a history of spontaneous abortion amongst primiparae and ≥ 3 previous livebirths amongst multiparae were associated with increases in the risks of birthweight below the lower percentiles and above the upper percentiles. The constrained odds ratios were not significantly different from 1.0, but there were differences in the odds ratios over the percentiles. For these two covariates the assumption that the observed frequencies in each category of birthweight standardised for gestational age could be explained by changes in the location of the distribution of an underlying continuous variable seems unreasonable. The second type of

BIRTHWEIGHT STANDARDISED FOR GESTATION AGE, TERM PRIMIPARAE 1980-82: COMPARISON OF LOGISTIC AND COMPLEMENTARY LOG-LOG LINK FUNCTIONS

COVARIATE	LOGISTIC LINK FUNCTION					COMPARISON LOG-LOG LINK FUNCTION				
	Marginal Data			Cross Classified Data		Marginal Data			Cross Classified Data	
	Odds Ratio	Likelihood Ratio Test	Goodness-of-fit	Adjusted Odds Ratio	Likelihood Ratio Test	Relative Hazards	Likelihood Ratio Test	Goodness-of-fit	Adjusted Relative Hazard	Likelihood Ratio Test
Married	1.00	206.31	9.45	1.00	67.88	1.00	189.43	26.32	1.00	75.00
Single	1.30(1.26,1.35)	(1)	(7)	1.18(1.14,1.23)	(1)	1.16(1.14,1.19)	(1)	(7)	1.11(1.09,1.14)	(1)
Social Class										
I-II	.86(.83,.89)	357.01	65.64	.93(.89,.96)	141.83	.94(.92,.96)	259.31	163.34	.98(.96,1.00)	104.49
III	1.00	(3)	(21)	1.00	(3)	1.00	(3)	(21)	1.00	(3)
IV-V	1.17(1.13,1.21)			1.13(1.09,1.17)		1.08(1.06,1.10)			1.06(1.04.1.08)	
Unknown	1.20(1.16,1.24)			1.15(1.11,1.19)		1.11(1.09,1.14)			1.09(1.07,1.11)	
Maternal Height										
<150 cm	2.22(2.06,2.38)	2426.52	59.34	2.16(2.01,2.32)	2234.24	1.60(1.54,1.67)	2085.00	400.86	1.58(1.51,1.65)	1951.48
150-164 cm	1.00	(2)	(14)	1.00	(2)	1.00	(2)	(14)	1.00	(2)
≥165 cm	.54(.52,.55)			.55(.53,.56)		.72(.71,.73)			.72(.71,.74)	
Maternal Age										
<18 yrs	.95(.90,1.01)	197.12	19.18	.81(.77,.86)	67.44	.97(.94,1.01)	150.62	65.67	.88(.86,.91)	66.95
18-24 yrs	1.00	(3)	(21)	1.00	(5)	1.00	(3)	(21)	1.00	(3)
25-34 yrs	.83(.81,.85)			.94(.91,.96)		.91(.90,.93)			.97(.95,.99)	
≥35 yrs	.80(.72,.87)			.91(.83,1.00)		.86(.81,.91)			.92(.87,.97)	
Previous Spontaneous Abortion										
0	1.00	3.46	25.61	1.00	1.25	1.00	8.17	20.89	1.00	2.82
1	.96(.91,1.00)	(2)	(14)	.98(.93,1.03)	(2)	.97(.94,.99)	(2)	(14)	.98(.95,1.01)	(2)
≥2	1.02(.91,1.14)			1.03(.92,1.16)		.95(.89,1.01)			.97(.90,1.03)	
Previous Induced Abortion										
0	1.00	0.03	8.07	1.00	0.11	1.00	0.19	7.91	1.00	.05
≥1	1.01(.95,1.06)	(1)	(7)	1.01(.97,1.16)	(1)	1.01(.98,1.04)	(1)	(7)	1.00(.97,1.04)	(1)

TABLE 4.41

polytomous regression model suggested in section 3.3.2, in which the distribution of a continuous variable underlying the categories of the response is assumed to vary in scale but not location, was investigated in relation to these two covariates.

Assuming, as before, that an underlying continuous variable has a logistic distribution so that, in the terminology of section 3.3.2,

$$\psi^{-1}(p_j) = \text{logit}(p_j),$$

and if the only difference between two groups is in scale parameters, b_1 and b_2, the following relationship holds between the $\{p_{1j}\}$ and $\{p_{2j}\}$, the proportions below cutpoints in either group

$$b_1/b_2 = \ln\{\text{logit}(p_{2j})/\text{logit}(p_{1j})\}, \quad 1 \leq j \leq k.$$

For values of p_j less than 0.5, logit(p_j) is negative and the right hand side cannot be expressed as the difference of link functions of probabilities as required in the algorithm for fitting the models. The logarithm of the ratio of logits can be computed as long as the two logits are of the same sign. If the log ratio between groups 1 and 2 is a constant, Δ, then

$$p_{2j} = \frac{p_{1j}^{\exp(\Delta)}}{1+p_{1j}^{\exp(\Delta)}}.$$

Tables 4.42 and 4.43 give the cumulative logits, their differences and the logarithm of their ratios, for the covariates previous spontaneous abortion (for primiparae) and previous livebirths. A comparison between women with ≥ 2 previous spontaneous abortions and women with none, showed that there was

TABLE 4.42

COMPARISON OF LOGITS OF BIRTHWEIGHT BELOW EACH PERCENTILE FOR HISTORY OF SPONTANEOUS ABORTION: TERM PRIMIPARAE

	5th	10th	25th	50th	75th	90th	95th
Cumulative Logits							
0	-3.108	-2.315	-1.239	-0.110	1.076	2.181	2.936
1	-3.075	-2.326	-1.274	-0.145	1.020	2.076	2.818
$\geqslant 2$	-2.728	-2.129	-1.074	-0.078	0.990	1.893	2.563
Difference between Logits for J and 0 Spontaneous Abortions							
J = 1	0.033	-0.011	-0.035	-0.035	-0.056	-0.106	-0.118
J = $\geqslant 1$	0.380	0.187	0.165	0.032	-0.087	-0.288	-0.373
Logarithm of Ratio of Logits for J and 0 Spontaneous Abortions							
J = 1	-0.011	0.005	0.028	0.278	-0.053	-0.050	-0.041
J = $\geqslant 1$	-0.131	-0.084	-0.143	-0.345	-0.084	-0.142	-0.136

TABLE 4.43

COMPARISON OF LOGITS OF BIRTHWEIGHT BELOW EACH PERCENTILE FOR PREVIOUS LIVEBIRTHS: TERM MULTIPARAE

	5th	10th	25th	50th	75th	90th	95th
Cumulative Logits							
0–2	-3.187	-2.383	-1.281	-0.093	1.043	2.107	2.850
$\geqslant 3$	-3.001	-2.244	-1.201	-0.089	0.934	1.952	2.655
Difference between Logits for 3 and 0–2 Previous Livebirths							
	0.186	0.139	0.080	0.005	-0.109	-0.154	-0.196
Logarithm of Ratio of Logits for 3 and 0–2 Previous Livebirths							
	-0.060	-0.060	-0.065	-0.050	-0.110	-0.076	-0.071

a monotonic trend in the difference of logits from 0.380 at the 5^{th} percentile, to 0.032 at the 50^{th} percentile, and to -0.373 at the 95^{th} percentile. Looking at the logarithms of the ratio of the logits, we see that all the values were negative, varying between -0.084 and -0.345, with average -0.152. For women with one previous abortion there was also a trend in the difference. The change in sign came at the 10^{th} percentile, and the assumption of no change in location was thus unreasonable. The logarithms of the ratio of the logits in this case showed an even greater range of values than their differences. The differences between the logits for multiparae with 0-2 and ≥ 3 previous livebirths, varied between 0.186 at the 5^{th} percentile and -0.196 at the 95^{th} percentile, while the logarithms of the ratio of the logits were similar, varying between -0.110 and -0.050 with average -0.070.

CHAPTER 5 : NEONATAL CLASSIFICATION BASED ON LATENT CLASS ANALYSIS

5.1 Introduction

A central issue in the allocation of hospital resources amongst regions or areas is the measurement of their differing need for medical services. One approach is to classify individuals into groups with similar medical characteristics and to use the prevalence of these groups as a basis for estimating the resources required by each area. This chapter is concerned with classification, the first stage in this approach, and investigates a measure of case-mix during the neonatal period between birth and discharge from hospital. Many techniques used to classify neonates are based on the categorisation of birthweight into a normal and several low weight groups. The difficulty in employing measures based solely on any one variable is that other aspects of health are ignored: for instance, normal weight infants may have problems or needs which also require specialised treatment. With the help of latent class analysis a measure based on eleven indicators of neonatal health was constructed.

Latent class analysis, introduced by Lazarsfeld in the 1950's (Lazarsfeld & Henry, 1962) is a method of breaking down a multi-dimensional contingency table into constituent subtables, the sums of frequencies of corresponding cells in these subtables equalling those of the original table. If each of the subtables has high frequency in a different region of the table it may be possible to use the assignment of the original cells to a subtable as a classification procedure. The computation required by the procedure, particularly when using many variables, is very heavy and the increasing capabilities of modern computers have greatly facilitated its application. There have been several applications of latent class analysis involving medical data: Miller et al (1962) constructed two classes of mentally retarded patients, one having much higher mortality for all durations of hospitalisation, and concluded that the method had possibilities as an alternative to standardisation in demographic studies: Skene (1978) used latent class analysis to approximate distributions when predicting outcome following severe head injury: Rindskopf & Rindskopf (1986) examined the sensitivity and specificity of various indicators of myocardial infarction within latent classes describing presence or absence of the disease: Kaldor & Clayton (1985) and Clayton (1985) used a similar approach when estimating risk when either the primary factors or confounding variables are measured with error in case-control studies.

5.2 Fitting the Latent Class Model

The data for analysis consisted of measurements of J (equal to 11) categorical variables, X_j, $j=1,\ldots,J$ on each infant; in vector form $\mathbf{X}$. Each X_j, $1 \leq j \leq J$ could take values 1 to C_j, where C_j, was the number of categories of the j^{th} variable. The data were amalgamated into I distinct variable combinations indexed i, $i \leq i \leq I$, with $\mathbf{X}$ taking the same values $\mathbf{x}_i=(x_{i1},\ldots,x_{iJ})$ on each of the w_i infants with combination i. The total number of infants, T, was therefore $\sum_{i=1}^{I} w_i$.

The latent class model is based on a mixture of underlying classes, within each of which the J variables are mutually independent. The number of classes K is assumed known. In practice several different values K are tried. The probabilities that $X_j=1$ within class k are given by parameters θ_{jlk}, $1 \leq j \leq J$,

$1 \leq l \leq C_j$, $1 \leq k \leq K$. The proportion, or prevalence, of class k in the population is given by the parameter λ_k, $1 \leq k \leq K$. If the value of the j^{th} variable, X_j, in the i^{th} combination is denoted by $l(ij)$, $1 \leq l(ij) \leq C_j$, the conditional probabilities of $\mathbf{x}_i$ within classes are given by

$$P(\mathbf{X}=\mathbf{x}_i | \text{class } k) = \prod_{j=1}^{J} \theta_{jl(ij)k},$$

and therefore

$$P(\mathbf{X}=\mathbf{x}_i) = \sum_{k=1}^{K} \lambda_k \prod_{j=1}^{J} \theta_{jl(ij)k}.$$

The likelihood from I distinct combinations, L, is given by

$$L(\theta,\lambda) = \prod_{i=1}^{I} \left[\sum_{k=1}^{K} \lambda_k \prod_{j=1}^{J} \theta_{jl(ij)k} \right]^{w_i},$$

where

$$\sum_{k=1}^{K} \lambda_k = 1, \quad \sum_{l=1}^{C_j} \theta_{jlk} = 1, \text{ for all } 1 \leq j \leq J,\ 1 \leq k \leq K.$$

The values of the parameters θ and λ which maximise the likelihood were found using the EM algorithm which involves two basic steps. First define $\mathbf{z}_i=(z_{i1},\ldots,z_{iK})$

$$z_{ik} = P(\text{class } k | \mathbf{X}=\mathbf{x}_i).$$

The first stage of the algorithm, the M-step, maximises $L(\theta,\lambda)$ for given $\mathbf{z}_i$s by

$$\hat{\theta}_{jlk} = \frac{\left[\sum_{i:x_{ij}=l} z_{ik} w_i + \frac{1}{C_j} \right]}{\sum_{i=1}^{I} z_{ik} w_i + 1},$$

and

$$\hat{\lambda} = \frac{\left[\sum_{i=1}^{I} z_{ik}w_i + \frac{1}{K}\right]}{T+1}.$$

The quantities $1/C_j$ and $1/K$ introduce a small degree of smoothing and ensure that the θs and λs do not become zero during the procedure. If the class membership is known, that is the z_{ik}s have values 0 or 1, the formula for $\hat{\theta}$ reduces to the usual maximum likelihood estimates for independent proportions within classes.

Bayes' theorem gives

$$z_{ik} = P(\text{class } k|\mathbf{X}=\mathbf{x}_i) = \frac{P(\mathbf{X}=\mathbf{x}_i|\text{class } k)P(\text{class } k)}{\sum_{k=1}^{K} P(\mathbf{X}=\mathbf{x}_i|\text{class } k)P(\text{class } k)}$$

and the second stage of the algorithm, the E-step, assumes values θ and λ so that

$$z_{ik} = \frac{\lambda_k \prod_{j=1}^{J} \theta_{jl(ij)k}}{\sum_{k=1}^{K} \lambda_k \prod_{j=1}^{J} \theta_{jl(ij)k}}.$$

The EM algorithm is an iterative process whereby the $\mathbf{z}_i$ are given starting values, $\mathbf{z}_i^0$ $1 \leq i \leq I$, in this case random Uniform(0,1)s scaled so that

$$\sum_{k=1}^{K} z_{ik}^0 = 1.$$

Estimates of θ and λ, θ^0 and λ^0, are obtained from the M-step formula and are substituted in the E-step for z_{ik} to obtain z_{ik}^1s. The process is repeated m times till the difference between successive values of the likelihood is less than a preset tolerance, 0.01 in the analyses here. The values of the parameters which maximise the likelihood in the m^{th}, final step of the iteration, θ^m and λ^m, are taken as the maximimum likelihood estimates of θ and λ. The procedure can converge to local maxima and should be repeated using several sets of starting values for **z**.

5.3 The 1- to 4-Class Models

The parameter estimates of the 1- to 4-class models are presented in Table 5.1 and values of the likelihood ratio statistic between each model and the complete model fitting the data exactly, are given in the first column of Table 5.2 and are labelled $-2\log\Lambda$. The 1-class parameters give the marginal relative frequencies of the variables amongst cases with complete data that were included in the analysis. The value of the $-2\log\Lambda$, 30,545, provides a test for the presence of association among the eleven variables (but most of the cells having low expected frequencies invalidates comparison to a χ^2 percentile). By examining the parameter estimates for the 2-class model it is possible to describe infants belonging to either class. For instance, the probability of an infant weighing >2,500 gms in IIa is 0.99, whereas in IIb an infant has a 0.48 probability of weighing 2,001-2,500 gms and a 0.15 probability of weighing 1,501-2,000 gms. In addition an infant in IIb has probability of 0.05 of death before discharge and generally much higher probabilities of most health problems. The 3-class model consists of IIIa describing healthy infants, and divides class IIb into two. Class IIIc, a class with characteristics associated with moderately low probabilities of assisted resuscitation and a

TABLE 5.1

PARAMETER ESTIMATES FOR THE 1- to 4-CLASS MODELS USING 1980 DATA AND COMPLETE CASES ONLY

Variable	No. of Levels	Levels***	1 Class	2 Classes		3 Classes			4 Classes			
			I	IIa	IIb	IIIa	IIIb	IIIc	IVa	IVb	IVc	IVd
Birthweight	4	2001-2500g	.05	.01	.48	.00	.18	.79	.00	.20	.78	.03
		1501-2000g	.01	.00	.15	.00	.26	.09	.00	.32	.09	.00
		≤1500g	.01	.00	.08	.00	.21	.01	.00	.25	.01	.00
Birthweight for gestation age	2	<10th centile	.10	.07	.43	.07	.19	.62	.07	.21	.62	.10
Apgar at 5 min.	2	<7	.02	.01	.12	.01	.26	.01	.00	.21	.01	.32
Resuscitation	3	Intermediate*	.09	.08	.19	.08	.26	.13	.07	.25	.13	.31
		By intubation	.03	.02	.17	.02	.33	.04	.01	.29	.03	.52
Assisted ventilation after 30 mins.	2	Present	.01	.00	.10	.00	.25	.00	.00	.29	.00	.01
Recurrent Apnoea	2	Present	.01	.00	.06	.00	.17	.00	.00	.20	.00	.00
Jaundice **	2	Present	.30	.28	.58	.28	.67	.49	.28	.71	.49	.32
Convulsions	2	Present	.00	.00	.03	.00	.07	.00	.00	.07	.00	.01
In tube feeding	2	Present	.03	.01	.30	.01	.60	.10	.01	.67	.10	.05
Dead at discharge	2	Present	.00	.00	.05	.00	.13	.00	.00	.15	.00	.00
Age at discharge	3	4-10 days	.80	.84	.34	.83	.09	.53	.83	.04	.50	.84
		>11 days	.08	.03	.61	.03	.79	.45	.03	.82	.46	.14
Frequency of Class			1.00	.92	.08	.92	.03	.05	.89	.03	.05	.04

* Mask + Intermittant Positive Pressure Ventilation, Drugs only, Other.

** >86 μmol/litre bilirubin

*** Parameters for all levels of a variable sum to 1, the 1st level of each variable is omitted without loss of information,

TABLE 5.2 MODEL STATISTICS

MODEL DESCRIPTION	$-2\log\Lambda$	FITTED PARAMETERS	DISTINCT CASES	TOTAL CASES
1980 Data				
Complete Cases only				
1 Class	30545	15	600	45426
2 Classes	11473	31	600	45426
3 Classes	7953	47	600	45426
4 Classes *	6052	63	600	45426
5 Classes *	4915	79	600	45426
6 Classes *	3367	85	600	45426
Complete and Incomplete Cases**				
Missing at Random	***	47	1041	52022
Best possible Values	9297	47	687	52022
Worst Possible Values	13444	47	769	52022
Half Samples**				
Sample 1	4329	47	407	22771
Sample 2	4347	47	434	22655
1978 Data				
3 Classes	7380	47	602	33858
3 Classes (1980 parameters)	8927	0	602	33858
4 Classes	4998	63	602	33858
4 Classes (1980 parameters)	6580	0	602	33858

* The values for 4 class models were obtained after 10 repeat analyses. Repeat analyses were not performed for 5 and 6 classes - these are probably not global maxima.

** Three class models.

*** In the missing at random model $-2\log\Lambda$ was calculated over all cells, a pattern with a missing value was treated as a separate cell contributing to the total value of $-2\log\Lambda$ which was therefore incorrect. This should not affect the E or M stages but only the stopping rule.

low Apgar score, but with birthweight predominantly in the category 2,001-2,500 gms and 62 per cent falling below the 10^{th} percentile of the birthweight distribution standardised for gestational age. Class IIIb has lower prevalence but higher probabilities of death at 0.13, low birthweight and other indicators of poor health than IIb. The 4-class model comprises three classes IVa, IVb and IVc that are similar to IIIa, IIIb and IIIc respectively. The fourth class IVd represents infants that are like IVa in most respects but they tend to require either intermediate resucitation or intubation and have a high probability of an Apgar score <7. These infants seem to be drawn mainly from class IIIa which has a higher prevalence than IVa.

Difficulties arose with multiple maxima at the stage of fitting four classes. It was not practical to carry out a sufficient number of repeat analyses to be reasonably confident of having achieved a global maxima for the 5- and 6-class models. The EM algorithm found two maxima in the likelihood of the 4-class model. The first with $-2\log\Lambda$ at 6,052 was found twice, while the second with a higher value of $-2\log\Lambda$ at 6,432 was found in the remainder of ten repeat analyses. The former of these was taken to be the overall maximum of the likelihood. The introduction of a fourth class in the latter model had the effect of dividing IIIa into two classes, which were differentiated by the categories of only one variable, jaundice. Consequently the method was unable to assign many of the $\mathbf{x}_i$s, including the two representing 63 per cent of the population, decisively to either class. In IVa-IVd, the classes of the former model, the probabilities of jaundice at 0.28, 0.71, 0.49 and 0.32 did not contribute as much to the distinction between each pair of classes as other variables.

The value of $-2\log\Lambda$ has little relevance for choosing the number of classes because many cells have low expected frequencies and the sample size is so large that any parsimonious model would be rejected. Furthermore, the asymptotic theory for the distribution of likelihood ratio statistics breaks down when testing the significance of a K^{th} class in the presence of K-1 (Aitkin et al, 1981). Aitkin et al compare several recently suggested approximations to the distribution of the likelihood ratio between one and two class models and perform simulation tests at the 5 per cent level. This approach was not feasible here because of the size of the data and cost of fitting models.

The choice of number of classes was made on pragmatic grounds. The 4-class model distinguished between the healthy infant and three additional classes, two representing moderately ill infants requiring various types of special neonatal care and the third very low birthweight infants with relatively high probabilities of poor outcome for most of the eleven indicators of neonatal health and a 0.15 probability of death. In trial runs with different categorisations and sets of variables similar results were obtained for up to four classes, whereas the 5- and 6-class models differed from classes IVa-IVd by the division of one or two classes on a single variable. Unlike the 4-class model the results for the five and six classes were therefore very dependent on the choice and categorisation of individual variables. For these reasons it was decided to focus only on the 3- and 4-class models even though $-2\log\Lambda$ continued to fall substantially as more classes were included. The 4-class model was recommended for the further study of newborn infants, but the effects of missing data and the stability of the technique were investigated using only three, since searching for maxima of the likelihood of the 4-class model involved extensive repeat analyses. Multiple maxima were not found for the 3-class model.

The assumption of independence within classes underlying the latent class model can rarely be the case in practice. The aim of this analysis was to approximate a large contingency table by classes which most closely reflect conditional independence. Checks were made on the independence of selected pairs of variables within classes. Association remained even after fitting six classes, but generally the pairwise associations within the four classes were considerably less than their marginal counterparts. A typical example of this reduction is the association between birthweight and death at discharge. The marginal test for this association had value 11,254 with 3 dgrees of freedom, while the corresponding values within classes were 195.39, 0.01, 9.71 and 0.00. The association between jaundice and birthweight below the 10^{th} percentile increased however; the marginal test of association had 12.85 with 1 degree of freedom, whereas the within class test values were 6.86, 147.27, 2.26 and 3.47

5.4 Missing Data and the Stability of the 3-Class Model

Approximately 13 per cent of the population had a missing value on at least one of the variables and it was possible that, had these values been recorded, substantially different classes would have emerged. In the absence of studies which have examined the reasons for missing data, this possibility was investigated by examining whether or not the results withstood various assumptions about the nature of the missing data. First, missing data were assumed to indicate the best possible value, secondly, the worst possible value, and thirdly, were assumed to occur at random, as defined by Rubin (1976). Under this last assumption the summations in the M-step formulae for θ and λ are performed over all cases, whether complete or with a missing value, and the quantities z_{ik} are calculated by multiplication over j with x_{ij} non-missing. The results of these procedures are presented in Table 5.3 under columns labelled BP, WP and MR respectively. Procedures BP and MR both found similar classes to the complete cases. Classes WPa and WPb were similar to IIIa and IIIb, but WPc, having high probabilities of low Apgar score and intubation, was more akin to IVd. These results reflect the fact that assigning the worst possible value, imposes the unrealistic assumption that 100 per cent of the missing data had adverse values.

The data set was very large and it was possible to take random halves and still have ample cases to estimate the 47 parameters of the 3-class model. Columns labelled S1 and S2 of Table 5.3 contain these parameter estimates; comparison with the results from the whole data set shows that the resultant classes were essentially the same.

The analyses were repeated on comparable data for 1978 and the results are shown in columns labelled 78 of Table 5.3. Class 78a is similar to IIa but has lower prevalence. Class 78b has distinctly higher probabilities of several indicators of poor health including death and a lower prevalence at 0.01 (compared to 0.03) than IIIb. Class 78c has lower probabilities of low birthweight and birthweight below the 10th percentile than IIIc but has higher probabilities of a low Apgar score and assisted resuscition than IIIc; 78c also has a higher prevalence at 0.12 (compared to 0.05 for IIIc). In general, however, the

TABLE 5.3

PARAMETER ESTIMATES FOR 3-CLASS MODELS

VARIABLE	CLASS a							CLASS b							CLASS c						
	IIIa	BPa	WPa	MRa	S1a	S2a	78a	IIIb	BPb	WPb	MRb	S1b	S2b	78b	IIIc	BPc	WPc	MRc	S1c	S2c	78c
Birthweight 2001-2500g	.00	.00	.02	.00	.00	.00	.00	.18	.12	.41	.12	.22	.15	.10	.79	.62	.03	.61	.83	.74	.38
1501-2000g	.00	.00	.00	.00	.00	.00	.00	.26	.23	.21	.24	.27	.25	.20	.09	.11	.00	.11	.09	.09	.08
≤1500g	.00	.00	.00	.00	.00	.00	.00	.21	.34	.14	.36	.19	.23	.43	.01	.01	.00	.01	.01	.01	.01
B.weight < 10th centile	.07	.06	.11	.07	.07	.07	.06	.19	.19	.39	.20	.19	.19	.23	.62	.51	.11	.52	.65	.59	.39
Apgar at 5 min < 7	.01	.01	.04	.01	.01	.01	.01	.26	.37	.17	.40	.21	.30	.46	.01	.02	.85	.02	.01	.02	.16
Intermed. resusc.	.08	.08	.08	.08	.08	.08	.05	.26	.23	.19	.24	.24	.27	.13	.13	.14	.00	.16	.14	.13	.13
Resusc. by intubation	.02	.02	.04	.02	.02	.02	.02	.33	.43	.25	.47	.31	.35	.50	.04	.05	.99	.06	.03	.04	.21
Assisted ventilation	.00	.00	.00	.00	.00	.00	.00	.25	.38	.18	.41	.22	.28	.50	.00	.00	.00	.01	.00	.00	.01
Recurrent Apnoea	.00	.00	.00	.00	.00	.00	.00	.17	.27	.14	.29	.15	.19	.41	.00	.00	.00	.00	.00	.00	.00
Jaundice	.28	.26	.28	.26	.28	.28	.29	.67	.56	.65	.57	.69	.65	.40	.49	.53	.09	.54	.48	.50	.46
Convulsions	.00	.00	.00	.00	.00	.00	.00	.07	.09	.09	.10	.07	.08	.07	.00	.01	.00	.01	.00	.00	.02
In tube feeding	.01	.01	.01	.01	.01	.01	.00	.60	.59	.45	.61	.59	.61	.57	.10	.19	.01	.19	.06	.14	.17
Dead at discharge	.00	.00	.00	.00	.00	.00	.00	.13	.25	.10	.27	.12	.14	.43	.00	.00	.00	.00	.00	.00	.00
Age 4-10 days	.83	.83	.84	.83	.83	.83	.84	.09	.11	.15	.09	.08	.09	.13	.53	.47	.71	.47	.55	.52	.47
" ≥11 days at discharge	.03	.03	.03	.03	.03	.03	.03	.79	.67	.76	.67	.81	.79	.51	.45	.51	.04	.51	.43	.46	.51
Frequency of Class	.92	.91	.89	.91	.92	.92	.86	.03	.02	.06	.02	.03	.03	.01	.05	.07	.05	.07	.05	.06	.12

Note: IIIa, IIIb, IIIc complete cases
BP = Best possible values
WP = Worst possible values
MP = Missing at random
S1 = Sample 1
S2 = Sample 2
78 = Complete cases 1978 data

interpretation of the classes was largely unaltered. When four classes were fitted to the 1978 data, differences in parameter estimates were again noted. For example the second 1978 class although less prevalent, had a higher probability of death at 0.40 (compared to 0.15 in 1980), but the general interpretation of the classes remained the same. The differences could reflect changes in an infant population whose neonatal death rate declined by 11 per cent between 1978 and 1980, or could be due to the increase in coverage of the SMR11 scheme between 1978 and 1980.

5.5 Use of the Classification

The classes have various uses in health service research. They could, for example, be used in estimating the need for hospital services, or in directly standardising utilisation of sevices within classes for a health board by the overall Scottish prevalence of classes. In this section a method is described of estimating quantities such as the prevalence of classes within health boards and the proportion of infants using some type of service within classes.

Let s, $1 \leq s \leq N_H$, represent the N_H live births in health board H. Membership probabilities (z_{sk}, $1 \leq k \leq 4$) of the four classes IVa-IVd can be calculated from the E-step formula for z_{ik} using the variable combination, $\mathbf{x}_s$, for infant s. Assuming data are missing at random, when infant s has a missing value on one or more variables $p(\mathbf{X}=\mathbf{x}_s|\text{class } k)$ is based on only those variables which are recorded. Each infant could then be assigned to the class with highest membership probability or remain unclassified if no one z_{sk} was substantially higher than the others. The certainty with which an individual is assigned can be used to weight its contribution in estimates of prevalence given by

$$P(\text{class } k \text{ in health board } H) = \sum_{s=1}^{N_H} \frac{z_{sk}}{N_H}.$$

When individuals are assigned to classes, the z_{sk}s are, in effect, replaced by 0 or 1 and the estimates reduce to relative frequencies within health boards. Weighted estimates of prevalence for the Greater Glasgow Health Board and Tayside Health Board are given in Table 5.4. Greater Glasgow is characterised by higher prevalences of classes IVb, IVc and IVd representing ill-health, than either Tayside or Scotland. The proportions of the four classes in Scotland vary slightly from those of the 4-class model in Table 5.1 because cases with missing values are included in the prevalence figures in Table 5.4.

Similar estimates can be obtained for the joint probabilities of each class and the use of a service by multiplying z_{sk} in the above formula by an indicator variable for the service. The within class average of a continuous variable, such as duration of stay, can also be estimated using the same procedure.

TABLE 5.4

CLASS PREVALENCE FOR GREATER GLASGOW AND TAYSIDE IN 1980

Health Board	Livebirths*	Class			
		VIa	IVb	IVc	IVd
Greater Glasgow	13661	.875	.036	.054	.035
Tayside	5082	.905	.026	.048	.022
Scotland	52022	.885	.038	.051	.027

* excluding congenitally malformed

CHAPTER 6 : CONCLUSIONS

6.1 The Regression Studies

6.1.1 The Strengths and Weaknesses of SMR2 Data

SMR2 information is used for a variety of purposes, including the construction of official statistics; hospital planning; and clinical and epidemiological research. Production of the data involves the co-operation of large numbers of clinical staff and record officers, and is expensive and time consuming. Taken together, the nature and scale of the operation limit the information that can be collected and consequently the data are not tailored to answer specific research questions. To compensate for any lack of detailed information, however, the SMR2 scheme is unique in Britain in providing population data for the obstetric period with approaching 100 per cent coverage. The SMR2 data thus have a considerable advantage over hospital based studies which may be unrepresentative of either the severity of patients, or the socio-economic group from which they come, and have many fewer cases; over studies based on smaller geographical units or single health boards which provide fewer cases and may be unrepresentative of socio-economic groups; and over the national birth surveys which also provide fewer cases and do not allow the examination of trends over consecutive years. Epidemiological studies based on routinely collected national data thus fill a gap left vacant by other types of study. The large sample sizes make possible the examination of rare events with sufficient power to detect quite small differences in rates. They tend, however, to suffer from a lack of control for specific confounding variables, and are not suited to the confirmation of novel hypotheses because when new variables are incorporated there is a time lag of several years before the data become available for analysis.

6.1.2 The Choice and Definition of Covariates

The regression analyses of chapter 4 exploited the range of information available on the SMR2 document in an exploratory study of low birthweight and gestational age. No one covariate was the focus of attention, and indeed in designing a project to examine some of the covariates much additional information would have been desirable. The relationship between some of the covariates and perinatal outcome has been well documented in earlier studies and the results of the analyses were generally consistent with this literature. Other covariates have not been the subject of many studies and in these cases the results should, perhaps, be regarded as initial findings to be confirmed by other studies. For example, a history of induced abortion has only rarely been examined using British data. Section 6.1.4 comprises a discussion of the results relating to this variable.

The covariates selected for study mainly describe the status of the woman prior to conception. The one exception, sex of infant, was, like the others, constant throughout pregnancy. An examination of the mediating effects on perinatal outcome of various pathological conditions occurring during pregnancy, available on the SMR2 document, was not undertaken and analysis was limited to the relationship between the chosen covariates and the three measures of perinatal outcome. The objective of this study was to document risk to pregnancy experienced by women in different demographic, obstetric and social groups rather than to identify specific causal influences.

The difficulty of including pathological conditions that arise during pregnancy is emphasized by the structure of the survival model for gestational age. A condition developing during the study period should properly be treated as an alternative pregnancy outcome competing with delivery, usually having an effect only on the subsequent hazard of delivery. Because the time (in terms of gestational age) at which such conditions occur is not recorded on SMR2 it is not possible to include them in this way. If they are considered as regressors in the proportional hazards model, the multiplicative effect is operational before the problem developed. This might be sensible if the time at which the condition is noticed relates to the emergence of a symptom, but the underlying risk is present throughout the study period. If a condition represents a new pregnancy environment and a consequent change in the risk of delivery its inclusion as a regressor with a proportional effect on the hazard thoughout the study period is not appropriate.

The regression analyses were limited by the size of data that the algorithms could handle and by the number of parameters that could be estimated. An important consideration when defining the covariates was to keep the number of categories, and hence the number of distinct covariate combinations which formed the cases of the analysis, and also the parameters in the models, to a minimum. In the case of the covariates measuring previous obstetric history, a finer categorisation would probably have added little to the findings because obstetric histories including high numbers of poor outcomes are rare. Women aged 25-34 and $\geq$35 years and primiparae aged <18 years experienced lower risks of an SGA infant than women aged 18-24 years, an unexpected finding. Since two of the categories of maternal age, 18-24 and

25-34 years, comprised the majority of the population a finer categorisation of age was possible and might have been helpful in explaining these results. Both social class and maternal height were associated with gradients in the risk of all three outcomes and examination of the five social classes separately or the inclusion of more height categories would probably have accentuated these results. The categorisation of maternal height concentrated on the high risk group of short women, <150 cm, who were examined separately but represented under 4 per cent of the population, whereas the category of tall women, $\geq$165 cm, represented approximately 25 per cent. Marital status could also have been coded differently by considering separated, divorced and widowed women either as a group on their own or by combining them with single rather than married women.

Two of the covariates, maternal age and height, were continuous and, as such, could have entered the model linearly or with quadratic or higher order terms. This would have substantially increased the number of distinct cases in the analysis, and the models considered here could not have been fitted. An alternative approach might be to treat birthweight, for example, as a continuous dependent variable and consider an ordinary linear regression. Linear regression can be performed using the covariance matrix of the data and the storage required depends on the number of parameters and not cases. Several assumptions concerning the distribution of birthweight underlie such an analysis which, as discussed in sections 6.1.6 and 6.1.8, are not met by the data.

6.1.3 Comparison of Adjusted and Unadjusted Risks

In general, adjusted parameter estimates were less extreme than their unadjusted values, and the increase in likelihood due to each covariate was smaller in the presence of controlling variables. The pattern of risk associated with most covariates, however, was not substantially changed after controlling, and it can be concluded that these covariates made some contribution to the association with the three perinatal outcomes. At one extreme, the risks associated with sex of infant were virtually unaltered by the presence of the other covariates in the regression, suggesting that this variable acts more or less independently of the others. The risks associated with single marital status and a history of caesarean section or ≥ 3 previous livebirths, on the other hand, showed the greatest attenuation. For marital status this may largely be due to the inclusion of social class, while other variables, possibly maternal age, may explain the attenuation of risks associated with the latter two covariates. Even in these cases, some degree of increased risk still persisted after adjustment.

6.1.4 Risks Following Induced Abortion.

Although, internationally, there have been many studies of pregnancy following induced abortion, no general pattern of risk has emerged, possibly because clinical practice and the type of women choosing to terminate pregnancy varies to the extent that it is necessary to assess the situation separately for each country. Few studies of pregnancy following induced abortion based on British data are available, and none is population based, so that SMR2 data are particularly valuable in establishing patterns of risk. A validation of the sensitive histories of induced abortion stated by women was not undertaken, but SMR2 data was shown in section 2.1.5 to be reasonably reliable in this area.

Primiparae with a history of induced abortion experienced increased risks of birthweight below 2,500 gms, 2,000 gms, 1,500 gms and 1,000 gms after adjusting for the other covariates. Examination of the hazards of preterm delivery and the risks of an SGA infant showed increased hazards for primiparae with a history of induced abortion, but if anything a marginal decrease in the constrained odds ratio of an SGA infant. Multiparae with a history of induced abortion also suffered increased risks of low birthweight and preterm delivery, but their risk of having an SGA infant was only slightly higher than that of multiparae with no history of induced abortion.

It might be expected that if induced abortion has no effect on subsequent pregnancy, the experience of primiparae with one induced abortion would be similar to that of women in their second pregnancy, who experience generally lower risks. In a separate study of preterm delivery and the birth of an SGA infant, Pickering & Forbes (1985) introduced a second control group of women whose only previous pregnancy had ended in a livebirth. Women in this control group experienced lower risks of preterm delivery and the birth of an SGA infant than primiparae with both none or one previous induced abortion.

The findings reflect the constraints imposed by the information available on SMR2. Although it was possible to control for several relevant variables, cigarette consumption, which has been included in several other studies (Meirik & Bergstrom, 1983; WHO Task Force, 1979) and is a known correlate of birthweight, was not available. It was also not possible to investigate whether abortions performed at different gestational ages or by different techniques had similar effects on subsequent pregnancies. Additional information on legally induced abortion is published for the Scottish population (Common Services Agency, 1975-82). The principal methods used for abortion in the years preceding the study were vacuum aspiration (72 per cent), dilatation and curretage (11 per cent) and hysterotomy (1.5 per cent). The remaining 15.5 per cent are classified as 'other' and include terminations induced by prostaglandins the most common method of abortion between three and six months. The same source also shows that 42 per cent of terminations were performed before the 10th week of gestation and 87 per cent before the 14th week.

6.1.5 Comparison with Longitudinal Studies

In section 1.2.7 the difficulties of interpreting studies based on longitudinal or cross-sectional data were outlined. Longitudinal studies are typified by a decline in risk with increasing pregnancy order within families of a given final size. The decline is partly caused by the tendency to terminate child bearing after a successful pregnancy. Cross-sectional studies reveal a different picture with women in their first pregnancy suffering higher risks than women in their second, and in their fourth and subsequent pregnancies women also experience increased risk. The risk associated with high pregnancy order in cross-sectional studies reflects the disproportionate numbers of 'poor reproducers' compensating for earlier losses. The regression analysis reported here was based on population data and the typical cross-sectional pattern of risks associated with high pregnancy order might therefore be anticipated.

Several aspects of this study may reduce the effects of pregnancy compensation. The analysis of multiparae did not include parity, per se, as a covariate. Rather the number of previous livebirths was considered. Women with three or more livebirths have initiated a further pregnancy in spite of their earlier successful births (although compensation for infant death is a possibility). The analyses were also controlled for history of perinatal death and spontaneous abortion, the factors that introduce bias to cross-sectional studies. In spite of these precautions, women with ≥ 3 previous livebirths experienced slightly increased risk of birthweight below 2,500 gms, preterm delivery and the birth of an SGA infant, and hence conformed to the traditional J-shaped pattern of cross-sectional studies.

Maternal age is correlated with pregnancy order and cross-sectional studies tend to reveal increased risks for both extremes of the age distribution. The same factors that confound studies of high parity apply to maternal age, and there is, additionally, a possibility of bias due to sub-fertility, since women taking longer to achieve pregnancy and hence reaching higher age groups, may experience poorer pregnancy outcome. In this study, primiparae and multiparae aged <18 years and ≥ 35 years experienced increased risk of low birthweight and preterm delivery after adjusting, with primiparae ≥ 35 years and multiparae <18 years experiencing particularly high risk. The risks experienced by older women may in part be due to congenital malformations, which were not excluded from the study because not all malformations are identified at the time SMR2 documents are completed. The lower risk of SGA infants associated with primiparae aged <18 years and multiparae aged ≥ 35 years is not readily explicable and is not confirmed by other studies.

6.1.6 Analysis of Birthweight

The findings relating to birthweight were generally consistent with earlier studies. The examination of the four birthweight cutpoints separately revealed that some covariates were associated with more extreme risks of birthweight below 2,500 gms than below the lower cutpoints. For example, maternal height was a major predictor of birthweight below 2,500 gms, but the estimated odds ratios of birthweight below 1,500 gms and 1,000 gms for women of height <150 cm or $\geq$165 cm were much reduced. Similarly, marital status and social class were associated with decreasing differential in risks of birthweight below 2,500 gms, 2,000 gms, 1,500 gms and 1,000 gms. In two cases, amongst female infants and multiparae with $\geq$3 livebirths, only the risks below 2,500 gms were increased; at the other cutpoints the confidence intervals were fairly wide but estimated odds ratios were generally close to 1.0.

In comparison, several aspects of obstetric history were associated with more extreme risks below 1,500 gms and 1,000 gms than below 2,500 gms. For example, the risk associated with $\geq$2 spontaneous abortions increased steadily (2.01, 2.76, 3.25, 4.35 for primiparae and 2.04, 2.69, 3.52, 3.86 for multiparae) in the analysis of birthweight below 2,500 gms, 2,000 gms, 1,500 gms and 1,000 gms. Similar trends were also observed following a history of one spontaneous abortion, induced abortion (although for multiparae the risk below 1,000 gms did not follow the trend) and previous perinatal death.

These trends in risk suggest that some factors are associated with some degree of lower birthweight, resulting in higher than expected numbers of births in the 2,000 to 2,499 gms category, but no appreciable increase in the risk of very low birthweight infants. Amongst these are marital status, social class, maternal height and ≥ 3 previous livebirths. These variables are socially determined or, like maternal height, may reflect the social background of the woman. One further factor, sex of infant, fell into this group being associated with some increase in the risk of moderately low (2,000 to 2,499 gms) birthweight. On the other hand, several aspects of obstetric history showed a different pattern, and were associated with much higher risks of very low birthweight infants, below 1,000 gms. Factors such as a history of perinatal death and spontaneous abortion, both of which indicate poor performance in earlier pregnancies, are of this second kind. A history of induced abortion also fell into this group, possibly reflecting cases where termination was a medical necessity rather than a form of late contraception.

6.1.7 Analysis of Preterm Foetal Survival Times

Proportional hazards models provide a natural approach to modelling foetal survival times which has been used only rarely to examine the relationship between covariates and gestational age. The majority of studies, in choosing to examine preterm delivery, have opted for a simple dichotomisation of gestational age. On the whole the results from this study might have been predicted from the literature on risk of preterm delivery. One study that does adopt a similar approach of examining hazards throughout pregnancy (Harris, 1981) provides a more direct comparison for some of the covariates. Substantially lower estimates of the relative hazard associated with maternal age, single marital status and previous perinatal loss were found by Harris than are reported in section 4.3. The discrepancy may be explained by the assumption in Harris's study that proportionality holds up to and beyond term. If the relative hazards are lower during the later weeks of pregnancy as was suggested here, their representation in the sample would swamp the pattern at earlier weeks.

The tests for full time dependency revealed one significant result, that for maternal age amongst primiparae. No covariate showed a significant linear time trend, although some covariates tended to have more extreme hazards in earlier weeks. For example, although the hazard during early weeks was high for primiparae with previous spontaneous abortions, by week 36 their hazard was similar to that of primiparae with no history of spontaneous abortion. Figure 4.6.b showed that the logarithm of the weekly hazards when a linear time trend was included in the model was a close match to the converging empirical hazards in 4.6.a. For this covariate one might expect such a pattern. The tendency to repeat obstetric outcome suggests that primiparae whose earlier pregnancies have ended spontaneously before the 28^{th} week are more likely to suffer increased hazard during roughly the same period than later in a subsequent pregnancy. The hazards of a spontaneous abortion or delivery around week 28 for these women might thus be expected to be greater than that of delivery around week 36. SMR2 information is available from late spontaneous abortions which were delivered in obstetric wards, and it may be possible to extend the study period to include weeks 20 to 28.

6.1.8 Analysis of Birthweight Standardised for Gestational Age

Although no previous studies have investigated changes in the distribution of birthweight standardised for gestational age using the approach adopted here, Scott et al (1981; 1982) and Ounsted et al (1985) examine the risks of the birth of an SGA or LGA infant. Scott et al (1982) reported increased risk of an LGA infant amongst women of parity ≥ 3, consistent with the increased risks of an LGA infant amongst multiparae with ≥ 3 livebirths in this study. After adjusting for factors that included smoking, maternal weight, net pregnancy weight gain, hypertension and pre-eclampsia (which was the most important predictor of SGA) no association between SGA and social class remained (Scott et al, 1981). Before adjusting social class did have a weak association with the risk of an SGA infant. In this study social class was associated with the birth of an SGA infant after adjustment, and it may be that some covariate in the earlier study but not included here partly explains the association. Of the possible adjusting variables, only pre-eclampsia was reported by Ounsted et al (1985) to be associated with increased risk of SGA but not LGA infants, which suggests that it may mediate between social class and the birth of an SGA infant. Older women experienced increased risk of an LGA infant (Scott et al, 1982), as did women ≥ 35 years in this study, but after adjusting for parity, maternal height and weight, siblings birthweight, smoking and net pregnancy weight gain the association disappeared. The unusual pattern of risk of an SGA infant associated with maternal age found in this study is not confirmed by the earlier reports.

By allowing covariate effects to vary over response categories in the analysis of birthweight standardised for gestational age it was possible to examine quite complex patterns of association. Some covariates were described by shifts in location, while others were associated with increased probability in either one or both tails of the distribution. This complexity can not easily be examined within a linear regression model of birthweight controlling for gestational age. For example, associations involving increased risks in both tails may result in no change in the mean of the distribution, and the relationship could only be found by allowing changes in the variance in a regression model. While if the association was primarily with one tail of the distribution, some change in the mean would be anticipated, but the full pattern would not be apparent within the constraints of a Normal error structure.

Two methods of examining unconstrained risks were compared. Firstly, odds ratios were estimated in models where each covariate in turn was unconstrained while the others were constrained to have equal risk at each percentile. In the second method separate binary regressions at each percentile were fitted and thus the risks for all the covariates simultaneously varied over percentiles. The two approaches produced virtually identical results. The former generalised polytomous model was much more expensive to fit (section 6.1.11), but had the advantage of permitting a likelihood ratio test for changes in the odds ratios over percentiles. The association between the covariates and birthweight standardised for gestational age was generally reduced after adjusting for the other covariates. The similarity between estimates when adjusted by constrained effects in the first approach or by unconstrained covariate effects in the second, shows that it is unnecessary to adjust for the precise association with each covariate even when the unconstrained odds provided a substantially better fit to the association.

6.1.9 Checking the Fit of a Binary Logistic Model

Plots of observed against fitted rates of birthweight below each cutpoint showed that the logistic model over-estimated risks amongst cases predicted to be at highest risk. In the analyses of birthweight below 2,500 gms less than 0.5 per cent of births were over-estimated, while in most other analyses up to 5 per cent were over-estimated. In the analysis of birthweight below 1,000 gms for multiparae, however, the risk experienced by over 20 per cent of the population was over-estimated. In this case the fit might be improved by including interactions, but when the one significant interaction, that between maternal height and perinatal death, was included in the model very little difference was made to the plot of observed against expected risk. By including sufficient interactions, possibly of third or higher order, a better fit would be achieved but this might prove computationally impractical. The rank order of the predicted risks were generally close to that of the observed risks which suggests that the logistic model could be used to identify a given percentage of women at highest risk for special attention. Where only small percentages were over-estimated by the model, the fitted parameters adequately describe the levels of risk experienced by the majority, but should not be used to estimate extremes of risk in the population. When a higher proportion are over-estimated the results are less reliable.

The simulations of the deviance and Pearson residuals described in section 4.2.3 represent a simpler version of an empirical probability plot. Each covariate combination was assumed to have the same mean taken to be the average of the fitted means over distinct covariate combinations. This average was used because a typical covariate combination has low frequency but higher risk than the overall population. Residuals were calculated directly from the true mean, without re-estimating the model for each simulation. The numbers above and below the nominal 5^{th} and 95^{th} percentiles from a Normal distribution were calculated and compared to the observed number of residuals from the original data. Thus only these two parts of the distribution were compared. Omitting to re-estimate the mean for each simulation probably makes little difference considering the enormous sample sizes and resulted in considerable savings in computer time. The assumption of a constant mean probably does affect the results. The cases with smaller means are less accurately approximated by a Normal distribution and more extreme residuals might be expected. The inclusion of such extreme residuals in the actual data is not compensated by cases with higher than average means, where the approximation by a Normal distribution is better. However, the covariate combinations with lowest rates of low birthweight tend to have high frequency and are few in number.

The final method of checking the logistic model considered here involved testing the adequacy of the link function within a 2-parameter family of functions. This approach was not very successful since it was not possible to estimate two separate parameters in any but the models for birthweight below 2,500 gms, and confidence intervals were very wide. Pregibon (1980) has suggested that the available data may often provide insufficiently detailed information concerning these parameters.

These model checks were not employed in relation to the polytomous regression models. It would have been possible to use methods for binary regression to test the adequacy of fit below each week of gestation or below each percentile in the analysis of birthweight standardised for gestational age. It is likely that the problems of over-estimation of risk for cases predicted to be at highest risk carry over to the polytomous regression models also.

6.1.10 Comparison of Logistic and Proportional Hazards Models

The choice between logistic and complementary log-log link functions in situations where the response occurs with small probability typically makes little difference to the estimated parameters (Brenn & Arnesen, 1985). The two link functions only diverge to a notable degree for probabilities greater than 0.1, and where the response occurs with probabilities above this level the link function does make a difference. For example, in the analysis of birthweight standardised for gestational age the probability below 5 of the 7 percentiles was above 0.1 and the choice of link function did influence the results.

The two models for preterm gestational age produced similar results and fitted equally well. Results were primarily presented for the proportional hazards model because they incorporated the passage of time underlying preterm delivery, and made possible the examination of changes in the risk of delivery at different periods of pregnancy.

The logistic model fitted the data on birthweight standardised for gestational age better than the proportional hazards model, but the major improvement in fit related to only one covariate, maternal height, and the proportional hazards model provided a marginally better fit to the association with two covariates. There were inadequacies with both models, and several of the covariates could not be explained by simple multiplicative changes in risk within either model. Although the logistic model did provide the better fit, this was not the only criterion for choosing between the two link functions. The main reason for using the logistic was pragmatic. It provided results in terms of increased risks of either tail of the distribution conforming with the standard notation, while the hazard at each percentile has little intuitive meaning.

6.1.11 Computing Costs of Regression Modelling

Two specialised FORTRAN programs were written for fitting the polytomous regression models. The first fitted proportional hazards model to grouped survival data, and allowed for the inclusion of linear time dependent terms. The second was programmed for logistic and proportional hazards models and was adapted to permit risks relating to any combination of covariates, including interactions, to depend on the response category. Although only the second program was strictly necessary, the duplication of the proportional hazards model did provide a check on the accuracy of the programming, and the results from the two programs were identical.

A detailed comparison of various methods of fitting polytomous regression models was not attempted. It is possible that, for example, by reducing the accuracy by which the linear equations are solved within each iteration of the algorithm, some saving in computer time could be achieved without substantially changing the final solution. The algorithm for the generalised polytomous regression model was deemed to have converged when the total difference between successive parameter values was less than 0.01, and that for the proportional hazards models when the Euclidean distance between parameter values was less than 0.01. An examination of the impact of varying the stopping rule may also result in savings in computer time. Alternative approaches to estimation could be considered. For example the models can, in theory, be fitted in GLIM. In practice, however, the workspace available in GLIM on the Glasgow computer was insufficient to store the larger of the two data set, that for multiparae, in the appropriate form of covariate data for each level of the response separately for all covariate combinations. Although savings in computer time may be possible, the computing costs by the methods used here do give a general idea of the expense of this type of modelling.

In Table 6.1 the costs of a variety of typical jobs, measured in seconds of CPU time, performed on the Glasgow University ICL 2988 mainframe are given. In the first rows the cost of fitting individual models for birthweight below 2,500 gms within GLIM are shown to be lower for primiparae than multiparae. This was a result of the lower number of parameters that were estimated from fewer cases for primiparae. The costs tended to be higher for the lower birthweight cutpoints where the data conveyed less information.

The next rows show computing costs for the proportional hazards and polytomous logistic models. These are considerably more expensive than the binary logistic models because of their

TABLE 6.1

COMPUTING COSTS OF TYPICAL REGRESSION JOBS ON THE GLASGOW UNIVERSITY COMPUTER

JOB DESCRIPTION	CPU Tune (seconds)	
	PRIMIPARAE	MULTIPARAE
GLIM, BIRTHWEIGHT < 2500 gms		
One covariate, unadjusted*	7	15
All covariates, simultaneously	12	40
All covariates + one interaction*	20	55
FORTRAN, GESTATIONAL AGE		
One covariate,** unadjusted	80	320
All covariates, simultaneously	540	3100
One linear dependent covariate** effect, unadjusted	450	1100
FORTRAN, BIRTHWEIGHT STANDARDISED FOR GESTATIONAL AGE		
One covariate**, constrained, unadjusted	45	110
All covariates, constrained	195	860
One covariate** unconstrained, others constrained	500	2000

* Averaged over all covariates/interactions
** Typical of a covariate with 3 levels

greater complexity and the number of parameters being fitted. The models for gestational age involved more covariates (sex of infant was excluded from the analyses of birthweight standardised for gestational age), and the models took longer to fit.

Including covariate effects that were dependent on the response category considerably increased the computing in both polytomous models. In the proportional hazards model only unadjusted linear dependent effects were considered, and these involved only one extra parameter for each level of the covariate above two. Computing costs for these models approached those of the models where all covariates were included in a proportional hazards model. In the example of a polytomous logistic model with covariate effects varying at all levels of the response (for a covariate with three levels illustrated in Table 6.1), an extra 12 parameters would be estimated compared to 18 and 21 parameters in the adjusted constrained models for primiparae and multiparae respectively. Computing times were more than doubled when unconstrained effects were included in the adjusted model. Binary logistic regressions below each percentile and polytomous logistic models in which each covariate alone was unconstrained produced virtually identical results, but in Table 6.1 it can be seen that there was a huge difference between the computing costs of the two approaches. The cost of repeating an adjusted binary regression at the 7 percentiles was around 84 seconds using the cost figures from the low birthweight model, while the repetition of unconstrained analyses took around 3,000 seconds for primiparae, and for multiparae the latter approach cost approaching 20,000 seconds (approximating total costs over all covariates from figures in Table 6.1).

In summary, the polytomous regression models were expensive to fit, especially when covariate effects were allowed to depend on the response level.

6.2 The Neonatal Classification

Latent class analysis was able to construct a classification of neonates which remained substantially unchanged when repeated on random halves of the data and under various assumptions about missing data. When the analysis was performed on two years of data the classes altered only slightly. The 4-class model contained one class of 'normal' infants occurring with high probability, and made a distinction between three other classes. The second class described infants with high probabilities of severe outcomes on most variables and in particular a 0.15 probability of death before discharge. The final two classes had low probabilities of death, the third class associated with symptoms reflecting moderately low birthweight and SGA infants, whereas infants in the fourth suffered from problems immediately after birth, as indicated by high probabilities of a low Apgar score or the need for resuscitation.

The classification was based on variables measured throughout the neonatal period, some of which may be influenced by neonatal care early in life. The alternative of using only indicators measured at delivery, which in turn may depend on treatment during labour, has the drawback of missing conditions with symptoms that appear later and require specialised neonatal care.

The SMR11 document remains open until the infant is discharged or transferred out of neonatal wards, so the scheme does not have a fixed time period. A condition arising after an infant has been transferred to a medical ward or discharged home, which could be within a few days of delivery, will not be reported on the SMR11 document, while the SMR11 of an infant kept under surveillance in a neonatal ward may contain information covering a period of several weeks. In practice very few infants discharged home develop serious problems, but transfer to other wards depend on individual hospital policy, and under reporting for these infants and could vary between hospitals.

The classes do not identify underlying sub-populations within each of which the eleven variables are strictly independent, but by using this formulation it proved possible to draw a distinction between various patterns of neonatal health. It is possible to take account of the arbitrariness of assigning an individual to a class by incorporating assignment probabilities. In measuring the case-mix within various regions, for example, estimates of class prevalence can be weighted by these probabilities thus making use of the uncertainty of class membership.

As with most statistical clustering techniques, there are no theoretical methods for choosing an appropriate number of classes. In some applications where interpretation of the classes is important, the purpose of the study may dictate an appropriate number of classes, or, as is the case here, the results may become unsatisfactory as more classes are included. Starting from the 1-class model each additional class up to the fourth was considered a useful division of the data, but beyond the 4-class model the results became very dependent on the choice and categorisation of the variables used, and multiple maxima were increasingly a problem. The log likelihood increased considerably between the 1- and 2-class models and by a reduced amount with each additional class up to the sixth. Although these values were not the major factor in choosing the number of classes, they did indicate that the 1- and 2-class models provided inadequate descriptions of the data.

In practice, the usefulness of the neonatal classification depends on its performance in describing hospital case-mix and the reliability of subsequent resource allocation. In this study several desirable aspects of the classification have been described but some drawbacks to its use have been suggested. A full description of its viability as a measure of case-mix and resource allocation has not been attempted here and remains the subject of further study.

APPENDIX 1

SMR2 DOCUMENT

MEDICAL IN CONFIDENCE

SCOTLAND MATERNITY DISCHARGE SHEET

SMR2(A)

281556 CT3 DUP 2-7

1. GENERAL INFORMATION CT1 SERIAL (2-7)

- Hospital Code — 8-12
- Hospital Case Reference Number — 13-22
- Surname — 23-34
- Forename — 35
- Second Initial — 36
- Maiden Name — 37-48
- Age* ____ Date of Birth 19 — 49-56
- Marital State — 57
- Home Address*
- Post Code — 58-64
- Occupation
  - – Patient — 65-67
  - – Husband — 68-70
- Date of Marriage — 71-76

CT2 DUP 2-7

- Obstetrician — 8-13
- Family Doctor — 14-19
- Type of Antenatal Care — 20

2. PREVIOUS PREGNANCIES

- Total Number — 21
- Spontaneous Abortions (Miscarriages) — 22
- Therapeutic Abortions — 23
- Caesarean Sections — 24
- Perinatal Deaths — 25
- Children now Living — 26

3. CURRENT PREGNANCY

- Date of Admission — 27-32
- Admitted From — 33
- Number of Previous Admissions to Any Hospital in this Pregnancy — 34
- Type of Admission — 35
- Date of Booking — 36-41
- Original Booking for Delivery — 42
- Blood Group ____ Rh — 43
- Height ____ ft ____ ins* — 44-46
- Type of Abortion — 47
- Management of Abortion — 48
- Sterilisation after Abortion — 49
- Principal Complication of Abortion — 50
- Last Menstrual Period — 51-56
- Estimated Gestation at Abortion or Delivery — 57-58
- Certainty of Gestation based on LMP — 59

4. MATERNAL DISCHARGE DATA

- Date of Discharge — 60-65
- Condition on Discharge — 66
- Discharged To — 67
- Category of Patient — 68
- Unit on Discharge — 69

* Entries at these items are for hospital use only.

5. RECORD OF LABOUR

- Method of Induction of Labour — 8
- Presentation at Delivery or start of Operative Delivery — Baby 1 — 9; Baby 2 — 10
- Mode of Delivery — Baby 1 — 11; Baby 2 — 12
- Duration of Labour (In Hours) — 13-14
- Sterilisation after Delivery — 15
- Date of Delivery — 16-21
- Number of Births this Pregnancy — 22
- Outcome of Pregnancy — Baby 1 — 23; Baby 2 — 24
- Birthweight (GMS) — Baby 1 — 25-28; Baby 2 — 29-32
- Apgar Score at 5 mins — Baby 1 — 33; Baby 2 — 34
- Sex — Baby 1 — 35; Baby 2 — 36

6. POSTNATAL RECORD OF INFANT(S)

- Special Care Baby Unit — Baby 1 — 37; Baby 2 — 38
- Baby Discharged To — Baby 1 — 39; Baby 2 — 40
- Case Record No. in this Hospital — Baby 1 — 41-50; Baby 2 — 51-60

To be specified by Clinician

CT4 DUP 2-7

- Underlying Cause of Stillbirth or Baby Death — Baby 1 — 8-11; Baby 2 — 12-15

7. MAIN CONDITION — 16-21

8. OTHER CONDITIONS — 22-27; 28-33; 34-39; 40-45; 46-51

9. OPERATION — 52-55

- National Use — 56-67
- Local Use — 68-79

SMR2 DOCUMENT (REVERSE)

KEY TO CODED ITEMS

SMR2(A)

Marital State [57]
1 = Never married (Single)
2 = Married
3 = Widowed
4 = Divorced
5 = Separated
8 = Other
9 = Not Known

Type of Antenatal Care [20]
0 = None
1 = GP Only
2 = GP care with specialist consultation
3 = Hospital Only
4 = GP and Hospital Shared
8 = Other
9 = Not Known

Admitted from [33]
0 = Not admitted
1 = Home
2 = Other hospital
3 = GP unit outwith this hospital
4 = Other speciality in this hospital

Type of Admission [35]
0 = Domiciliary (Not Admitted)
1 = Abortion (includes threatened abortion and ectopic pregnancy)
2 = Pregnant but not in labour
3 = In Labour
4 = Born before arrival
5 = Admitted after delivery at home
6 = Admitted after delivery in any hospital
8 = Other (e.g. doubtfully pregnant)

Original Booking for Delivery [42]
0 = Not booked prior to this admission
1 = Booked for Home delivery
2 = This Hospital (Consultant Unit)
3 = This Hospital (GP Unit)
4 = Other Hospital (Consultant Unit)
5 = Other Hospital (GP Unit)
9 = Not Known

Blood Group [43]
1 = O Rh −ve
2 = O Rh +ve
3 = A Rh −ve
4 = A Rh +ve
5 = B Rh −ve
6 = B Rh +ve
7 = AB Rh −ve
8 = AB Rh +ve
9 = Not Known

Type of Abortion [47]
0 = Threatened Abortion (still pregnant on discharge)
1 = Spontaneous or incomplete abortion
2 = Missed abortion
3 = Hydatidiform mole
4 = Therapeutic Abortion
5 = Suspected illegal abortion
6 = Failed therapeutic abortion
7 = Ectopic Pregnancy
8 = Unspecified abortion

Management of Abortion [48]
0 = Not operative (i.e. management of threatened or spontaneous complete abortion)
1 = D+C
2 = Vacuum aspiration
3 = Hysterotomy
4 = Prostaglandin (all forms)
5 = Amniotic infusion (other than Prostaglandin)
8 = Other (including ectopic pregnancy)
9 = Not stated

Sterilisation after Abortion [49]
0 = None
1 = Laparoscopy
2 = Laparotomy
3 = Laparoscopy Other hospital
4 = Laparotomy Other hospital
8 = Other
9 = Not stated

Principal Complication of Abortion [50]
0 = None
1 = Haemorrhage
2 = Sepsis
3 = Trauma to Cervix or uterus
4 = Damage to bowel
5 = Retained products requiring re-evacuation
8 = Other
9 = Not stated

Certainty of Gestation [59]
0 = Not applicable
1 = Certain
2 = Uncertain
9 = Not known

FULL INSTRUCTIONS FOR COMPLETING THIS FORM ARE CONTAINED IN THE SMR 2 MANUAL.

Condition on Discharge [66]
0 = Domiciliary Delivery
1 = Still pregnant
2 = Aborted (all types of completed abortion)
3 = Delivered
4 = Post natal care only
5 = Pregnancy not confirmed
8 = Other (e.g. known missed abortion)

Discharged to [67]
0 = Domiciliary Delivery
1 = Home Care
2 = Other hospital - GP maternity unit
3 = Other hospital - specialist maternity unit
4 = Other hospital or institution
5 = Other unit in this hospital
6 = Died (PM)
7 = Died (No PM)
8 = Other

Category of Patient [68]
1 = Amenity
2 = Paying
3 = NHS
7 = Special arrangement (see manual)

Unit on Discharge [69]
1 = Obstetric (Consultant)
2 = Obstetric (General Practitioner)
3 = Home or Other confinement not admitted to hospital
4 = Day Case (for definition see manual)
9 = Other or Not Known

Method of Induction of labour [8]
0 = None
1 = ARM
2 = Oxytocics
3 = ARM + Oxytocics
4 = Prostaglandins
5 = ARM + Prostaglandins
6 = Prostaglandins + Oxytocics
7 = Prostaglandins + ARM + Oxytocics
8 = Other
9 = Not Known

Presentation at Delivery or start of Operative Delivery (Baby 1 and Baby 2) [9], [10]
1 = Occipito - anterior
2 = Occipito - posterior
3 = Occipito - lateral
4 = Breech
5 = Face/brow
6 = Shoulder
7 = Cord
8 = Other
9 = Not Known

Mode of Delivery (Baby 1 and Baby 2) [11], [12]
0 = Normal, spontaneous vertex, vaginal delivery, occipito - anterior.
1 = Cephalic vaginal delivery with abnormal presentation of head at delivery, without instruments, with or without manipulation.
2 = Forceps, low application, without manipulation, forceps delivery NOS
3 = Other forceps delivery. Forceps with manipulation. High forceps. Mid forceps
4 = Vacuum extraction ventouse.
5 = Breech delivery, spontaneous assisted or unspecified partial breech extraction
6 = Breech extraction. Breech extraction; NOS or Total or Version with breech extraction.
7 = Elective (planned) Caesarean Section.
8 = Emergency, other and unspecified Caesarean Section
9 = Other and unspecified method of delivery

Sterilisation after Delivery [15]
0 = None
1 = Laparoscopy
2 = Laparotomy
3 = Laparoscopy other hospital
4 = Laparotomy other hospital
8 = Other
9 = Not stated

Outcome of Pregnancy (Baby 1 and Baby 2) [23], [24]
1 = Live birth
2 = Still birth
3 = Live birth died < 7 days
4 = Live birth died 7-28 days
5 = Live birth died after 28 days

Sex (Baby 1 and Baby 2) [35], [36]
1 = Male
2 = Female
8 = Other or Not Known

Special Care Baby Unit (Baby 1 and Baby 2) [37], [38]
0 = Not Admitted
1 = Admitted for up to 48 hours
2 = Admitted for more than 48 hours
9 = Not known

Baby Discharged to (Baby 1 and Baby 2) [39], [40]
1 = Home
2 = Remaining in Special Care Baby Unit
3 = Special Care Baby Unit but home with mother
4 = Transfer to Other Hospital
5 = Other Unit in same hospital
6 = Foster Home
7 = Local Authority Care
8 = Healthy baby remaining in unit after mother's discharge
9 = Dead

SMR11 DOCUMENT

Kindly return completed computer record to Information Services Division, Edinburgh.

SMR11 (ABBREVIATED)

NEONATAL RECORD (Revised 1980) ....HOSPITAL

Card 1 — 1

1. Hospital Code No. — 2-6
2. Case Reference No. — 7-16
3. Maternal Reference No. — 17-26
4. Maternal Hospital No. — 27-31
5. Surname — 32-43

Infant Forenames....................................

Home Address....................................

........................Telephone No............

8. Ward........ Consultant or G.P .... — 46-51

Family Doctor....................................

Address....................................

........................Telephone No............

BIRTH RECORD

10. Place........................ 1. This Hospital 2. G.P. Hospital 3. Other Hospital 4. Home 5. Other 9. N/K — 53

E.D.D. .... Sure Not Sure N/K

11. Gestation (by dates)........................ — 54-55

12. Time Hrs. Date of Birth — 56-61

13./14. Apgar Score 1 min./5 min. (for 9 and 10 enter 9, N/K=0) — 62-63

17. Resuscitation
1. Nil (including clear airway, mask O_2)
2. Mask+IPPV
3. Intubation+IPPV (no drugs)
4. Intubation+IPPV (with drugs)
5. Drugs only Specify:
8. Other
9. N/K — 68

18. Birth Weight (g) — 69-72

Singleton/Twin/Triplet 1 12 123 — 73
Number:
Order — 74

21. Sex 1. Male 2. Female 9. N/K — 75

22. O.F.C cm (Third day) — 76-78

29. Transfers (2 only)
0. None
1. Special Care Nursery
2. Paediatric Medical
3. Paediatric Surgical
4. Cardio-thoracic unit
5. Neurol/Surgery
6. Other
9. N/K — 88, 89

MATERNAL RECORD

Health in Pregnancy Normal Abnormal

Specify....................................

....

Mode of Delivery
0. Vertex
1. Manipulation
2. Breech
3. Forceps Low
4. Forceps Other
5. Ventouse
6. Caesarean
8. Other
9. N/K

Local Option — 90

31. — 91-101, 102-112

Card 2 — 1

Hospital Code No. — DUPLICATE — 2-6

Case Reference No. — 7-16

32. Jaundice (bilirubin mg%; µmol/l) — 17
1. Absent
2. Mild (5-11.9; 86-204) or not measured
3. Moderate (12.0-19.9; 205-340)
4. Severe (20.0+, 342+)
9. N/K

Absent	Present	N/K
1	2	9

34. Significant Hypotonia — 19
37. Convulsions — 22
38. Recurrent Apnoea — 23
39. Assisted Ventilation after 30 mins. — 24
40. Feeding Difficulty (Tube Feeding) — 25

DISCHARGE RECORD

44. Condition 1. Normal 2. Doubtful 3. Cong. Abn. 4. Dead 5. Dead P.M. 6. Cong Abn. & Other 8. Other 9. N/K — 29

45. Discharge (Final)
1. Home with Mother
2. Home after Mother
3. To care of relative
4. Transfer to other Hospital (Medical)
5. Transfer to other Hospital (Social)
6. Transfer to Residential/Foster Care
7. Dead
8. Other
9. N/K — 30

46/47. Age (Days)/Weight — 31-36

DIAGNOSIS and PROCEDURES (Cardiff Code)

48. .... — V 3 — 37-42
49. 43-48
50. 49-54
51. 55-60
52. 61-66
53. 67-72
54. 73-78
55. 79-84
56. 85-90
57. 91-96

Operations (two only) 58. 97-100 59. 101-104

60. FOLLOW UP 1. None 2. This Hospital 3. Elsewhere 4. Multiple — 105

61. Local Option — 106-112

Comments

G S & Co Ltd. 53-2349 10/79

PLEASE NOTE: COMPLETION OF SHADED BOXES IS OPTIONAL

SMR11 DOCUMENT (REVERSE)

NEONATAL RECORD – SMR11 (ABBREVIATED)

Notes for completion of this form

General

1. This form should be completed for every newborn infant discharged or transferred from hospital. The top copy should be sent in monthly batches to Room B020, Information Services Division, Trinity Park House, South Trinity Road, Edinburgh EH5 3SQ. The second copy to the General Practitioner and the back copy is retained in the Hospital Case Records.
2. Use legible block capitals. A ball point pen should be used.
3. In those instances where the key to the code is not given on the form, you may write comments in the spaces provided in addition to coding.
4. The use of separate temperature and weight charts will be necessary.
5. Items which do not have numbered coding boxes will not be included on the computer record. Shaded boxes need not be filled.

Administrative and Clinical Data – Coding Instructions

Card 1

Box	Item	Instructions
1	Card No. 1	This item is pre-printed as "1"
2-6	Hospital code No.	Enter the 5-digit official code
7–16	Case Reference No.	Enter a unique 10-digit code in the range 1–9999999999. If the number is less than 10 digits the remaining boxes should be completed by leaving blank the spaces following the number. e.g. 1 2 3 4 5 6
17–26	Maternal Reference No.	(as above)
27-31	Maternal Hospital No	Enter the 5-digit official code
32-43	Surname	Enter the current surname, starting with the left-hand box. Apostrophes and hyphens should be placed in separate boxes. Names containing more than 12 letters should be entered as follows D U R H A M - R O B E R T S O N
46–51	Paediatric Consultant or G.P	Enter a 6-digit code in the range 000001-999999.
53	Place of Birth	Enter a single digit code in the range 1–5 If not known, enter "9".
54–55	Gestation	Enter the number of weeks (min - 20, max. -45) If not known, enter "99"
56–61	Date of Birth	Enter a 6-digit date, e.g. 010180 for 1st January 1980.
62-63	Apgar Scores (at 1 and 5 minutes)	Each box must contain a number in the range 0–9, where "0" means not done. For scores of 9 or 10 enter code "9".
68	Resuscitation	Enter a single digit in the range 1–5, 8 or 9.
69–72	Birth Weight	Enter the weight in grams, e g. '0980' (min. = 0500, max. = 9998). If not known enter 9999
73–74	Singleton/Twin/Triplet	Enter in box 73 if the child was a singleton (1), a twin (2) or a triplet (3) Enter in box 74 the order this child was born, e g the second child born in triplets = 32. Enter 99 in the boxes if more than triplets are born.
75	Sex	Enter a single digit code in the range 1-2. If not known, enter '9'.
76–78	O F C.	Enter the measurement in cm (preferably measured on third day of life). Min. - 10 0, max = 49.9. If not known enter 99 9".
88-89	Transfers	Enter single digit codes in the range 0–6 in each box If not known, enter "9" Both boxes should be completed
90–112	Local Options	Boxes 90–112 are for local use only. The code entered in each box should be in the range 0–9 Unused boxes should be left blank.

Card 2

Box	Item	Instructions
1	Card No. 2	This item is pre-printed as "2"
2–6	Board, District, Institution, Type	These items must be entered identically to the items entered card 1, boxes 2–16.
7-16	Case Reference No	(as above)
17	Jaundice	Enter a single digit code in the range 1–4 If not known, enter "9"
19-25	Findings – Any Age	Enter a single digit code in each box in the range 1–2. If not known, enter '9'.
29	Condition	Enter a single digit code in the range 1–6, 8. If not known, enter "9"
30	Discharge to	Enter a single digit code in the range 1–8 If other or not known, enter '9'.
31-32	Age (Days) on Discharge	Enter two digits representing the age in days. e.g 07 for 7 days. If the age is more than 99 days, enter 99'.
33–36	Weight on Discharge	Enter the weight in grams, e.g. '0980' (min. = 0500, max = 9998). If greater than 9998 gm. enter 9998 If not known, enter 9999
37–42) 43–48) 49–54) 55–60) 61–66) 67–72) 73–78) 79–84) 85–90) 91–96)	Diagnosis and Procedures	The first diagnosis, Classification of infant, Single twin, etc., must be completed, although the others may be left blank The codes to be used are in the range V30 to V39 with 0 1 and 2 used to indicate place of birth The V 3 have been preprinted, leaving the remaining two digits to be filled in.
97–100) 101–104)	Operations	Enter code in accordance with the Office of Population Censuses and Surveys code of operations in the range 0010 to 9999
105	Follow up	Enter a single digit code in the range 1–4 If not known enter 9"
106–112	Local Option	Boxes 106–112 are for local use only The code entered in each box should be in the range 0-9 Unused boxes should be left blank

List of Abbreviations

EDD	– Estimated date of delivery	Cong Abn	– Congenital Abnormality
IPPV	– Intermittent positive pressure ventilation	N/K	– Not known

APPENDIX 2

SCOTTISH BIRTHWEIGHT STANDARDS 1975-79

Taken from Smalls & Forbes (1983)

MALES PRIMIGRAVIDAE

Gestation Weeks	Centile 5	10	25	50	75	90	95
28	0.87	0.91	1.05	1.29	1.53	1.85	2.12
29	0.85	0.94	1.11	1.37	1.63	2.00	2.29
30	0.90	1.02	1.22	1.50	1.78	2.16	2.47
31	1.00	1.15	1.37	1.66	1.95	2.34	2.66
32	1.14	1.32	1.55	1.84	2.15	2.53	2.84
33	1.32	1.51	1.75	2.05	2.36	2.73	3.02
34	1.52	1.72	1.97	2.27	2.58	2.93	3.20
35	1.74	1.94	2.20	2.49	2.80	3.14	3.38
36	1.96	2.16	2.42	2.71	3.02	3.33	3.55
37	2.18	2.37	2.63	2.92	3.22	3.52	3.72
38	2.38	2.56	2.83	3.11	3.41	3.69	3.87
39	2.56	2.73	2.99	3.28	3.57	3.84	4.00
40	2.71	2.87	3.12	3.41	3.69	3.96	4.13
41	2.81	2.96	3.20	3.49	3.78	4.06	4.24
42	2.86	3.00	3.23	3.53	3.82	4.12	4.33

FEMALES PRIMIGRAVIDAE

Gestation Weeks	Centile 5	10	25	50	75	90	95
28	0.81	0.85	1.00	1.09	1.31	1.81	2.48
29	0.80	0.88	1.07	1.23	1.49	1.98	2.54
30	0.85	0.97	1.18	1.40	1.68	2.16	2.63
31	0.96	1.10	1.33	1.59	1.89	2.34	2.74
32	1.11	1.26	1.51	1.79	2.11	2.53	2.87
33	1.29	1.46	1.71	2.01	2.33	2.72	3.01
34	1.50	1.66	1.93	2.23	2.55	2.91	3.16
35	1.71	1.88	2.15	2.45	2.76	3.09	3.31
36	1.93	2.10	2.37	2.65	2.96	3.27	3.47
37	2.14	2.30	2.57	2.85	3.15	3.43	3.62
38	2.33	2.49	2.75	3.02	3.31	3.58	3.75
39	2.49	2.65	2.90	3.16	3.45	3.71	3.88
40	2.61	2.77	3.01	3.27	3.56	3.82	3.98
41	2.69	2.84	3.08	3.35	3.63	3.91	4.06
42	2.70	2.86	3.08	3.37	3.66	3.97	4.12

SCOTTISH BIRTHWEIGHT STANDARDS 1975-79 (contd.)

MALES MULTIPARAE

Gestation Weeks	Centile 5	10	25	50	75	90	95
28	0.97	1.05	1.12	1.34	1.47	1.98	2.55
29	0.91	1.03	1.17	1.38	1.58	2.09	2.60
30	0.93	1.08	1.28	1.49	1.74	2.23	2.69
31	1.02	1.19	1.43	1.64	1.93	2.40	2.81
32	1.16	1.35	1.61	1.84	2.15	2.59	2.96
33	1.35	1.54	1.82	2.06	2.38	2.80	3.12
34	1.57	1.76	2.05	2.30	2.62	3.01	3.29
35	1.80	2.00	2.28	2.55	2.87	3.23	3.47
36	2.04	2.23	2.51	2.80	3.10	3.44	3.66
37	2.27	2.46	2.73	3.03	3.33	3.64	3.83
38	2.48	2.66	2.92	3.24	3.53	3.82	4.00
39	2.66	2.83	3.09	3.41	3.70	3.98	4.15
40	2.79	2.96	3.21	3.53	3.83	4.10	4.28
41	2.87	3.02	3.29	3.59	3.91	4.19	4.38
42	2.87	3.02	3.30	3.59	3.94	4.24	4.45

FEMALES MULTIPARAE

Gestation Weeks	Centile 5	10	25	50	75	90	95
28	0.96	0.99	1.11	1.22	1.48	2.26	2.91
29	0.87	0.97	1.14	1.30	1.57	2.24	2.83
30	0.87	1.02	1.22	1.42	1.70	2.29	2.82
31	0.94	1.12	1.36	1.58	1.87	2.38	2.86
32	1.07	1.28	1.53	1.78	2.07	2.52	2.94
33	1.25	1.46	1.73	1.99	2.29	2.69	3.06
34	1.46	1.67	1.95	2.22	2.52	2.88	3.20
35	1.70	1.90	2.17	2.45	2.75	3.08	3.36
36	1.93	2.12	2.40	2.68	2.98	3.29	3.53
37	2.16	2.34	2.61	2.89	3.20	3.49	3.70
38	2.37	2.54	2.80	3.08	3.39	3.67	3.86
39	2.54	2.71	2.96	3.25	3.55	3.83	4.00
40	2.67	2.83	3.08	3.37	3.67	3.95	4.11
41	2.74	2.90	3.15	3.44	3.75	4.02	4.19
42	2.73	2.91	3.15	3.46	3.77	4.04	4.23

APPENDIX 3

McCULLAGH'S REGRESSION ALGORITHM (CONTD.)

```
      PROGRAM LOGIT
C MCCULLAGH (1980) MODEL WITH LOGISTIC LINK FUNCTION
C ESTIMATES PARAMETERS THETA AND BETA BY NEWTON-RAPHSON
C METHOD WITH FISHER SCORING. THETAS ARE REFERENCE
C LOGITS AND BETAS ARE REGRESSION COEFFICIENTS.
C CONVERGENCE WHEN TOTAL DIFFERENCE BETWEEN PARAMETER
C VALUES AT SUCCESSIVE ITERATIONS LESS THAN TOLERANCE.
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION IREAD(50),ICAT(20)
      COMMON /PARS/ NK,NP,NCASE
      COMMON /CHAN/ NCHIN,NCHOUT
      COMMON /RULES/ TOL,ICON
C ICAT - NUMBER OF LEVELS OF EACH COVARIATE
      DATA (ICAT(J),J=1,6) /2,3,2,3,4,4/
      CALL METERINIT
C NCHIN - INPUT CHANNEL
      NCHIN=1
C NCHOUT - OUTPUT CHANNEL
      NCHOUT=6
C NK - NUMBER OF LEVELS OF RESPONSE
      NK=8
C NCASE - NUMBER OF CASES TO BE READ IN
      NCASE=356
C NREAD - NUMBER OF COVARIATES
      NREAD=6
      WRITE(6,'(1X,''NO OF VARIABLES'')')
      READ(5,*) NVIN
C READ NVIN  NUMBER OF ITEMS IN ARRAY IREAD DESCRIBING THE
C THE COVARIATES, INTERACTIONS AND RESPONSE DEPENDENT
C TERMS IN THE MODEL
      WRITE(6,'(1X,''VARIABLES TO BE USED'')')
      IF(NVIN.EQ.0) GO TO 30
      READ(5,*) (IREAD(J),J=1,NVIN)
      DO 10 J=1,NVIN-1
      IF(IREAD(J).EQ.0.AND.IREAD(J+1).LE.0) GO TO 20
   10 CONTINUE
      IF(IREAD(NVIN).EQ.0) GO TO 20
      GO TO 30
   20 WRITE(NCHOUT,'(2X,''IREAD '',50(1X,I3))')
     +(IREAD(J),J=1,NVIN)
      WRITE(NCHOUT,'(2X,''INCORRECT MODEL'')')
      STOP
   30 CONTINUE
      IREAD(NVIN+1)=1000000
C ICONV - SET TO 1 ON CONVERGENCE
      ICONV=0
C TOL - TOLERANCE FOR TOTAL PARAMETER CHANGE AT CONVERGENCE
      TOL=.01
      CALL READAT(NREAD,IREAD,ICAT)
      CALL INIT(IREAD,ICAT)
      CALL ICL9CECPUTIME(X)
      WRITE(6,'(5X,''READ IN & INITIALISATION COMPLETED'',
     +'' , CPU = '',F10.2)') X
C NIT - NUMBER OF ITERATIONS
      NIT=0
  100 NIT=NIT+1
      WRITE(6,'(5X,''ITERATION '',I3)') NIT
      CALL ARHS
      CALL ICL9CECPUTIME(X)
      WRITE(6,'(5X,''A AND RHS COMPUTED , CPU = '',F10.2)') X
      CALL SOLVE
      CALL ICL9CECPUTIME(X)
      WRITE(6,'(5X,''EQUATIONS SOLVED , CPU = '',F10.2)') X
      CALL CONV(ICONV)
      IF(ICONV.EQ.0) GO TO 100
      CALL OUTPUT(NIT,NVIN,IREAD)
      STOP
      END
      SUBROUTINE READAT(NREAD,IREAD,ICAT)
C READ DATA AND CREATE DUMMY VARIABLES FOR SPECIFIED MODEL
```

```
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION IREAD(50),ICAT(20)
      DIMENSION IX(1299,59),Z(1299,10),N(1299),ITEM1(20),ITEM2(10)
      COMMON /PARS/ NK,NP,NCASE
      COMMON /DATA/ IX,Z,N
      COMMON /CHAN/ NCHIN,NCHOUT
      NP=0
      IR=0
C CALCULATE NP - NUMBER OF REGRESSION PARAMETERS IN MODEL
    1 CONTINUE
      IR=IR+1
      IF(IREAD(IR).EQ.1000000) GO TO 3
      IADD=ICAT(IREAD(IR))-1
      IMULT=1
      IF(IREAD(IR+1).NE.0) GO TO 4
      IMULT=ICAT(IREAD(IR+2))
      IR=IR+2
    4 IADD=IADD*IMULT
      IF(IREAD(IR+1).LT.0) GO TO 2
      NP=NP+IADD
      GO TO 1
    2 IR=IR+1
      NP=NP+IADD
      IF(IREAD(IR+1).GE.0) GO TO 1
      GO TO 2
    3 CONTINUE
      NC=NCASE
      NCASE=0
      DO 100 I=1,NC
      READ(NCHIN,71) NTEMP,(ITEM1(J),J=1,NREAD),
     +(ITEM2(J),J=1,NK)
   71 FORMAT(5X,I6,1X,6I1,10(1X,I5))
C OMIT CASES WITH MISSING DATA
      DO 30 J=1,NREAD
      IF(ITEM1(J).GT.ICAT(J)) GO TO 100
   30 CONTINUE
      NCASE=NCASE+1
C CREATE CUMULATIVE RELATIVE FREQUENCIES FOR EACH CASE IN Z
      N(NCASE)=NTEMP
      ZN=0
      DO 10 J=1,NK-1
      Z(NCASE,J)=ZN+FLOAT(ITEM2(J))
   10 ZN=Z(NCASE,J)
      DO 11 J=1,NK-1
   11 Z(NCASE,J)=Z(NCASE,J)/FLOAT(NTEMP)
      Z(NCASE,NK)=1.
C CREATE DUMMY VARIABLES IN IX
      IR=0
      IWR=1
   20 CONTINUE
      IR=IR+1
      IVAR=IREAD(IR)
      IF(IVAR.EQ.1000000) GO TO 25
      JNEST=1
      IF(IREAD(IR+1).NE.0) GO TO 26
      IVAR2=IREAD(IR+2)
      JNEST=ICAT(IVAR2)
      IR=IR+2
   26 IF(IREAD(IR+1).LT.0) GO TO 22
C IF A COVARIATE IS NESTED CREATE SEPARATE SETS OF DUMMYS
C FOR EACH LEVEL OF STRATIFYING VARIABLE
      DO 21 J2=1,JNEST
      IX2=1
      IF(JNEST.EQ.1) GO TO 27
      IF(ITEM1(IVAR2).NE.J2) IX2=0
   27 DO 21 J=2,ICAT(IVAR)
      IX(NCASE,IWR)=0
      IF(ITEM1(IVAR).EQ.J.AND.IX2.EQ.1) IX(NCASE,IWR)=1
   21 IWR=IWR+1
      GO TO 20
```

```
C IF RESPONSE DEPENDENT TERMS ARE INCLUDED DUMMYS ARE SET TO
C -1*(LEVEL NUMBER OF RESPONSE AT WHICH TERMS ENTER MODEL)
   22 IR=IR+1
      DO 24 J2=1,JNEST
      IX2=1
      IF(JNEST.EQ.1) GO TO 28
      IF(ITEM1(IVAR2).NE.J2) IX2=0
   28 DO 24 J=2,ICAT(IVAR)
      IX(NCASE,IWR)=0
      IF(ITEM1(IVAR).EQ.J.AND.IX2.EQ.1) IX(NCASE,IWR)=IREAD(IR)
   24 IWR=IWR+1
      IF(IREAD(IR+1).LT.0) GO TO 22
      GO TO 20
   25 CONTINUE
      IF(I.GT.6) GO TO 100
  100 CONTINUE
      RETURN
      END
      SUBROUTINE INIT(IREAD,ICAT)
C SET INITIAL VALUES FOR PARAMETERS
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION IX(1299,59),Z(1299,10),N(1299)
      DIMENSION T1(20,5),T2(20,5),W(10),PI(10)
      DIMENSION TH(10),B(59),G(10),Y(10),VIN(10),DY(10)
      DIMENSION IREAD(50),ICAT(20)
      COMMON /DATA/ IX,Z,N
      COMMON /GAMYV/ G,Y,VIN,DY
      COMMON /THEBE/ TH,B
      COMMON /PARS/ NK,NP,NCASE
C INITIALISE MATRICES
      DO 10 I=1,20
      DO 10 J=1,5
      T1(I,J)=0
   10 T2(I,J)=0
      DO 15 J=1,10
      PI(J)=0
   15 Y(J)=0
      RTOT=0
C CALCULATE POPULATION FREQUENCIES
      DO 16 I=1,NCASE
      RN=FLOAT(N(I))
      RTOT=RTOT+RN
      ZO=0
      DO 16 J=1,NK
      PI(J)=PI(J)+RN*(Z(I,J)-ZO)
   16 ZO=Z(I,J)
      WTOT=0
      TEMP=0
C CALCULATE POPULATION WEIGHTS
      DO 17 J=1,NK-1
      TEMP=TEMP+PI(J)
      W(J)=TEMP*(RTOT-TEMP)*(PI(J)+PI(J+1))
   17 WTOT=WTOT+W(J)
      DO 18 J=1,NK-1
   18 W(J)=W(J)/WTOT
      DO 100 I=1,NCASE
      RN=FLOAT(N(I))
      TEMP=0
      DO 20 J=1,NK-1
C CALCULATE WEIGHTED EMPIRICAL LOGIT FOR EACH CASE
      X=(Z(I,J)*RN+.5)/(RN-RN*Z(I,J)+.5)
      RLIN=DLOG(X)
      Y(J)=Y(J)+Z(I,J)*RN
      TEMP=TEMP+RLIN*W(J)
   20 CONTINUE
      IR=0
      IWR=1
   30 CONTINUE
      IR=IR+1
      IVAR=IREAD(IR)
```

```
      IF(IVAR.EQ.1000000) GO TO 100
      JNEST=1
      IF(IREAD(IR+1).NE.0) GO TO 40
      IR=IR+2
      JNEST=ICAT(IREAD(IR))
   40 IREF=1
C ADD WEIGHTED LOGIT TO APPROPRIATE CATEGORY TOTAL FOR EACH
C COVARIATE IN MODEL, IGNORING INTERACTIONS AND RESPONSE
C DEPENDENT EFFECTS
      DO 50 J=1,JNEST
      DO 50 K=2,ICAT(IVAR)
      IF(IX(I,IWR).EQ.0) GO TO 50
      T1(IVAR,K)=T1(IVAR,K)+RN*TEMP
      T2(IVAR,K)=T2(IVAR,K)+RN
      IREF=0
   50 IWR=IWR+1
      IF(IREF.NE.1) GO TO 60
      T1(IVAR,1)=T1(IVAR,1)+RN*TEMP
      T2(IVAR,1)=T2(IVAR,1)+RN
   60 IF(IREAD(IR+1).GT.0) GO TO 30
      IR=IR+1
   70 IF(IREAD(IR+1).GT.0) GO TO 30
      IWR=IWR+(ICAT(IVAR)-1)*JNEST
      IR=IR+1
      GO TO 70
  100 CONTINUE
C SET INITIAL THETA VALUES TO POPULATION LOGITS
      DO 110 K=1,NK-1
      TH(K)=(Y(K)+.5)/(RTOT-Y(K)+.5)
  110 TH(K)=DLOG(TH(K))
      IR=0
      IWR=1
  130 CONTINUE
      IR=IR+1
      IVAR=IREAD(IR)
      IF(IVAR.EQ.1000000) GO TO 150
      JNEST=1
      IF(IREAD(IR+1).NE.0) GO TO 180
      JNEST=ICAT(IREAD(IR+2))
      IR=IR+2
C SET INITIAL BETA VALUES TO TOTAL WEIGHTED LOGITS IN
C EACH CATEGORY MINUS WEIGHTED LOGITS IN REFERENCE
C CATEGORY FOR EACH COVARIATE, INITIAL PARAMETER VALUES
C EQUAL FOR EACH LEVEL OF STRATIFYINGC COVARIATE IN
C AN INTERACTION
  180 BREF=T1(IVAR,1)/T2(IVAR,1)
      DO 120 J2=1,JNEST
      DO 120 J=2,ICAT(IVAR)
      B(IWR)=T1(IVAR,J)/T2(IVAR,J)-BREF
  120 IWR=IWR+1
      IF(IREAD(IR+1).GE.0) GO TO 130
      IR=IR+1
C INITIAL BETA VALUES EQUAL AT DIFFERENT RESPONSE LEVELS
C FOR RESPONSE DEPENDENT EFFECTS
  170 IF(IREAD(IR+1).GE.0) GO TO 130
      DO 160 J2=1,JNEST
      DO 160 J=2,ICAT(IVAR)
      B(IWR)=T1(IVAR,J)/T2(IVAR,J)-BREF
  160 IWR=IWR+1
      IR=IR+1
      GO TO 170
  150 CONTINUE
      WRITE(6,'(2X,''THETAS'')')
      WRITE(6,'(5(2X,F10.6))') (TH(K),K=1,NK-1)
      WRITE(6,'(2X,''BETAS'')')
      WRITE(6,'(5(2X,F10.6))') (B(K),K=1,NP)
      G(NK)=1.
      DY(NK)=DEL(G(NK))
      RETURN
      END
```

```
      SUBROUTINE ARHS
C CALCULATE MATRIX OF EXPECTED SECOND DERIVATIVES AND
C RIGHT HAND SIDE OF NEWTON-RAPHSON EQUATION.
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION QJR(69),A(69,69),RHS(69),DY(10)
      DIMENSION IX(1299,59),Z(1299,10),N(1299),G(10),Y(10),VIN(10)
      COMMON /DATA/ IX,Z,N
      COMMON /GAMYV/ G,Y,VIN,DY
      COMMON /SIMUL/ A,RHS
      COMMON /PARS/NK,NP,NCASE
C INITIALISE MATRICES
      DO 10 I=1,NK+NP-1
      RHS(I)=0
      DO 10 J=1,NK+NP-1
   10 A(I,J)=0
      DO 500 I=1,NCASE
      RN=FLOAT(N(I))
      CALL GYVCAL(I)
C CALCULATE QJR CORRESPONDING TO THETAS
      DO 100 J=1,NK-1
      DO 20 K=1,NK-1
      QJR(K)=0
      IF(K.NE.J) GO TO 21
      QJR(K)=DY(J)
   21 IF(K.NE.J+1) GO TO 20
      QJR(K)=-G(J)*DY(J+1)/G(J+1)
   20 CONTINUE
C CALCULATE QJR CORRESPONDING TO BETAS
      DO 30 K=1,NP
      TEMP=0
      IF(IX(I,K).GT.0.OR.-IX(I,K).EQ.J) TEMP=1.
      TEMP1=0
      IF(IX(I,K).GT.0.OR.-IX(I,K).EQ.J+1) TEMP1=1.
      QJR(NK-1+K)=DY(J)*TEMP-(G(J)/G(J+1))*DY(J+1)*TEMP1
   30 CONTINUE
      QJ=Y(J)*DY(J)-(G(J)/G(J+1))*DY(J+1)*Y(J+1)
      DO 80 IR=1,NK+NP-1
C RIGHT HAND SIDE IN VECTOR RHS
      ADD=RN*VIN(J)*QJR(IR)*(QJ+Z(I,J)-G(J)*Z(I,J+1)/G(J+1))
      RHS(IR)=RHS(IR)+ADD
      DO 80 IS=1,IR
C EXPECTED SECOND DERIVATIVES IN BOTTOM TRIANGLE OF MATRIX A
      A(IR,IS)=A(IR,IS)+RN*VIN(J)*QJR(IR)*QJR(IS)
   80 CONTINUE
  100 CONTINUE
  500 CONTINUE
      RETURN
      END
      SUBROUTINE GYVCAL(I)
C CALCULATE Y - LINEAR COMPONENT OF MODEL, G - INVERSE
C LINK FUNCTION OF LINEAR COMPONENT, VIN - RATE OF CHANGE OF
C PHI WITH GAMMA, DY - RATE OF CHANGE OF GAMMA WITH Y.
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION TH(10),B(59),IX(1299,59),Z(1299,10),N(1299)
      DIMENSION G(10),Y(10),VIN(10),DY(10)
      COMMON /DATA/ IX,Z,N
      COMMON /GAMYV/ G,Y,VIN,DY
      COMMON /THEBE/ TH,B
      COMMON /PARS/ NK,NP,NCASE
      DO 100 J=1,NK-1
      X=TH(J)
      DO 10 K=1,NP
      TEMP=0
      IF(IX(I,K).GT.0.OR.-IX(I,K).EQ.J) TEMP=1.
   10 X=X+B(K)*TEMP
      G(J)=RINLIN(X)
      Y(J)=X
  100 CONTINUE
      DO 200 J=1,NK-1
      VIN(J)=G(J+1)/(G(J)*(G(J+1)-G(J)))
```

```
      DY(J)=DEL(G(J))
  200 CONTINUE
      RETURN
      END
      FUNCTION RLINK(X)
C CALCULATES LINK FUNCTION
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      RLINK=DLOG(X/(1.-X))
      RETURN
      END
      FUNCTION DEL(X)
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DEL=X*(1.-X)
      RETURN
      END
      FUNCTION RINLIN(X)
C CALCULATES INVERSE LINK FUNCTION
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      RINLIN=DEXP(X)
      RINLIN=RINLIN/(1.+RINLIN)
      RETURN
      END
      SUBROUTINE SOLVE
C SOLVE NETWON-RAPHSON EQUATION AX(N+1)=RHS
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION A(69,69),RHS(69),P(69)
      COMMON /SIMUL/ A,RHS
      COMMON /PARS/ NK,NP,NCASE
      N=NK-1+NP
      IFAIL=0
      CALL F01BXF(N,A,69,P,IFAIL)
      IF(IFAIL.EQ.0) GO TO 10
      WRITE(6,'(2X,''F01BXF , IFAIL = '',I1)') IFAIL
      STOP
   10 CALL F04AZF(N,1,A,69,P,RHS,69,IFAIL)
      IF(IFAIL.EQ.0) GO TO 20
      WRITE(6,'(2X,''F04AZF , IFAIL = '',I1)') IFAIL
      STOP
   20 RETURN
      END
      SUBROUTINE CONV(ICONV)
C SET NEW WORKING VALUES FOR THETA AND BETA AND CALCULATE
C CONVERGENCE CRITERION, TOTAL DIFFERENCE BETWEEN NEW AND
C OLD PARAMETER VALUES
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION A(69,69),RHS(69),TH(10),B(59)
      COMMON /SIMUL/ A,RHS
      COMMON /THEBE/ TH,B
      COMMON /PARS/ NK,NP,NCASE
      COMMON /RULES/ TOL,ICON
      DIFF=0
C THETA VALUES
      DO 10 J=1,NK-1
      DTEMP=TH(J)-RHS(J)
      IF(DTEMP.LT.0) DTEMP=-DTEMP
      IF(DTEMP.GT.DIFF) DIFF=DTEMP
      TH(J)=RHS(J)
   10 CONTINUE
C BETA VALUES
      DO 20 J=1,NP
      DTEMP=B(J)-RHS(NK-1+J)
      IF(DTEMP.LT.0) DTEMP=-DTEMP
      IF(DTEMP.GT.DIFF) DIFF=DTEMP
      B(J)=RHS(NK-1+J)
   20 CONTINUE
      IF(DIFF.LT.TOL) ICONV=1
      WRITE(6,'(2X,''THETAS'')')
      WRITE(6,'(5(2X,F10.6))') (TH(K),K=1,NK-1)
      WRITE(6,'(2X,''BETAS'')')
      WRITE(6,'(5(2X,F10.6))') (B(K),K=1,NP)
```

```
      WRITE(6,'(2X,''MAXIMUM PARAMETER CHANGE = '',F10.7)') DIFF
      RETURN
      END
      SUBROUTINE LOGLIK(RLL,RDEV)
C CALCULATES LOG LIKELIHOOD AND CASES WITH MAXIMUM AND MINIMUM
C LINEAR COMPONENTS
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION IX(1299,59),Z(1299,10),N(1299)
      DIMENSION G(10),Y(10),VIN(10),DY(10),AVG(10)
      COMMON /DATA/ IX,Z,N
      COMMON /MAXMIN/ AVG,IMAX,YMAX,IMIN,YMIN
      COMMON /GAMYV/ G,Y,VIN,DY
      COMMON /PARS/ NK,NP,NCASE
      RLL=0
      RDEV=0
      YMIN=10000.
      YMAX=-100000.
      DO 20 J=1,NK-1
   20 AVG(J)=0
      DO 100 I=1,NCASE
      CALL GYVCAL(I)
C MAXIMUM AND MINIMUM CASES
      DO 30 J=1,NK-1
   30 AVG(J)=AVG(J)+G(J)
      IF(Y(1).LE.YMAX) GO TO 40
      YMAX=Y(1)
      IMAX=I
   40 IF(Y(1).GE.YMIN) GO TO 50
      YMIN=Y(1)
      IMIN=I
   50 CONTINUE
      T=0
      F=0
C CALCULATES INCREMENTS TO THE LOG LIKELIHOOD
      DO 10 J=1,NK-1
      T=T+Z(I,J)*DLOG(G(J)/(G(J+1)-G(J)))
      T=T-Z(I,J+1)*DLOG(G(J+1)/(G(J+1)-G(J)))
      IF(Z(I,J).EQ.Z(I,J+1)) GO TO 10
      IF(Z(I,J).EQ.0.0) GO TO 15
      F=F+Z(I,J)*DLOG(Z(I,J)/(Z(I,J+1)-Z(I,J)))
   15 F=F-Z(I,J+1)*DLOG(Z(I,J+1)/(Z(I,J+1)-Z(I,J)))
   10 CONTINUE
      RDEV=RDEV+FLOAT(N(I))*F
      RLLINC=FLOAT(N(I))*T
      RLL=RLL+RLLINC
      RLLINC=-2.*RLLINC
  100 CONTINUE
      RLL=-2.*RLL
      RDEV=RLL+2.*RDEV
      DO 60 J=1,NK-1
   60 AVG(J)=AVG(J)/NCASE
      RETURN
      END
      SUBROUTINE INVERT
C CALCULATES INFORMATION MATRIX AT SOLUTION
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION A(69,69),RHS(69),B(69,69)
      COMMON /SIMUL/ A,RHS
      COMMON /PARS/ NK,NP,NCASE
      CALL ARHS
      N=NK-1+NP
      IFAIL=0
      DO 10 IR=1,NK+NP-2
      DO 10 IS=IR+1,NK+NP-1
   10 A(IR,IS)=A(IS,IR)
C MATRIX INVERSION
      CALL F01ACF(N,X02AFF(IT),A,69,B,69,RHS,NCORR,IFAIL)
      IF(IFAIL.EQ.0) GO TO 20
      WRITE(6,'(2X,''F01ACF , IFAIL = '',I1)') IFAIL
      STOP
```

```
   20 CONTINUE
      DO 30 IR=1,NK+NP-1
      DO 30 IS=1,IR
   30 A(IS,IR)=A(IR+1,IS)
      RETURN
      END
      SUBROUTINE OUTPUT(NIT,NVIN,IREAD)
C OUTPUT PARAMETERS AND STATISTICS
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION IX(1299,59),Z(1299,10),N(1299),IREAD(50)
      DIMENSION B(59),TH(10),A(69,69),RHS(69),AVG(10)
      COMMON /THEBE/ TH,B
      COMMON /DATA/ IX,Z,N
      COMMON /MAXMIN/ AVG,IMAX,YMAX,IMIN,YMIN
      COMMON /SIMUL/ A,RHS
      COMMON /PARS/ NK,NP,NCASE
      COMMON /CHAN/ NCHIN,NCHOUT
      WRITE(NCHOUT,'(5X,''NUMBER OF CASES IN ANALYSIS '',I8)')
     +NCASE
      WRITE(NCHOUT,'(5X,''NUMBER OF RESPONSE LEVELS '',I3)') NK
      WRITE(NCHOUT,'(5X,''NUMBER OF PARAMETERS '',I6)') NP
      WRITE(NCHOUT,'(5X,''MODEL''//5X,50(1X,I3))')
     +(IREAD(J),J=1,NVIN)
      WRITE(NCHOUT,'(2X,''CONVERGENCE AFTER ITERATION '',I2)') NIT
      CALL LOGLIK(RLL,RDEV)
      WRITE(NCHOUT,70) RLL,RDEV
   70 FORMAT(2X,'LOG LIKELIHOOD RATIO = ',G12.7,
     +//,2X,'RESIDUAL DEVIANCE = ',G12.7)
      CALL INVERT
C CALCULATES AND WRITES OUT PARAMETERS AND CONFIDENCE INTERVALS
      DO 10 J=1,NK-1+NP
      IF(J.EQ.1) WRITE(NCHOUT,72)
      IF(J.EQ.NK) WRITE(NCHOUT,73)
      IF(J.LE.NK-1) PAR=TH(J)
      IF(J.GT.NK-1) PAR=B(J-NK+1)
      SD=1.96*SQRT(A(J,J))
      CL=PAR-SD
      CU=PAR+SD
      IF(J.GT.NK-1) GO TO 20
      ETH=RINLIN(PAR)
      ECL=RINLIN(CL)
      ECU=RINLIN(CU)
      GO TO 10
   20 ETH=DEXP(PAR)
      ECL=DEXP(CL)
      ECU=DEXP(CU)
   10 WRITE(NCHOUT,71) PAR,CL,CU,ETH,ECL,ECU
C WRITES OUT AVERAGE, MAXIMUM AND MINIMUM RISK CASES
      WRITE(NCHOUT,74) (AVG(J),J=1,NK-1)
      XMAX=YMAX-TH(1)
      WRITE(NCHOUT,75) IMAX,XMAX,N(IMAX),(IX(IMAX,J),J=1,NP)
      XMIN=YMIN-TH(1)
      WRITE(NCHOUT,76) IMIN,XMIN,N(IMIN),(IX(IMIN,J),J=1,NP)
      CALL ICL9CEDATE(DATE)
      CALL ICL9CECPUTIME(X)
      WRITE(NCHOUT,'(30X,A8,5X,''CPU = '',F10.2)') DATE,X
      RETURN
   71 FORMAT(10X,F6.3,2X,'(',F6.3,',',F6.3,')',
     +5X,F6.3,'(',F6.3,',',F6.3,')')
   72 FORMAT(2X,'THETAS',2X,2X,'LOGISTIC SCALE',7X,
     +5X,2X,'INVERSE LINK SCALE')
   73 FORMAT(2X,'BETAS',3X,2X,'LOGISTIC SCALE',7X,
     +5X,2X,'EXPONENTIAL SCALE')
   74 FORMAT(2X,'CASE AVERAGE GAMMAS '//2X,20(F6.4,2X))
   75 FORMAT(/2X,'MAXIMUM RISK CASE = ',I6,5X,'LINEAR RISK = ',
     +F7.4,5X,'FREQUENCY = ',I4//2X,39(I2,1X))
   76 FORMAT(/2X,'MINIMUM RISK CASE = ',I6,5X,'LINEAR RISK = ',
     +F7.4,5X,'FREQUENCY = ',I4//2X,39(I2,1X))
      END
```

APPENDIX 4

PRENTICE & GLOECKLER'S PROPORTIONAL HAZARDS ALGORITHM

```
      PROGRAM PHAZ
C PRENTICE & GLOECKLER'S (1978) MODEL FOR GROUPED SURVIVAL DATA.
C ESTIMATES PARAMETERS GAMMA AND BETA BY NEWTON-RAPHSON. GAMMAS
C ARE REFERENCE LOG-LOG(SURVIVAL PROBABILITIES IN EACH GROUP),
C BETAS ARE REGRESSION COEFFICIENTS. CONVERGENCE WHEN EUCLIDEAN
C DISTANCE BETWEEN SUCCESSIVE PARAMETER VALUES LESS
C THAN TOLERANCE.  THIS VERSION CAN HANDLE TIME
C DEPENDENT VARIABLES
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION IREAD(20)
      COMMON /PARS/ NVAR,NWEEK,NCASE,TOL
      COMMON /CHAN/ NCHIN,NCHOUT
      COMMON /READIN/ NREAD,IREAD
      CALL METERINIT
      WRITE(6,'(1X,''NO OF VARIABLES'')')
C NVAR - NUMBER OF PARAMETERS IN ANALYSIS
      READ(5,*) NVAR
C NWEEK - NUMBER OF SURVIVAL CATEGORIES
      NWEEK=10
C NCASE - NUMBER OF DATA CASES
      NCASE=1521
C NREAD - NUMBER OF PARAMETERS IN MODEL
      NREAD=16
      WRITE(6,'(1X,''VARIABLES TO BE USED'')')
C IREAD - ARRAY CONTAINING COVARIATES ON DATA FILE IN MODEL
      READ(5,*) (IREAD(J),J=1,NVAR)
C TOL - TOLERANCE FOR EUCLIDEAN DISTANCE ON CONVERGENCE
      TOL=.01
C NCHIN - INPUT CHANNEL
      NCHIN=8
C NCHOUT - OUTPUT CHANNEL
      NCHOUT=13
      CALL MAIN
      STOP
      END
      SUBROUTINE MAIN
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION S11(20),S12(20,20),S22(20,20),T11(20)
      DIMENSION B(20),G(20)
      COMMON /PARIT/ B,G
      COMMON /PARS/ NVAR,NWEEK,NCASE,TOL
      COMMON /INFI/ S11,S12,S22,T11
C NIT - NUMBER OF ITERATIONS
      NIT=1
C ICON - SET TO 1 ON CONVERGENCE
      ICON=0
      CALL READAT
      CALL METER
      WRITE(6,'(5X,'' DATA READ IN'')')
      CALL START
      CALL METER
      WRITE(6,'(5X,'' PARAMATERS INITIALISED'')')
C NEWTON RAPHSON ITERATION BEGINS
  100 CONTINUE
      WRITE(6,'(5X,'' ITERATION '',I3)') NIT
      CALL SCORE
      CALL METER
      WRITE(6,'(5X,'' SCORE STATISTIC CALCULATED'')')
      CALL INVERT
      CALL METER
      WRITE(6,'(5X,
     +'' INFORMATION CALCULATED AND INVERTED'')')
      CALL UPDATE
      CALL METER
      WRITE(6,'(5X,'' PARAMETER VALUES UPDATED'')')
      CALL CONVGE(ICON,DIFF)
      IF(ICON.EQ.1) GO TO 200
      NIT=NIT+1
      GO TO 100
  200 CALL OUTPUT(NIT,DIFF)
```

```
      RETURN
      END
      SUBROUTINE SCORE
C CALCULATES CURRENT SCORE
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION SCG(20),SCB(20),B(20),G(20),IREAD(20)
      DIMENSION IX(1700,20),IGES(1700,20),H(20)
      COMMON /PARIT/ B,G
      COMMON /PARS/NVAR,NWEEK,NCASE,TOL
      COMMON /SCOR/ SCG,SCB
      COMMON /READIN/ NREAD,IREAD
      COMMON /COVAR/ IX
      COMMON /WEEKS/ IGES
      FUNB(X)=X*DEXP(-X)/(1.-DEXP(-X))
C INITIALISE ARRAYS
      DO 5 I=1,20
    5 SCG(I)=0
      DO 6 I=1,20
    6 SCB(I)=0
      DO 500 I=1,NCASE
      DO 100 K=1,NWEEK
      HK=0
      BHK=0
      ZB=0
      IF(NVAR.EQ.0) GO TO 8
      DO 7 J=1,NVAR
      Z=IX(I,J)
      IF(IX(I,J).LT.0) Z=IX(I,J+IX(I,J))*K**(-IX(I,J))
    7 ZB=ZB+B(J)*Z
    8 IF(K.EQ.NWEEK) GO TO 30
      H(K)=DEXP(G(K)+ZB)
      HK=H(K)
      BHK=FUNB(HK)
   30 CONTINUE
C CALCULATE SCORE CORRESPONDING TO GAMMAS
      DO 50 J=1,NWEEK-1
      IF(J-K) 10,20,50
   10 SCG(J)=SCG(J)-H(J)*IGES(I,K)
      GO TO 50
   20 SCG(J)=SCG(J)+BHK*IGES(I,K)
   50 CONTINUE
C CALCULATE SCORE CORRESPONDING TO BETAS
      IF(NVAR.EQ.0) GO TO 100
      DO 60 J=1,NVAR
      HSUM=0
      DO 55 K1=1,K-1
      Z=IX(I,J)
      IF(IX(I,J).LT.0) Z=IX(I,J+IX(I,J))*K1**(-IX(I,J))
   55 HSUM=HSUM+Z*H(K1)
      Z=IX(I,J)
      IF(IX(I,J).LT.0) Z=IX(I,J+IX(I,J))*K**(-IX(I,J))
      SCB(J)=SCB(J)+IGES(I,K)*(Z*BHK-HSUM)
   60 CONTINUE
  100 CONTINUE
  500 CONTINUE
      WRITE(6,'(5X,''SCORE STATISTIC'')')
      WRITE(6,'(2X,''GAMMA'')')
      WRITE(6,*) (SCG(I),I=1,K)
      IF(NVAR.EQ.0) RETURN
      WRITE(6,'(2X,''BETA'')')
      WRITE(6,*) (SCB(I),I=1,NVAR)
      RETURN
      END
      SUBROUTINE INVERT
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION S11(20),S12(20,20),S22(20,20),T11(20)
      DIMENSION W(20,20)
      COMMON /INFI/ S11,S12,S22,T11
      COMMON /PARS/ NVAR,NWEEK,NCASE,TOL
      CALL CALCI
```

```
      CALL METER
      WRITE(6,'(5X,'' MATRIX S CALCULATED'')')
      IF(NVAR.EQ.0) GO TO 55
C CALCULATE E=S22-S21*INV(S11)*S12
      DO 5 I=1,NVAR
      DO 5 J=I,NVAR
      DO 5 K=1,NWEEK-1
    5 S22(I,J)=S22(I,J)-S12(K,I)*S12(K,J)/S11(K)
      IF(NVAR.GT.1) GO TO 6
      S22(2,1)=1./S22(1,1)
      GO TO 10
    6 EPS=X02AAF(X)
      IFAIL=0
C INVERT LOWER DIAGONAL MATRIX
      CALL F01ACF(NVAR,EPS,S22,20,W,20,T11,L,IFAIL)
      IF(IFAIL.EQ.0) GO TO 10
      WRITE(6,'(5X,''IFAIL = '',I1)') IFAIL
   10 CONTINUE
      CALL METER
      WRITE(6,'(5X,'' MATRIX INVERTED'')')
C CALCULATE F
      DO 20 I=1,NWEEK-1
      DO 20 J=1,NVAR
   20 S12(I,J)=S12(I,J)/S11(I)
C COPY INVERSE INTO UPPER TRIANGULAR S22
      DO 30 I=1,NVAR
      DO 30 J=I,NVAR
   30 S22(I,J)=S22(J+1,I)
      DO 40 I=2,NVAR
      DO 40 J=1,I-1
   40 S22(I,J)=S22(J,I)
C CALCULATE INV(S11)+F*INV(E)*FT IN LOWER TRIANGLE S22
      DO 50 I=1,NWEEK-1
      DO 50 J=1,NVAR
      W(I,J)=0
      DO 50 K=1,NVAR
   50 W(I,J)=W(I,J)+S12(I,K)*S22(K,J)
   55 CONTINUE
      DO 60 I=1,NWEEK-1
      DO 60 J=1,I
      S22(I+1,J)=0
      IF(I.EQ.J) S22(I+1,J)=1./S11(I)
      IF(NVAR.EQ.0) GO TO 60
      DO 65 K=1,NVAR
   65 S22(I+1,J)=S22(I+1,J)+W(I,K)*S12(J,K)
   60 CONTINUE
      IF(NVAR.EQ.0) RETURN
C MOVE -F*INV(E) FROM W TO S12
      DO 70 I=1,NWEEK-1
      DO 70 J=1,NVAR
   70 S12(I,J)=-W(I,J)
      RETURN
      END
      SUBROUTINE UPDATE
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION B(20),G(20),SCG(20),SCB(20),BLAST(20),GLAST(20)
      DIMENSION S11(20),S12(20,20),S22(20,20),T11(20)
      COMMON /PARIT/ B,G
      COMMON /PARS/ NVAR,NWEEK,NCASE,TOL
      COMMON /INFI/ S11,S12,S22,T11
      COMMON /SCOR/ SCG,SCB
      COMMON /LAST/ BLAST,GLAST
C UPDATE GAMMA ESTIMATES
      DO 50 I=1,NWEEK-1
      G(I)=GLAST(I)
      DO 10 K=1,NWEEK-1
      IF(I.GE.K) G(I)=G(I)+S22(I+1,K)*SCG(K)
      IF(I.LT.K) G(I)=G(I)+S22(K+1,I)*SCG(K)
   10 CONTINUE
      IF(NVAR.EQ.0) GO TO 50
```

```
      DO 20 K=1,NVAR
   20 G(I)=G(I)+S12(I,K)*SCB(K)
   50 CONTINUE
      IF(NVAR.EQ.0) GO TO 200
C UPDATE BETA ESTIMATES
      DO 100 I=1,NVAR
      B(I)=BLAST(I)
      DO 60 K=1,NWEEK-1
   60 B(I)=B(I)+S12(K,I)*SCG(K)
      DO 70 K=1,NVAR
      IF(I.LE.K) B(I)=B(I)+S22(I,K)*SCB(K)
      IF(I.GT.K) B(I)=B(I)+S22(K,I)*SCB(K)
   70 CONTINUE
  100 CONTINUE
  200 CONTINUE
      WRITE(6,'(5X,'' NEW ESTIMATES'')')
      WRITE(6,*) (G(I),I=1,NWEEK-1)
      IF(NVAR.EQ.0) RETURN
      WRITE(6,*) (B(I),I=1,NVAR)
      RETURN
      END
      SUBROUTINE CONVGE(ICON,DIFF)
C CALCULATE EUCLIDEAN DISTANCE BETWEEN SUCCESSIVE
C PARAMETER VALUES AND RESET OLD VALUES.
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION B(20),G(20),BLAST(20),GLAST(20)
      COMMON /PARIT/ B,G
      COMMON /PARS/ NVAR,NWEEK,NCASE,TOL
      COMMON /LAST/ BLAST,GLAST
      DIFF=0
   10 DO 20 I=1,NWEEK-1
      DIFF=DIFF+(G(I)-GLAST(I))**2
   20 GLAST(I)=G(I)
      IF(NVAR.EQ.0) GO TO 50
      DO 30 I=1,NVAR
      DIFF=DIFF+(B(I)-BLAST(I))**2
   30 BLAST(I)=B(I)
   50 CONTINUE
      DIFF=SQRT(DIFF)
      WRITE(6,'(5X,''DIFF = '',F10.6)') DIFF
C TEST FOR CONVERGENCE
  500 IF(DIFF.LT.TOL) ICON=1
      RETURN
      END
      SUBROUTINE CALCI
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION S11(20),S12(20,20),S22(20,20),T11(20),IREAD(20)
      DIMENSION IGES(1700,20),IX(1700,20),B(20),G(20),H(20)
      COMMON /PARIT/ B,G
      COMMON /INFI/ S11,S12,S22,T11
      COMMON /PARS/ NVAR,NWEEK,NCASE,TOL
      COMMON /READIN/ NREAD,IREAD
      COMMON /WEEKS/ IGES
      COMMON /COVAR/ IX
      FUNB(X)=X*DEXP(-X)/(1.-DEXP(-X))
      DO 10 J1=1,NWEEK-1
      S11(J1)=0
      IF(NVAR.EQ.0) GO TO 10
      DO 5 J2=1,NVAR
    5 S12(J1,J2)=0
   10 CONTINUE
      IF(NVAR.EQ.0) GO TO 15
      DO 20 J1=1,NVAR
      DO 20 J2=J1,NVAR
   20 S22(J1,J2)=0
   15 CONTINUE
      DO 500 I=1,NCASE
      DO 25 J1=1,NWEEK-1
   25 T11(J1)=0
      DO 500 K=1,NWEEK
```

```
      ZB=0
      IF(K.EQ.NWEEK) GO TO 22
      IF(NVAR.EQ.0) GO TO 23
      DO 21 J=1,NVAR
      Z=IX(I,J)
      IF(IX(I,J).LT.0) Z=IX(I,J+IX(I,J))*K**(-IX(I,J))
   21 ZB=ZB+B(J)*Z
   23 H(K)=DEXP(G(K)+ZB)
   22 CONTINUE
C CALCULATE S11 INCREMENT FOR INDIVIDUALS (I,K)
      HK=0
      DIK=0
      IF(K.EQ.NWEEK) GO TO 50
      HK=H(K)
      E=DEXP(-HK)
      DIK=FUNB(HK)*(E+HK-1.)/(1.-E)
   50 DO 60 J=1,NWEEK-1
      IF(J-K) 30,40,60
   30 T11(J)=H(J)
      S11(J)=T11(J)*IGES(I,K)+S11(J)
      GO TO 60
   40 T11(J)=DIK
      S11(J)=T11(J)*IGES(I,K)+S11(J)
   60 CONTINUE
      IF(NVAR.EQ.0) GO TO 90
C CALCULATE S12 INCREMENT FOR INDIVIDUALS (I,K)
      DO 70 J1=1,NWEEK-1
      DO 70 J2=1,NVAR
      Z=IX(I,J2)
      IF(IX(I,J2).LT.0) Z=IX(I,J2+IX(I,J2))*J1**(-IX(I,J2))
   70 S12(J1,J2)=S12(J1,J2)+Z*T11(J1)*IGES(I,K)
C CALCULATE S22 INCREMENT FOR INDIVIDUALS (I,K)
      DO 80 J1=1,NVAR
      DO 80 J2=J1,NVAR
      HSUM=0
      DO 85 J=1,K-1
      Z1=IX(I,J1)
      IF(IX(I,J1).LT.0) Z1=IX(I,J1+IX(I,J1))*J**(-IX(I,J1))
      Z2=IX(I,J2)
      IF(IX(I,J2).LT.0) Z2=IX(I,J2+IX(I,J2))*J**(-IX(I,J2))
   85 HSUM=HSUM+Z1*Z2*H(J)
      Z1=IX(I,J1)
      IF(IX(I,J1).LT.0) Z1=IX(I,J1+IX(I,J1))*K**(-IX(I,J1))
      Z2=IX(I,J2)
      IF(IX(I,J2).LT.0) Z2=IX(I,J2+IX(I,J2))*K**(-IX(I,J2))
   80 S22(J1,J2)=S22(J1,J2)+IGES(I,K)*(Z1*Z2*DIK+HSUM)
   90 CONTINUE
  500 CONTINUE
      RETURN
      END
      SUBROUTINE READAT
C READ DATA AND CREATE COVARIATES FOR THE PROGRAM.
C LINEAR (AND HIGHER ORDER) DEPENDENCY ON SURVIVAL CATEGORY
C INCORPORATED BY CREATING COVARIATES WITH NEGATIVE VALUES.
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION IX(1700,20),IGES(1700,20),IREAD(20),ITEM(20)
      COMMON /PARS/ NVAR,NWEEK,NCASE,TOL
      COMMON /COVAR/ IX
      COMMON /WEEKS/ IGES
      COMMON /CHAN/ NCHIN,NCHOUT
      COMMON /READIN/ NREAD,IREAD
      DO 100 I=1,NCASE
      READ(NCHIN,71) (ITEM(J),J=1,NREAD),
     +(IGES(I,J),J=1,NWEEK)
   71 FORMAT(2X,16I1,17(1X,I5))
      DO 10 J=1,NVAR
      IF(IREAD(J).GT.0) IX(I,J)=ITEM(IREAD(J))
      IF(IREAD(J).LT.0) IX(I,J)=IREAD(J)
   10 CONTINUE
  100 CONTINUE
```

```
      RETURN
      END
      SUBROUTINE START
C STARTING PARAMETER VALUES
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION B(20),G(20),BLAST(20),GLAST(20)
      DIMENSION IGES(1700,20),TEMP(20)
      COMMON /PARIT/ B,G
      COMMON /PARS/ NVAR,NWEEK,NCASE,TOL
      COMMON /LAST/ BLAST,GLAST
      COMMON /WEEKS/ IGES
      RTOT=0
C SET STARTING VALUES FOR GAMMA TO POPULATION VALUES
      DO 30 J=1,NWEEK
   30 TEMP(J)=0
      DO 10 I=1,NCASE
      DO 10 J=1,NWEEK
      TEMP(J)=TEMP(J)+FLOAT(IGES(I,J))
   10 RTOT=RTOT+FLOAT(IGES(I,J))
      DO 20 J=1,NWEEK-1
      GLAST(J)=DLOG(DLOG(RTOT/(RTOT-TEMP(J))))
      G(J)=GLAST(J)
   20 RTOT=RTOT-TEMP(J)
      IF(NVAR.EQ.0) RETURN
C SET STARTING VALUES FOR BETA TO ZERO
      DO 60 J=1,NVAR
      B(J)=0
   60 BLAST(J)=0
      RETURN
      END
      SUBROUTINE OUTPUT(NIT,DIFF)
C WRITES OUTPUT STATISTICS AND JOB INFORMATION
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION S11(20),S12(20,20),S22(20,20),T11(20)
      DIMENSION B(20),G(20),TEM(20),IREAD(20)
      CHARACTER*8 DATE
      COMMON /PARIT/B,G
      COMMON /PARS/ NVAR,NWEEK,NCASE,TOL
      COMMON /INFI/ S11,S12,S22,T11
      COMMON /CHAN/ NCHIN,NCHOUT
      COMMON /READIN/ NREAD,IREAD
      CALL ICL9CEDATE(DATE)
      CALL ICL9CECPUTIME(X)
      WRITE(NCHOUT,'(30X,A8,5X,''CPU = '',F8.0)') DATE,X
      WRITE(NCHOUT,71) NVAR,NWEEK,NCASE,TOL
   71 FORMAT(/5X,'£ OF COVARIATES = ',I3,
     +/5X,'£ OF WEEKS = ',I3,
     +/5X,'£ OF CASES = ',I5,
     +/5X,'TOLERENCE = ',G11.4)
      WRITE(NCHOUT,'(/5X,''IREAD= '',20(:I2,'','','')))')
     +(IREAD(J),J=1,NVAR)
      WRITE(NCHOUT,72) NIT,DIFF
   72 FORMAT(//10X,'CONVERGENCE AFTER ITERATION ',I5,
     +/15X,'ACTUAL LL DIFFERENCE = ',G11.4)
      CALL LOGLIK(RLL)
      RLL=-2.*RLL
      NOPAR=NWEEK-1+NVAR
      WRITE(NCHOUT,77) RLL,NOPAR
   77 FORMAT(/5X,'-2*LOG-LIKELIHOOD = ',G14.7,2X,'WITH ',I2,' D.F')
      WRITE(NCHOUT,'(//5X,''PARAMETER ESTIMATES'')')
      WRITE(NCHOUT,'(/6X,''GAMMAS'',7X,''STANDEV'',7X,''ALPHA'')')
C CALCULATES AND WRITES OUT CONFIDENCE INTERVALS FOR GAMMA
      DO 10 I=1,NWEEK-1
      TALP=DEXP(-DEXP(G(I)))
      SD=SQRT(S22(I+1,I))
      C1=DEXP(-DEXP(G(I)-1.96*SD))
      C2=DEXP(-DEXP(G(I)+1.96*SD))
   10 WRITE(NCHOUT,73) G(I),SD,TALP,C1,C2
   73 FORMAT(/5X,3(G11.4,2X),2X,'(',G11.4,'-',G11.4,')')
      IF(NVAR.EQ.0) GO TO 12
```

```
      WRITE(NCHOUT,'(/6X,''BETAS '',7X,''STANDEV'',
     +6X,''EXP(BETA)'')')
C CALCULATES AND WRITES OUT CONFIDENCE INTERVALS FOR BETA
      DO 11 I=1,NVAR
      SD=SQRT(S22(I,I))
      EXB=DEXP(B(I))
      C1=DEXP(B(I)-1.96*SD)
      C2=DEXP(B(I)+1.96*SD)
   11 WRITE(NCHOUT,73) B(I),SD,EXB,C1,C2
   12 WRITE(NCHOUT,'(/5X,''CORRELATION MATRIX OF GAMMAS'')')
C WRITES OUT CORRELATION MATRIX FOR GAMMAS
      ITEM1=1
   15 CONTINUE
      LIM=ITEM1+4
      IF(LIM.GT.NWEEK-1) LIM=NWEEK-1
      WRITE(NCHOUT,76) (I,I=ITEM1,LIM)
      DO 50 I=ITEM1,NWEEK-1
      K=1
      LIM=ITEM1+4
      IF(LIM.GT.I) LIM=I
      DO 55 J=ITEM1,LIM
      TEM(K)=S22(I+1,J)/SQRT(S22(I+1,I)*S22(J+1,J))
   55 K=K+1
   50 WRITE(NCHOUT,75) I,(TEM(J),J=1,LIM-ITEM1+1)
      IF(LIM.EQ.NWEEK-1) GO TO 20
      ITEM1=ITEM1+5
      GO TO 15
   20 CONTINUE
      IF(NVAR.EQ.0) GO TO 100
      WRITE(NCHOUT,'(/5X,''CORRELATION MATRIX OF BETAS'')')
C WRITES OUT CORRELATION MATRIX FOR BETAS
      ITEM1=1
   30 CONTINUE
      LIM=ITEM1+4
      IF(LIM.GT.NVAR) LIM=NVAR
      WRITE(NCHOUT,76) (I,I=ITEM1,LIM)
      DO 60 I=ITEM1,NVAR
      K=1
      LIM=ITEM1+4
      IF(LIM.GT.I) LIM=I
      DO 65 J=ITEM1,LIM
      TEM(K)=S22(J,I)/SQRT(S22(I,I)*S22(J,J))
   65 K=K+1
   60 WRITE(NCHOUT,75) I,(TEM(J),J=1,LIM-ITEM1+1)
      IF(LIM.EQ.NVAR) GO TO 80
      ITEM1=ITEM1+5
      GO TO 30
   80 CONTINUE
      WRITE(NCHOUT,
     +'(/5X,''CORRELATION MATRIX OF GAMMAS & BETAS'')')
C WRITES OUT CORRELATION MATRIX BETWEEN GAMMA AND BETA
      ITEM1=1
   90 CONTINUE
      LIM=ITEM1+4
      IF(LIM.GT.NVAR) LIM=NVAR
      WRITE(NCHOUT,76) (I,I=ITEM1,LIM)
      DO 110 I=1,NWEEK-1
      K=1
      DO 115 J=ITEM1,LIM
      TEM(K)=S12(I,J)/SQRT(S22(I+1,I)*S22(J,J))
  115 K=K+1
  110 WRITE(NCHOUT,75) I,(TEM(J),J=1,LIM-ITEM1+1)
      IF(LIM.EQ.NVAR) GO TO 100
      ITEM1=ITEM1+5
      GO TO 90
  100 CONTINUE
   75 FORMAT(2X,I2,2X,5(1X,G11.4))
   76 FORMAT(/6X,5(3X,I2,7X))
      STOP
      END
```

```
      SUBROUTINE LOGLIK(RLLR)
C CALCULATES LOG LIKELIHOOD OF SOLUTION
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION B(20),G(20),IX(1700,20),IGES(1700,20),IREAD(20)
      COMMON /PARIT/ B,G
      COMMON /PARS/ NVAR,NWEEK,NCASE,TOL
      COMMON /READIN/ NREAD,IREAD
      COMMON /COVAR/ IX
      COMMON /WEEKS/ IGES
      RLLR=0
      DO 500 I=1,NCASE
      HSUM=0
      DO 40 K=1,NWEEK
      ZB=0
      RLINC=0
      IF(NVAR.EQ.0) GO TO 20
      DO 10 J=1,NVAR
      Z=IX(I,J)
      IF(IX(I,J).LT.0) Z=IX(I,J+IX(I,J))*K**(-IX(I,J))
   10 ZB=ZB+B(J)*Z
   20 IF(K.EQ.NWEEK) GO TO 30
      HK=DEXP(G(K)+ZB)
      RLINC=DLOG(1.-DEXP(-HK))
   30 RLINC=RLINC-HSUM
      RLLR=RLLR+RLINC*IGES(I,K)
   40 HSUM=HSUM+HK
  500 CONTINUE
      RETURN
      END
```

REFERENCES

Aitkin M & Clayton D (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. Applied Statistics; 29: 156-163.

Aitkin M, Anderson D & Hinde J (1981). Statistical modelling of data on teaching styles. Journal of the Royal Statistical Society, A; 144: 419-461.

Alberman E, Roman E, Pharoah POD & Chamberlain G (1980). Birthweight before and after a spontaneous abortion. British Journal of Obstetrics and Gynaecology; 87: 275-280.

Alberman E (1981). Sociobiological factors and birth weight in Great Britain. Preterm Labor. Edited by Elder MG & Hendricks CH. Butterworths International Medical Reviews; Obstetrics and Gynaecology 1: London.

Anderson JA (1984). Regression and ordered categorical variables. Journal of the Royal Statistical Society B; 46: 1-30.

Apgar V (1953). A proposal for a new method of evaluation of the newborn infant. Anesthesia and Analgesia; 32: 260.

Baird D (1974). The epidemiology of low birth weight: changes in incidence in Aberdeen, 1948-72. Journal of Biosocial Science; 6: 323-341.

Baird D (1985). Changing problems and priorities in obstetrics. British Journal of Obstetrics and Gynaecology; 92: 115-121.

Baker RJ & Nelder JA (1978). The GLIM system, release 3. Oxford: Numerical Algorithms Group.

Bakketeig LS, Seigel DG & Sternhal PM (1978). A fetal-infant life table based on single births in Norway, 1967-73. American Journal of Epidemiology; 10: 216-225.

Bakketeig LS & Hoffman HJ (1979). Perinatal mortality by birth order within cohorts based on sibship size. British Medical Journal; 2: 693-696.

Bakketeig LS, Hoffman HJ & Harley EE (1979). The tendency to repeat gestational age and birth weight in successive births. American Journal of Obstetrics and Gynecology; 135: 1086-1103.

Bakketeig LS & Hoffman HJ (1981). Epidemiology of preterm birth: results from a longitudinal study of births in Norway. Preterm Labor. Edited by Elder MG & Hendricks CH. Butterworths International Medical Reviews 1: London.

Bakketeig LS & Hoffman HJ (1983). The tendency to repeat gestational age and birth weight in successive births, related to perinatal survival. Acta Obstetrica Gynecologica Scandinavia; 62: 385-392.

Bartholomew DJ (1980). In discussion of McCullagh P (1980). Regression models for ordinal data. Journal of the Royal Statistical Society; 42: 109-142.

van den Berg BJ & Oechsli FW (1984). Prematurity. Perinatal Epidemiology. Edited by Bracken MB. Oxford University Press.

Billewicz WZ (1973). Some implications of self-selection for pregnancy. British Journal of Preventive and Social Medecine; 27: 49-52.

Bjerkedal T & Erickson JD (1983). Association of birth outcome with subsequent fertility. American Journal of Obstetrics and Gynecology; 147: 399-414.

Brenn T & Arnesen E (1985). Selecting risk factors: a comparison of discriminant analysis, logistic regression and Cox's regression model using data from the Tromso heart study. Statistics in Medicine; 4: 413-423.

Buerkins P, Delvoye P, Wollast E & Robyn C (1984). Epidemiology of pregnancies with unknown last menstrual period. Journal of Epidemiology and Community Health; 38: 79-80.

Butler NR & Alberman ED (1969). Perinatal Problems. E & S Livingstone; Edinburgh.

Carr-Hill RA & Pritchard C (1985). The Development and Exploitation of Empirical Birthweight Standards. Stockton Press; Macmillan: London.

Chamberlain R, Chamberlain G, Howlett B & Claireux A (1975). British Births 1970 Vol I: The First Week of Life. William Heinemann Medical Books Ltd, London.

Chiswick ML (1985). Intrauterine growth retardation. British Medical Journal; 291: 845-848.

Clayton D (1985). Using test-retest reliability data to improve estimates of relative risk: and application of latent class analysis. Statistics in Medicine; 4: 445-455.

Cole SK (1980). Scottish maternity and neonatal records. Perinatal Audit and Surveillance. Edited by Chalmers I & McIlwaine G. Royal College of Obstetrics and Gynaecologists: London.

Cole SK (1983). Evaluation of a neonatal discharge record as a monitor of congenital malformations. Community Medicine; 5: 21-30.

Common Services Agency for the Scottish Health Service, Information Services Division (1975-82). Scottish Health Statistic. HMSO: London.

Copas JB (1983). Plotting p against x. Applied Statistics; 32: 25-31.

Cox DR (1970). The Analysis of Binary Data. Methuen: London.

Cox DR (1972). Regression models and life tables. Journal of the Royal Statistical Society B; 34: 187-203.

Fienberg SE & Gong DG (1984). In discussion to Landwehr JM, Pregibon D & Shoemaker AC. Graphical methods for assessing logistic regression models. Journal of the American Statistical Society; 79: 61-83.

Forbes JF, Boddy FA, Pickering RM & Wiley MM (1982). Perinatal mortality in Scotland: 1970-9. Journal of Epidemiology and Community; 36: 282-288.

Forbes JF & McKellar D (1985). Personnel Communication.

Fortney JA & Whithorne EW (1982). The development of an index of high-risk pregnancy. American Journal of Obstetrics and Gynecology; 143: 501-508.

Frank PI, Kay CR, Lewis TLT & Parish S (1985). Outcome of pregnancy following induced abortion. Report from the joint study of the Royal College of General Practictioners and the Royal College of Obstetricians and Gynaecologists. British Journal of Obstetrics and Gynaecology; 92: 308-316.

French FE & Bierman JM (1962). Probabilities of fetal mortality. Public Health Reports; 77: 835-847.

Golding J, Butler NR & Newcombe RG (1983). Analysis of completed reproductive histories: a cautionary tale. Journal of Epidemiology and Community Health; 37: 78-81.

Hall MH & Carr-Hill R (1982). Impact of sex ratio on onset and the management of labour. British Medical Journal; 301: 677-403.

Hall MH, Carr-Hill RA, Fraser C, Campbell D & Samphier ML (1985a). The extent and antecedents of uncertain gestation. British Journal of Obstetrics and Gynaecology; 92: 445-451.

Hall MH & Carr-Hill RA (1985b). The significance of uncertain gestation for obstetric outcome. British Journal of Obstetrics and Gynaecology; 92: 452-460.

Harlop S, Shiono PH, Ramcharan S, Berendes H & Pellegrin F (1979). A prospective study of spontaneous fetal losses after induced abortion. New England Journal of Medicine; 301: 677-681.

Harris JE (1982). Prenatal medical care and infant mortality. Economic Aspects of Health. Edited by Fuchs VR. University of Chicago Press.

Hastie TJ (1984). In discussion to Landwehr JM, Pregibon D & Shoemaker AC. Graphical methods for assessing logistic regssion models. Journal of the American Statistical Society; 79: 61-83.

Hellier JL & Goldstein H (1979). The use of birthweight and gestation to assess perinatal mortality risk. Journal of Epidemiology and Community Health; 33: 183-185.

Hoffman HJ, Stark CR, Lundin FEJr & Ashbrook JD (1974). Analysis of birth weight, gestational age, and fetal viability, U.S. Births 1968. Obstetrical and Gynecological Reviews; 29: 651-681.

Hogue CJR, Cates WJr & Tietze C (1982). The effects of induced abortion on subsequent reproduction. Epidemiologic Review; 4: 66-94.

Holford TR (1980). The analysis of rates and of survivorship using log-linear models. Biometrics; 36: 299-305.

Illsley R & Mitchell RG (1984). The developing concept of low birth weight and the present state of knowledge. Low Birth Weight. A Medical, Pyschological, and Social Study. Edited by Illsley R & Mitchell RG. Wiley & Sons: Chichester.

Joint Committee of the Royal College of Obstetricians and Gynaecologists and the Population Investigation Committee (1948). Maternity in Great Britain. Oxford University Press; London.

Kalbfleisch JD & Prentice RL (1980). The statistical Analysis of Failure Time Data. J Wiley & Sons: New York.

Kaldor J & Clayton D (1985). Latent class analysis in chronic disease epidemiology. Statistics in Medicine; 4: 327-335.

Kay R & Little S (1986). Assessing the fit of the logistic model: a case study of children with the haemolytic uraemic syndrome. Applied Statistics; 35: 16-30.

Laird N & Olivier D (1981). Covariance analysis of censored survival data using log-linear analysis techniques. Journal of the American Statistical Association; 76: 234-240.

Landwehr JM, Pregibon D & Shoemaker AC (1984). Graphical methods for assessing logistic regression models. Journal of the American Statistical Association; 79: 61-83.

Lazarsfeld PF & Henry NW (1962). Latent Structure Analysis. Houghton Miffin: New York.

Leridon H (1976). Facts and artifacts in the study of intra-uterine mortality: a reconstruction from pregnancy histories. Population Studies; 30: 319-355.

Macintyre S (1978). Some notes on record taking and making in an antenatal clinic. Sociological Reviews; 26: 595-611.

Mackenzie IZ & Hillier K (1977). Prostaglandin-induced abortion and outcome of subsequent pregnancies: a prospective controlles study. British Medical Journal; 2: 1114-1117.

McCullagh P (1980). Regression models for ordinal data. Journal of the Royal Statistical Society B; 42: 109-142.

McCullagh P & Nelder JA (1983). Generalized Linear Models. Monographs on statistics and Applied Probability; Chapman and Hall: London.

Meirik O & Bergstrom R (1983). Outcome of delivery subsequent to induced vacuum-aspiration abortion in nulliparous women. Acta Obstetrica Gynecologica Scandinavia; 62: 449-509.

Mellin GW (1962). Fetal life tables. Journal of the American Medical Association; 180: 11-14.

Miller KR, Sabagh G & Dingman NW (1962). Latent structure analysis and differential mortality. Journal of the American Statistical Association; 57: 430-438.

Naeye RL & Dixon JB (1978). Distortions in fetal growth standards. Pediatric Research; 12: 987-991.

Nelder JA & Wedderburn RWM (1972). Generalized linear models. Journal of the Royal Statistical Society A 135: 370-384.

Neligan GA, Kolvin I, Scott DMcL & Garside RF (1976). Born Too Soon Born Too Small. Spastics International Medical Books; Heinemann: London.

Newcombe RG & Chalmers IG (1981). Assessing the risk of preterm labour. Preterm Labor. Edited by Elder MG & Hendrick CH. Butterworths International Medical Reviews; Obstetrics and Gynecology: London.

Ounsted M, Moar VA & Scott A (1985). Risk factors associated with small-for-gestational-age infants. British Journal of Obstetrics and Gynaecology; 92: 226-232.

Paneth N, Wallenstein, Kiely JL & Susser M (1982). Social class indicators and mortality in low birth weight infants. American Journal of Epidemiology; 116: 364-375.

Peters TJ, Golding J, Butler NR, Fryer JG, Lawrence CJ & Chamberlain GVP (1983). Plus ca change: predictors of birthweight in two national studies. British Journal of Obstetrics and Gynaecology; 90: 1040-1045.

Pickering RM & Forbes (1984). A classification of Scottish infants using latent class analysis. Statistics in Medicine; 3: 249-259.

Pickering RM & Forbes JF (1985). Risk of pre-term delivery and small for gestational age infants following abortion: a population study. British Journal of Obstetrics and Gynaecology; 92: 1106-1112.

Pickering RM, Murray GD & Forbes JF (1986). Pre-term foetal life time in Scotland. Population Studies; 40: 115-127.

Pickering RM (1987a). Maternal characteristics and the distribution of birthweight standardised for gestational age. Journal of Biosocial Science.

Pickering RM (1987b). Relative risks of low birthweight in Scotland : 1980-82. Social Paediatric and Obstetric Research Unit: University of Glasgow.

Pregibon D (1980). Goodness of link test for generalized linear models. Applied Statistics; 29: 15-24.

Pregibon D (1981). Logistic regression diagnostics. The Annals of Statistics; 9: 705-724.

Prentice RL & Gloeckler LA (1978). Regression analysis of grouped survival data with application to breast cancer data. Biometrics; 34: 57-67.

Prentice RL & Pyke R (1979). Logistic disease incidence models and case-control studies. Biometrics; 66: 403-411.

Registrar General Scotland (1983). Annual Report. HMSO: Edinburgh.

Resseguie LJ (1973). Influence of age, birth order and reproductive compensation on stillbirth ratios. Journal of Biosocial Sciences; 5: 443-452.

Resseguie LJ (1974). Pregnancy wastage and age of mother amongst the Amish. Human Biology; 46: 633-639.

Resseguie LJ (1976). Comparison of longitudinal and cross-sectional analysis: maternal age and stillbirth ratio. American Journal of Epidemiology; 103: 551-559.

Resseguie LJ (1977). The artifactual nature of effects of maternal age on risk of stillbirth. Journal of Biosocial Science; 9: 191-200.

Richardson JA & Dixon G (1976). Effects of legal termination on subsequent pregnancy. British Medical Journal; 1: 1303-1304.

Rindskopf D & Rindskopf W (1986). The value of latent class analysis in medical diagnosis. Statistics in Medicine; 5: 21-27.

Roman E, Doyle P, Beral V, Alberman E & Pharoah P (1978). Fetal loss, gravidity, and pregnancy order. Early Human Development; 2: 131-138.

Roman E (1984). Fetal loss rates and their relation to pregnancy order. Journal of Epidemiology and Community Health; 38: 29-35.

Rubin DB (1976). Inference and missing data. Biometrika; 63: 581-592.

Scott A, Moar V & Ounsted M (1981). The relative contributions of different maternal factors in small-for-gestational-age pregnancies. European Journal of Obstetrics, Gynecology and Reproductive Biology; 12: 157-165.

Scott A, Moar V & Ounsted M (1982). The relative contributions of different maternal factors in large-for-gestational-age pregnancies. European Journal of Obstetrics, Gynecology and Reproductive Biology; 13: 269-277.

Shapiro S, Jones EW and Densen PM (1962). A life table of pregnancy terminations and correlates of fetal loss. Milbank Memorial Fund Quaterly; 40: 7-45.

Skene AM (1978). Discrimination using latent structure models. Compstat 78. Proceedings in Computational Statistics. Physica-Verlag, Vienna.

Smalls M & Forbes JF (1983). Centile Values of Birthweight for Gestational Age in Scottish Infants. Social Paediatric and Obstetric Research Unit: University of Glasgow.

Smalls M, McKellar D & Pickering RM (1987). Linking neonatal and maternity discharge records. Community Medicine.

Strobino BR, Kline J, Shrout P, Susser M & Warburton D (1980). Recurrent abortion: definition of a sydrome. Human Embryonic and Fetal Death. Edited by Hook EB & Parker IH. Academic Press; New York.

Terris AM & Glasser M (1974). A life table analysis of the relation of prenatal care to prematurity. American Journal of Public Health; 64: 869-875.

Thompson R & Baker RJ (1981). Composite link functions in generalized linear models. Applied Statistics; 30: 125-131.

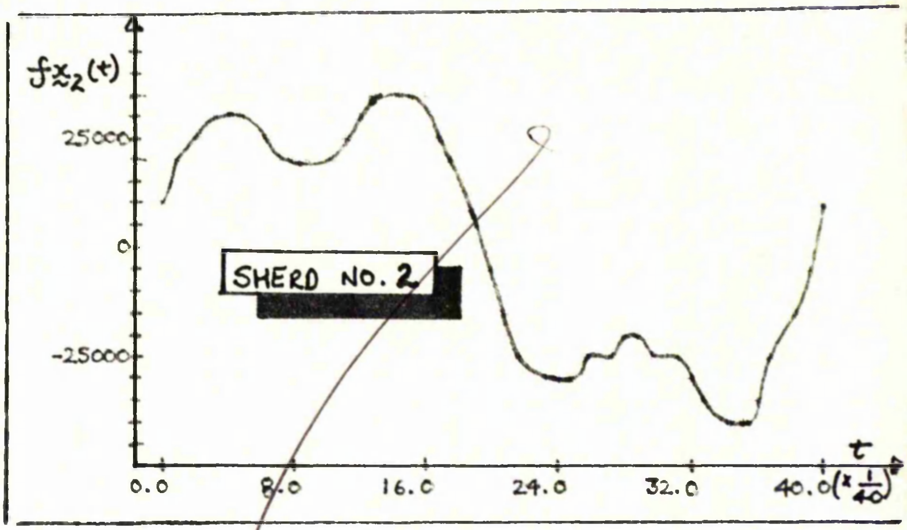
Thomson AM, Billewicz WZ & Hytten FE (1968). The assessment of fetal growth. Journal of Obstetrics and Gynaecology of the British Commonwealth; 75: 903-916.

Wilcox AJ (1981). Birth weight, gestation, and the fetal growth curve. American Journal of Obstetrics and Gynecology; 139: 863-867.

Wilcox AJ & Gladen BC (1982). Spontaneous abortion: the role of heterogeneous risk and selective fertility. Early Human Development; 7: 165-178.

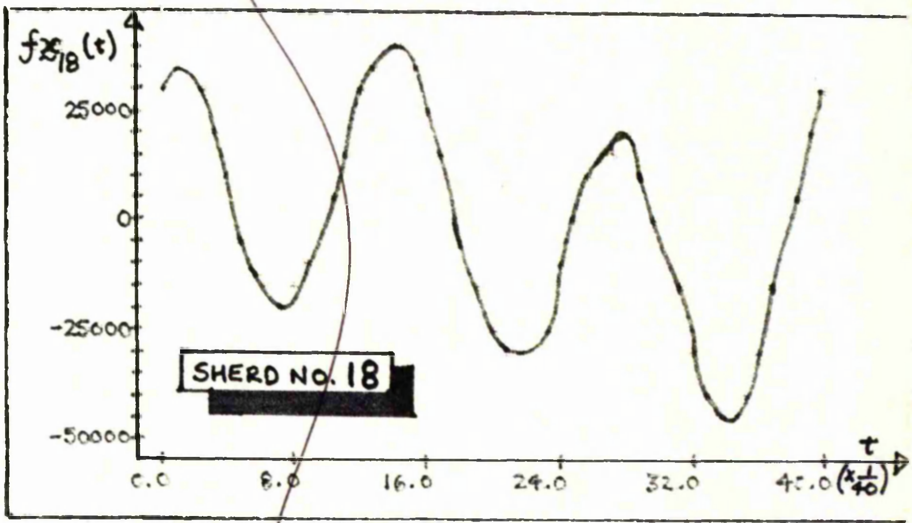
World Health Organization Task Force on the Sequelae of Abortion. Gestation, birthweight and spontaneous abortion in pregnancy after induced abortion (1979). Lancet; 1: 142-145.

Yudkin PL & Baras M (1983). A new approach to assessing the effect of birth order on outcome of pregnancy. Journal of Biosocial Science; 15: 307-316.



GRAPH 13

The Andrews Curve for sherd 2



GRAPH 14

The Andrews Curve for sherd 18

References

Bartholomew 1984 JRSSB 46 120-123

1980 42 293-321

McCullagh 42 109-142

889 8251
Ext 300
1hr

Steve McCarthy, Whittingham Gdns
Dept of Med.

Bashmore Sample 100
sources stress / strain
meths coping
other variable relationships

stress instrument = objective
70 items - ratings of how
stressful 0-3
0, 1, 2, 3

5 types
control, management, relations
+? +?

factor structure ?

Nursing study similar work
Gray-Toft + Anderson

Questions of interest : group comparisons
eg protected v remand v convicted

Possible project
Bartholomew - Latent Variables

Said - **bottom** line need to believe
that the raw score value 0, 1, 2, 3
is interval - not just ordinal
but probably scope for modifying
~~raw~~ method
- shld have referenced Bartholomew
but did reference Pickering