CHAPTER 6

6. SOURCES OF DATA AND METHODS OF ANALYSIS

6.1 INTRODUCTION

This chapter discusses the sources of data and methods used in this study. The chapter is divided into three sections. The first section describes the sources of data. The second and third sections describe the procedures and methods used in the study. The main statistical method used was the binary logistic regression.

6.2 SOURCES OF DATA

This section provides a brief description of the data used in the study. The Demographic and Health Survey III (DHS) data sets for Kazakhstan (1999), Uzbekistan (1996) and Kyrgyzstan (1997) were the main sources of data used. For a detailed description of these surveys, the interested reader should consult the DHS reports of the three countries. The purpose of the surveys was to develop single integrated sets of data for the governments of the respective republics to use in the planning of effective policies and programmes in the areas of health and nutrition. Permission was sought from and granted by Macro International Inc. to conduct a secondary analysis of selected variables for the purposes of this study. Selected variables included, among others, breastfeeding practises, nutrition, vaccination coverage and episodes of illness among children under the age of three years, anaemia levels, and weight and height measurements for both the children and the mothers.

The DHS data sets provide one of the best sources of data for the study of child malnutrition worldwide. In fact, the DHS provides far more detailed data on child malnutrition than any relevant data available from other sources. This is particularly true for developing countries such as the Central Asian Republics, which had either no data or scanty data on a range of demographic and health issues covered by the DHS.

6.2.1 Target population

The survey respondents included women aged 15-49 and all children younger than 60 months. In Uzbekistan and Kyrgyzstan only children below the age of 36 months were included in the surveys, while in Kazakhstan the survey included all children below the age of 60 months. The target population for analysis therefore included all children the ages of 0 and 35 months at the time of the surveys, together with their mothers, so that results would be comparable.

6.2.2 Sample design

(a) Sampling frame

The DHS for the three countries employed a nationally representative probability sample of women aged 15-49 and included approximately 4 000 women in each country. Complex sample designs were used. Selected regions, (both urban and rural), which are divided into oblasts[14], were included in the survey. In the urban areas, the sampling frame was the list of therapeutical uchastoks,[15] mainly cities. For the small towns, each town was divided into smaller segments, which were treated as if they were uchastoks. In the rural areas, the sampling frame was the list of all villages in the whole country. The primary sampling units (PSUs) were raions,[16] which were selected with probabilities proportional to population size.

The sample was selected in two stages. In the urban areas, the primary sampling units selected during the first sampling stage corresponded to the uchastoks. In the field, large uchastoks were divided into smaller segments, only one of which was selected for the survey. A complete listing of all the households residing in each selected segment was carried out. The lists of households obtained were used as the frame for the second stage sampling – in each cluster households were selected with the women aged between 15 and 49 being eligible to be interviewed.

[14] Oblasts are provinces.
[15] Each city is divided into uchastoks (health blocks), each of which is the responsibility of one physician. People living in the uchastoks go to a designated health centre for service.
[16] Raions are districts.

In the rural areas, the first stage-sampling units were the villages. In the field, very large villages that had been selected were divided into smaller segments of which only one was selected. The nonproportional distribution of the sample in the different survey regions meant that sampling weights were applied to the data.

(b) Questionnaires

Two main questionnaires with similar requirements were used for the DHS in these countries – the Household Questionnaire and the Individual Questionnaire. In Kazakhstan a third questionnaire, the Male Questionnaire, was also used. The questionnaires were based on the model survey instruments developed in the DHS programme. During consultations with specialists in the areas of reproductive health, child health and nutrition, the questionnaires were adapted to the data needs of the three countries. The questionnaires were all developed in English and translated into both Russian and one of the major languages in each country. After a pre-test, the questionnaires were further modified.

The Household Questionnaire was used to enumerate all the usual members and visitors in the sample households and to collect information relating to the socioeconomic position of the household. In the first part of the Household Questionnaire, information was collected on age, gender, educational attainment, marital status, and relationship to the head of the household of each person listed as a household member or visitor. A primary objective of this first part of the Household Questionnaire was to identify women eligible for the individual interview. The second part of the Household Questionnaire included questions relating to the dwelling unit, such as the number of rooms, the flooring material, the source of water, and the type of toilet facilities. This second part also included questions on the availability of a variety of consumer goods.

The Individual Women's Questionnaire was used to collect information from women aged 15-49. These women were asked questions on the following major topics:

- background characteristics
- pregnancy history
- outcome of pregnancies and antenatal care and attendance with more detailed information collected for postnatal care and attendance that occurred in the three years immediately preceding the surveys
- child health and nutrition practices
- child immunisation and episodes of diarrhoea and respiratory illness
- knowledge and use of contraception
- marriage and fertility preferences
- background of husband and woman's occupation
- anthropometry of mothers and children
- haemoglobin measurement of women and children

(c) Training and fieldwork

The questionnaires were pre-tested in all three countries. A number of females were selected and trained over a two-week period at different centres within these three countries. Training consisted of classroom lectures and practice, as well as practice interviews conducted in the field. The pre-test included one week of interviews in an urban area and one week in a rural area. The pre-test interviewers were retained to serve as supervisors and field editors for the main survey. All interviewers were female, while most of the supervisors and technicians were male. Interviewers were selected based on their performance during the training period. Teams consisting of between eight and nine members carried out the data collection. Each team consisted of a team supervisor, one editor, one household interviewer, four to five individual female interviewers, and one medical technician (responsible for height and weight measurement, and anaemia testing).

(d) Response rates

In all three countries the overall response rate, which was the product of the household and the individual interviews, was 97 percent for the individual interviews and 99 percent for the household interviews (see Table 6.1 below). The principal reason for non-response was a failure to find an eligible woman at home after repeated visits to the households.

(e) Limitations of the study

While it is well documented that the DHS limits data errors through rigorous cleaning and interrogation of the data, it is, however, still possible to find errors in the data. A global assessment of DHS data collected in 22 countries did not, however, detect any serious errors that could have an effect on demographic estimates (Desai & Alva, 1998). A systematic bias and/or rounding-off errors are possible while reading and recording the height and weight of a child. Age heaping and digit preference also constitute possible measurement errors. In addition, this study is based on cross-sectional data, which implies that it is not always possible to determine the direction of causal relationships explicitly.

Table 6.1 Results of the household and individual interviews

	Kazakhstan (1999)			Kyrgyzstan (1997)			Uzbekistan (1996)		
Results	**Urban**	**Rural**	**Total**	**Urban**	**Rural**	**Total**	**Urban**	**Rural**	**Total**
Household interviews									
Households sampled	4 311	1 990	6 301	1 757	2 064	3 821	2 228	1 717	3 945
Households found	4 038	1 922	5 960	1 686	2 009	3 695	2 099	1 664	3 763
Households interviewed	3 939	1 905	5 844	1 668	2 004	3 672	2 062	1 641	3 703
Household response rate	97.5	99.1	98.1	98.8	99.8	99.4	98.2	98.6	98.4
Individual interviews									
Number of eligible women	2 989	1 917	4 906	1 517	2 437	3 954	2 388	2 156	4 544
Number of eligible women interviewed	2 927	1 873	4 800	1 485	2 363	3 848	2 306	2 109	4 415
Eligible women response rate	97.9	97.7	97.8	97.9	97.0	97.3	96.6	97.8	97.2

Sources: UZDHS 1996, KRDHS 1997, KDHS 1999

6.3 ANALYTICAL FRAMEWORK

Univariate analyses of all variables relevant to the study of undernutrition[17] were carried out for each country. A joint model for the three countries was developed. This joint model was based on the assumption that the Central Asian countries are similar in a political, demographic and socioeconomic sense. The surveys were also carried out over a relatively short period of time in all three countries.

The assumption of a joint model was further supported by the absence of a significant interaction between country and other variables in the analysis. Multicollinearity of the predictors was assessed using the standardised methods as described by Pallant (2003). Thereafter bivariate analyses of the nutritional status of children in terms of demographic and background characteristics were carried out. Variables that indicated a significant association in the bivariate tabulations were tested for inclusion in the models for multivariate analyses. Multivariate logistic regression analysis was carried out separately to study the chances of underweight and stunted children being included in the study. Wasting was excluded in the analysis because very few independent variables showed any significant relationship with the dependent variables. Thereafter models predicting determinants of undernutrition were postulated. The forward selection procedure with a significance level of 0.5 for the Wald statistic was used. All potential interactions were included one by one as proposed by Hosmer and Lemenshow (2000). All those variables that appeared significant were used in the logistic regression. Data analysis was carried out using statistical package SPSS version 14.

6.3.1 Operational definitions of variables

The selection of potential predictors of child malnutrition was based on a literature review. The independent variables used in the analysis refer to different categories to which they could be said to belong, namely, socioeconomic and environmental (maternal education, number of household members, household

[17] In this study, the term "undernutrition" refers collectively to stunting, underweight and wasting.

wealth, number of children under the age of five years, toilet and water facilities), and biodemographic and health (child age, birth weight, length of preceding birth interval, number of immunisations, breastfeeding, child haemoglobin level, birth order, heard of ORT, BMI, ethnicity and delivery assistance).

The central variable in this study, which has already been introduced in chapter one is undernutrition, which is composed of height-for-age (*HTA*), weight-for-age (*WTA*) and weight-for-height (*WHT*).

(a) Undernutrition (dependent variable)

Height-for-age (*HTA*): Standard deviations from the reference median (stunting) treated as a dichotomous variable. Weight-for-age (*WTA*): Weight-for-age standard deviations from the reference median (underweight) treated as a dichotomous variable. Weight-for-height (*WHT*)*:* Weight-for-height standard deviations from the reference median (wasting) treated as a dichotomous variable.

(b) Socioeconomic and environmental factors

Three educational categories were defined for the mother *(MEDUC): 0-Part two* which includes *no education* and *primary education* (six years duration), *secondary education* (eleven years duration) and *higher education* (university degree). Household wealth *(HOWE):* a household wealth index consisting of four levels – very *low, medium, high* and very *high* – was created to be used as proxy for economic conditions of the household. Economic conditions of the household included ownership or possession of the following items: a car, bicycle, telephone, radio, television, electricity, main floor material, refrigerator and motorcycle. The number of people per household *(NOHOUSE)* was recoded into three categories: *1-4 persons*, *5-7 persons* and *more than 7 persons.* Three categories were defined for the number of children per household under the age of five *(NOCH): one child or no other child, two children and more than two children.* As far as environmental variables were concerned two variables only –

toilet and water facilities – revealed a significant association with the nutritional variables. Toilet facility *(TOIL)* was defined in terms of two categories; *flush toilet* and *pit toilet*.

The original variable source of drinking water *(WATER)* consisted of different categories. This variable was recoded into a binary variable: having access to piped water or unpiped water.

(c) Biodemographic and health factors

A set of four variables was defined for the variable child age *(CAGE): 0-5 months, 6-11 months, 12-23 months* and *24-35 months.* Birth weight (*BWE*) was recoded into a binary variable: below 2.5kg and above 2.5 kg. Four categories were defined for the variable length of the preceding birth interval *(BINT): first order births, interval less than 24 months, between 24 and 47 months* and *above 47 months.* According to WHO all children should have received at least eight immunisations by age of three years. Three categories for the number of immunisation *(NIMM)* received were created: *0-3 immunisations, 4-7 immunisations* and *8 immunisations.* Duration of breastfeeding in months *(BREAST),* measured as the number of months a child has been breastfed between the date of birth and the date of the interview, was divided into three categories: *0-6 months, 7-12 months* and *13-35 months*. The original haemoglobin *(HAEMO)* levels for the child were retained: severe, moderate, mild and none. Birth order *(BORD)* number which gives the order in which the children were born is measured as an interval variable, 1, 2, 3, & 3+.

Health variables included the number of immunisations, whether the mother had heard of ORT, and whether or not the mother had received professional health assistance during delivery of child, were included in the survey. Three categories of the number of immunisation *(NIMM)* received were created: *0-3 immunisations, 4-7 immunisations* and *8 immunisations.* The original variable assessing the type of assistance received during delivery *(DEL)* consisted of different categories. These were recoded into *professional* (doctor and/or nurse) and *non-professional* assistance (traditional birth attendant, relative and/or friend). Ethnicity *(ETHN)* was an additional characteristic of the mothers. Five categories were identified as the major ethnic groups in the study – *Kazakhs, Uzbeks, Kyrgyz, Russian and other*. The health and nutritional status of the mother was measured (in the DHS) using the body mass index *(BMI)*. Three categories of the BMI were created: *<18.5, 18.5-29.9 and >30*.

All variables used in this analysis had nearly complete responses with less than 5% missing values. Data from all three countries was merged and the variable country was included in the model.

6.3.2 Omitted "Traditional" explanatory variables

There are certain "traditional" variables that have been omitted in the analyses constructed in this study and these deserve mention. They include place of residence, mother's age, father's age, mother's occupation, father's occupation, father's education, and prenatal care. A number of these variables had a significant association and/or inverse relationship with the nutritional status of the child at country level, but this disappeared after the data was pooled.

(a) Place of residence: It is a well-established fact that, in any developing country, malnutrition among urban dwellers is generally much lower than that of rural dwellers. This is simply because of urban bias development. Most high quality health and educational services, housing and water services, and well-paid jobs in developing countries are concentrated in the urban areas. Thus, the inverse relationship between child malnutrition and urban or rural residence does not reflect the effects of the type of residence per se, but other factors associated with it. A number of recent studies have either suggested or found no significant inverse association between urban residence and child malnutrition within a multivariate context. Other factors such as education, marital status, family size, occupation, and unemployment may be underlying factors, which account for urban-rural differentials (see Chapter 3, p61). Trial runs for bivariate association showed similar results.

(b) Ages of mother and father: Babies born to young and to old mothers, compared to those born to mothers of intermediate age, for example, 25-35, have a higher risk of malnutrition. Several studies have found that malnutrition decreases with maternal age. Younger mothers experience higher child malnutrition levels as compared to older mothers. This may be due to a number of reasons such as lack of experience or undeveloped physiology. However, the mother's age may also reflect socioeconomic status, including status in the household hierarchy, and this may constitute one of the underlying factors, which account for the significant effect observed of the mother's age on child malnutrition. The age of the mother did not reveal any correlation with child malnutrition. Similarly, the age of the father did not reveal any correlation with child malnutrition (see Chapter 3, p52).

(c) Occupation and employment of mother and father: Parental occupational status especially that of the mother is expected to have an effect on the health status of children. According to Delpeuch et al., (2000) a salaried mother might be in a position to devote her income to her own and her children's needs, thus resulting in the improved nutritional status of her children.

The father's occupation and employment has been found to be positively associated with child nutritional status. Single country data analyses revealed an association between the father's occupation and employment and child nutritional status, but this disappeared with pooled data. Neither occupation nor employment shows any association with child malnutrition.

(d) Education of father: The father's education has been found to have a positive and statistically significant effect on child nutritional status in a number of studies conducted in developing countries (see Chapter 3, p68). However, the educational level of the father has been found to affect child health and nutritional status mainly through its income generating properties (Glewwe 1999), did not show any significant association with child nutritional status in the three countries studied

(e) Prenatal care: It is expected that prenatal care will have a positive and significant effect on the nutritional status of children because of the routine health examinations and parenting skills given to mothers during pregnancy. Infants and children of mothers who received prenatal care, either from a medical doctor or a nurse, were more likely to survive (see Chapter 3, pp56-57).

6.4. REGRESSION ANALYSES

The outcome measure of interest in the study on the nutritional status of children is the prevalence of undernutrition. The best method for the analysis of cross-sectional data is to treat anthropometrical indicators as continuous variables, and to focus on patterns of covariation rather than on the odds of being in one discrete category rather than another (McMurray, 1996:161). Certain authors prefer to use cut-off points to transform Z-scores or percentages of the reference median into dichotomous dependent variables so that odds ratios may be estimated with logistic regression models (Timaeus & Lush, 1995). However, this approach does not guarantee the reliability of the data; especially because the approach makes questionable assumptions to the effect that, the data allow meaningful classification to either side of a cut-off point, and that there are real differences between children on either side of the cut-off point.

An appropriate analytical technique for this purpose is linear regression, in terms of which categorical independent variables are converted into dichotomous dummy variables. With linear regression, it is possible to avoid the need to classify cases to either side of a cut-off point while the odds ratios derived from logistic regression might compromise the analysis (McMurray, 1996:164). However, multiple regression estimates of covariation with growth attainment, as well as logistic regression, which looked at the relative risk of being 2 Standard Deviations (SD) below the reference median, were carried out.

Logistic regression is the form of analysis most suited to studying the variability in the prevalence of undernutrition The theoretical advantages of using logistic regression are two-fold: its weighted form of analysis stabilises the variance, and the logit transform adjusts for skewness in the distribution of prevalence. It is possible to see whether a distribution is skewed by comparing the values of the mean and the median. For a normal distribution, the mean and the median are numerically identical. As the distribution becomes progressively more skewed the difference between the median and the mean increases. The one disadvantage for logistic regression lies in its complexity – logistic regression coefficients are in units of log odds ratios, which could compromise the analysis (Mascie-Taylor, 2003). An alternative method would be to work with a simpler transformation than the logit, such as the logarithmic transformation. This adjusts for skewness but does not stabilise the variance as effectively (Pelletier, et al., 1993: 1130-1133).

6.4.1 Logistic regression

The logistic regression model examines the estimates and the probability that an event may or may not occur. This model requires far fewer assumptions than the linear regression model. The logistic model may be written in terms of the odds of an event occurring. The odds of an event occurring are defined as the ratio of the probability that it will occur to the probability that it will not occur. As discussed earlier the percentage of children expected to fall above -2SDs is only 2.3 percent, which indicates severe obesity, and 13.6 percent between -1.00 and –1.99 indicating moderate obesity. In developing countries only a very small percentage (3.3%) of children under the age of five years are classified as obese, as the majority are either nourished or malnourished. Owing to this small percentage of obese children studies of child nutrition normally examine the odds of a child being nourished or malnourished. The logistic regression model looked at the odds of cases in various categories being 2SDs below the reference median. An odds ratio of less than one indicates lower odds relative to the category for that variable, and a value greater than one indicates higher odds.

The logistic regression model in terms of the log odds, which is called a logit, is written as follows:

$$\log\left[\frac{P(event)}{1-P(event)}\right]=\alpha+\beta_1X_1+\beta_2X_2+\dots\dots\dots\dots\dots\dots\dots+\beta_kX_k$$

where P (event) is the probability that an event will occur.

Since it is easier to think of odds rather than log odds the logistic equation may be written in terms of odds as follows:

$$\left[\frac{P(event)}{1-P(event)}\right]=e^{\alpha+\beta_1X_1+\beta_2X_2+\dots\dots\dots\dots\dots\dots\dots+\beta_kX_k}$$

and

P (event) = $\dfrac{1}{1+e^{-z}}$ where z = $\alpha+\beta_1X_1+\beta_2X_2+\dots\dots\dots\dots\dots\dots\dots+\beta_kX_k$

6.5 SUMMARY

These then are the tools for the analysis that is discussed in Chapter 7. The variables selected for inclusion in the models are justified by what is known about their influence on child malnutrition in other populations as reviewed in chapter three. In this study, these are the most explanatory variables. However, the claim is by no means made that these are the best variables for explaining child malnutrition in Central Asia as a whole, especially in view of the fact that only three of the five countries were included in this study. For the moment it is thought that, with the evidence provided in Chapters 4 and 5 about these variables, they are the most suitable explanatory variables. Chapters 7 and 8 present the discussion of the impact of the selected independent variables on child malnutrition estimated using binary logistic regression. The following chapter discusses the effects of these variables on child malnutrition.