Chapter 4

Data and methods

4.1 Introduction

The present chapter describes the data and methods that are used in this study. The shift of focus from statistical associations with socioeconomic variables to the causal factors and mechanisms of morbidity and mortality requires different data sources than the ones traditionally used in demography. The present study tries to expand and complement demographic data sources by using data from sources in the health sciences. The study adopts a multi-method approach by combining data from secondary sources and data from a newly-conducted hospital survey.

The study has two main research questions that were presented in Chapter 1. The first question is seeking to find the most important causal factors in loss and death during gestation, birth, and the neonatal period in the developed world. The second question is about the significance of these factors within the framework of an epidemiologic transition, as based on the current situation in regions in transition that are approaching the later stages of the epidemiologic transition. As a result, the study can be divided into two parts that are each based on a different approach. In the present chapter, Section 4.2 describes the data sources, methods, and difficulties related to the first question concerning the developed world. In this, an important topic is the concept of relative importance and the calculation of this relative importance. Subsequently, Section 4.3 discusses data and methods for the second question on regions in transition.

4.2 Part I: The developed world

What are the most important causal factors of loss and death during gestation, birth, and the neonatal period in the developed world? In this, the first research question, the term ‘developed world’ refers to regions in the later stages of the epidemiologic transition, i.e. low-mortality countries. Various subquestions are related to the main question: What are important factors? How may they be ordered and classified? What is the relative importance of selected factors as measured at the level of the individual? What is the relative importance of selected factors as measured at the level of the population? Does relative importance change over the life course – from gestation through the perinatal period and the neonatal period? And, are the selected factors related?

In Chapter 3, the first two subquestions have already been addressed as well as a preview of the last subquestion. The other questions will be answered in the proceeding chapters. The research providing the answers to these questions is based on the data and methods that are described below. However, before any discussion of data or methodology is
possible, the research questions of this first part require further specification and several terms have to be operationalised.

### 4.2.1 Specification and Operationalisation

The concepts of loss and death during gestation, birth, and the neonatal period were already clarified in Chapter 3 (Section 3.2.3). The present study is interested in ‘natural’ loss and death and, therefore, does not include induced abortion (IA) as one of the adverse outcomes. Further, perinatal death is split into its two components: stillbirth and early neonatal death. Combined reporting of all perinatal deaths is believed to conceal important distinctions between the components (Kalter 1991). Moreover, the populations at risk of the two events differ: unborn foetuses as against liveborn neonates. Therefore, the subdivision of perinatal mortality into stillbirth and early neonatal mortality would provide more accuracy. As a result, the relative importance of causal factors over the life course will be studied here on the basis of spontaneous abortion or loss (SA), stillbirth (SB), early neonatal death (ENND), and neonatal death (NND).

For simplification purposes, the focus here is on singletons only since multifetal or multiple pregnancies are known to have an elevated risk of morbidity and mortality. Complications that are more frequently encountered in twins or other multiplets include: abortion, perinatal mortality, low birth weight, preterm birth, foetal growth retardation, malformations, complicated labour, and abnormal foetal presentation (Cunningham et al. 1993). Some complications even pertain exclusively to multiplets, such as conjoined twins, intertwining of umbilical cords, vascular communications between foetuses, and interlocking twins at birth. Thus, the study is restricted to singletons in terms of selected risk factors and intermediate outcomes. Nevertheless, some of the valuable secondary data sources (e.g. study reports, articles) that are discussed in the present study do include multiplets.

The term ‘developed world’ has to be defined clearly before any gathering of data or information is undertaken. In the current study, the term refers to regions in the later stages of the epidemiologic transition, i.e. low-mortality countries. The Global Burden of Disease Study (GBD) divided the world into eight geographic regions to assess regional patterns in morbidity and mortality (World Bank 1993) (see also Chapter 2). Appendix B provides an overview of these eight regions. The Established Market Economies (EME) and the Formerly Socialist Economies of Europe (FSE) were seen as comprising the ‘developed’ part of the world. However, data in Chapter 2 demonstrated that the two regions show differences with regard to life expectancy at birth, mortality rates (including neonatal and perinatal mortality rate), and the causes of neonatal and infant mortality. The countries of the EME region are most advanced in terms of the epidemiologic transition. Furthermore, induced abortion has been a major method of birth control and family planning in Eastern Europe (Ketting 1993) and this could possibly confuse any analysis of spontaneous foetal loss. In the present study, the developed world will therefore be operationalised as consisting only of the EME region. The EME region includes all countries, with the exception of Turkey, that were members of the Organization for Economic Cooperation and Development (OECD) at the time of the GBD, plus a number of small high-income economies in Europe (World Bank 1993, p.xi).
CHAPTER 4: DATA AND METHODS

A few words are required on the issue of how to express the relative importance of the selected factors in relation to loss/death, both at the individual level and the population level. From the perspective of the individual, it would be interesting to know the personal risk (cf. incidence) of loss or death by selected adverse pregnancy/birth outcome, cf. the case fatality ratio. This absolute risk can be compared to the risk in non-exposed or unaffected individuals in order to determine the relative risk (RR). From the population perspective, it would be interesting to know the frequency (cf. prevalence) of the pregnancy/birth outcome within the population as well as the proportion of losses or deaths that are due to, or can be attributed to, the risk factor. When only the exposed or affected population is taken into consideration, such a measure of attribution is labelled attributable risk among the exposed or AR(E). When the focus is on the total population consisting of both exposed and non-exposed individuals, the measure is referred to as the population attributable risk (PAR) or the etiologic fraction (EF). Section 4.2.2, below, briefly describes and discusses these measures as well as some other basic, albeit important, concepts related to risk and frequency. The distinction between the individual level and the population level implies two types of interventions in the underlying processes. One type of intervention that would improve the survival chances of individuals affected by the risk factor which would then also have an effect on mortality figures in the total population. The second type of intervention would lower the prevalence of the risk factor in the population.

4.2.2 CONCEPTS AND MEASURES RELATED TO FREQUENCY AND RISK

In the present book, several concepts and measures are applied to indicate levels of loss and mortality, and to assess the importance and impact of risk factors. Since the study is situated at the boundary between demography and epidemiology, this section clarifies certain terms to avoid confusion at a later stage. The subsections below include a discussion of the following concepts and measures: ratio, proportion, rate, incidence, prevalence, relative risk, attributable risk among the exposed, and etiologic fraction.

Ratio, proportion, and rate

Ratio, proportion, and rate are basic concepts that underlie many measures related to frequency and risk. However, the terms are regularly misused and this can lead to confusion. Therefore, a brief description of each of these concepts is provided below. Most of the information is based on Elandt-Johnson (1975), Elandt-Johnson and Johnson (1980), Kleinbaum et al. (1982), Palmore and Gardner (1983), Schulman et al. (1988), Selvin (1991), Young (1998), and Preston et al. (2001).

A ratio is the result of dividing one quantity by another (\(y / x\)). The numerator and denominator of a ratio are two separate and distinct quantities. An example is sex ratio in which the number of males is divided by the number of females. Another example is foetal death or stillbirth ratio, which is calculated as the number of stillbirths divided by the number of live births.

A proportion is a type of ratio in which the numerator is included in the denominator \((y / [y+x])\). For example, the proportion of males in a population is calculated by dividing the number of males by the total population, i.e. the total number of females and males. As a
result, proportions are dimensionless or unitless and their values always lie between zero and one. Further, a percentage is a proportion that has been multiplied by the constant 100 and is therefore expressed per 100. Proportion is closely related to the concept of probability. As with a proportion, a probability is unitless with a value between zero and one. It refers to risk or the chance that a certain event will occur, within a specified time interval.

Probabilities and proportions only refer to time in an indirect way. On the other hand, the concept of rate is commonly used to reflect the rapidity of change in a phenomenon per unit of time. More generally, a rate may be defined as a measure of change in one quantity relative to a change per unit of another quantity ($\Delta y / \Delta x$). In practice, the latter quantity is usually time. For example, the speed of a car expressed in kilometres per hour, is a rate. Another example is the number of new cases of a disease during a specified time period divided by the person-time at risk. This measure is a so-called incidence rate (see the following subsection), which is comparable to the occurrence-exposure rate in demography. In general, rates can have values from zero up to infinity. According to Miller and Homan (1994), rate does not represent the concept of risk, unlike probability, but it reflects ‘force’. Rate refers to a population while risk may be interpreted on the individual level (Kleinbaum et al. 1982).

The term ‘rate’ is frequently confused with proportion and probability. Since rates are often small and based on unit time intervals (e.g. one year), the values for a rate and a probability are more or less indistinguishable (Selvin 1991). Many so-called mortality ‘rates’ are in fact not true rates but have the structure of proportions or probabilities. Important examples include foetal death rate, stillbirth rate, perinatal mortality rate, early neonatal mortality rate, neonatal mortality rate, postneonatal mortality rate, and infant mortality rate (see also Section 2.3.3). They indicate the number of deaths per 1,000 births or 1,000 live births. Perinatal mortality rate and foetal death or stillbirth rate should be distinguished from perinatal mortality ratio and foetal death or stillbirth ratio. In these cases, the ratio is calculated as the number of deaths divided by the number of live births whereas the rate divides the number of deaths by the number of total births, i.e. live births plus stillbirths (WHO 1993). Though they are not true rates, the mortality figures above are named labelled ‘rates’ by convention (for example, see WHO 1993). In the present book, I will refer to them by their conventional names.

In life table analysis, there is a clear distinction between probabilities (\(\mu q_x\)) and rates (\(\mu m_x\)). Probabilities in the standard life table are calculated as:

\[
\mu q_x = \frac{\delta d_x}{R_x}
\]

(4.1)

or the number of events (i.e. deaths) during interval \([x, x+h]\) divided by the population at risk or the risk set. On the other hand, rates are calculated as:

\[
\mu m_x = \frac{\delta d_x}{\delta L_x}
\]

(4.2)

or the number of events (i.e. deaths) during interval \([x, x+h]\) divided by person-time lived between age \(x\) and age \(x+h\). For a more extensive discussion of life table methodology, please refer to Chapter 5.
Incidence and prevalence

In epidemiology, the basic measures of disease frequency are incidence and prevalence. Although the distinction between these two concepts seems rather straightforward at first sight, they also lead to confusion and debate. The majority of information below is derived from Elandt-Johnson (1975), Schulman et al. (1988), Selvin (1991), Rockett (1994), and Young (1998).

Incidence reflects the number of new cases occurring in a population during a specified period of time. The key concept is a change in status, which is not necessarily from healthy to sick but may also be from alive to dead (Young 1998, p.25). Incidence may be expressed as a rate. An incidence rate is calculated as the number of new cases that develop in a population initially free of the outcome during a specified time period divided by the sum of individual person-times at risk during that time period. Besides a rate, incidence can also be expressed as a proportion: incidence proportion, in epidemiology also referred to as cumulative incidence. Incidence proportion expresses the proportion of individuals in a population who develop the outcome (e.g. disease, injury, death) within a specified period of observation. The measure is calculated as the number of new cases that develop in a population initially free of the outcome during a specified time period divided by the number of persons at risk at the beginning of that time period.

While incidence is about new cases, prevalence focuses on existing cases and refers to status rather than to change. As Young (1998) states: “Incidence is about becoming, whereas prevalence is about being something or having something” (p.25-26). Prevalence describes the proportion of individuals in a population who have a certain health problem or characteristic at a particular point in time or during a specific time period. It is calculated as the number of existing cases at a given time divided by the total population at that time. The denotation of time can either refer to a particular point in time (i.e. point prevalence) or to a specific time period (i.e. period prevalence). In addition, it should be noted that time can refer to a real time, such as a specific date, or to a specific point or moment in life, such as birth, death, or hospital discharge. Finally, it should be noted that although in the literature prevalence is frequently called rate, it is in fact always a proportion, which is clearly indicated by its denominator.

Prevalence and incidence are mathematically related. Prevalence equals incidence multiplied by the average duration of the disease, or:

\[ P = I \times D \]  

where \( P \) is prevalence, \( I \) is incidence, and \( D \) is the average duration of the disease. This relationship holds when the disease is rare (i.e. \( P \) is low) and incidence is stable over time.

Despite their apparent clear differences in definition, on occasion the use of the two concepts results in confusion and debate. For instance, with respect to autopsies and congenital anomalies, see Young (1998). Some authors refer to the frequency of congenital anomalies among births or live births as ‘incidence’. They often argue that the measure denotes the number of new cases in the population (e.g. Myrianthopoulos 1985 cited by Cornel et al. 1993a). On the other hand, several authors – including Leck (1983), Schulman et al. (1988), Cornel et al. (1993a), and Young (1998) – have argued that ‘prevalence
Early life changes

...proportion’ is the only correct term in the case of congenital anomalies at birth. These researchers have stronger arguments, which will be discussed here only briefly. The proportion of births or live births with congenital anomalies in fact reflects prevalence at the time of birth. Although all cases are new to the population their anomalies were already present before birth. In other words, they entered the study population being affected already (Cornel et al. 1993a, p.14). The actual incidence of anomalies would be very complicated to assess since normally the total population at risk (i.e. the total number of conceptions) as well as the total number of new cases in the population are unknown. They are virtually inaccessible to observation, partly because of undetected and unreported spontaneous abortions. In addition, many anomalies may originate only during a brief period in gestation (e.g. organogenesis), and as a result incidence could only be studied during this short period.

In the present study, some frequency measures and proportions are also confusing when it comes to the distinction between incidence and prevalence. In general, the proportion of cases with a specific risk factor (i.e. adverse pregnancy/birth outcome) among foetuses in utero, induced abortions (IA), spontaneous abortions (SA), stillbirths (SB), live births (LB), all births (B), or (early) neonatal deaths (ENND, NND) should be regarded as a prevalence proportion. Conversely, the proportion of cases that end in spontaneous abortion, stillbirth, or neonatal death should, in general, be considered as an incidence, or more specifically, an incidence proportion. This measure of incidence is comparable to the case-fatality ratio, proportion, or ‘rate’ in epidemiology, which is calculated as the number of cases of a disease that end in death divided by the total number of cases of that disease within the population (Selvin 1991, p.8). Nevertheless, complications arise in relation to low birth weight (LBW), preterm birth, and birth asphyxia.

The weight and gestational age of a child already exist before birth. Yet, intrinsic to their definitions, birth weight and gestational age at birth are not assessed until after birth (live or stillbirth). Consequently, LBW and preterm birth cannot be related to foetuses in utero, induced abortions, or spontaneous abortions. Measures of prevalence and incidence in this respect are non-existent. Furthermore, there is confusion with regard to stillbirth. It would be improper to label the proportion of stillbirths in LBW or preterm cases ‘incidence’. After all, the population at risk in the denominator refers to newborn babies rather than foetuses in utero. Since incidence measures are required for the calculation of the relative risk (see the subsection below), it is not possible to calculate relative risks of stillbirth for LBW and preterm newborns.

The situation concerning birth asphyxia (BA) is similar. In line with its definition, BA is also not to be assessed before birth and again, it seems improper to speak of the incidence of stillbirth. Nevertheless, BA is probably a very important causal factor of intrapartum stillbirth. Hence, an attempt will be made in Chapter 7 to estimate incidence of intrapartum stillbirth among cases asphyxiated at birth. Additional confusion arises because BA itself could also be regarded as an event that takes place during the birth process. Therefore, many authors refer to the proportion of BA cases among births (stillbirths and/or live births) as an incidence. However, there are some technical complications in specifying the population at risk. It is important to note that, strictly speaking, antepartum or macerated stillbirths should be excluded from the denominator since they are not at risk of experiencing BA. Furthermore,
the distinction between incidence and prevalence is very much dependent on the adopted perspective. During birth, BA is an event, or a change in status, whereas shortly after birth BA is a characteristic or reference to status. Here, the prevalence and incidence proportions have nearly the same values. In general, prevalence and incidence proportions are roughly equal when the duration of the ‘disease’ is short (Selvin 1991, p.8), as can be derived from equation (4.3). In the present study, BA is regarded as a characteristic of the newborn, and the proportion of BA cases among births or live births as prevalence. However, the book will generally refer to this measure with more neutral expressions such as ‘proportion’ and ‘frequency’.

Relative risk and attributable risk

The strength of association between exposure to a risk factor and a specific disease or outcome (e.g. death) may be expressed by various measures, including the easy interpretable ‘relative risk’. The effect or impact of the exposure in a population may be quantified by the ‘attributable risk among the exposed’ and the ‘etiologic fraction’. This subsection discusses these measures in more detail. Most of the information below is based on Selvin (1991), Rockett (1994), Fletcher et al. (1996), Jekel et al. (1996), Young (1998), and Kramer et al. (2000).

The relative risk (RR) is the ratio of incidence of the outcome in exposed persons to the incidence among non-exposed persons. The measure may be a rate ratio based on true incidence rates or a risk ratio based on incidence proportions. The relative risks that are estimated in the present study are all risk ratios. The risk ratio is calculated as:

\[ RR = I_1 / I_0 \]  

(4.4)

where \( I_1 \) is the incidence (proportion) of the event among those exposed or affected, and \( I_0 \) is the incidence (proportion) in the reference group, i.e. among those who were not exposed or affected. In the present study, the event or outcome of interest is loss/death and therefore, equation (4.4) could be rewritten as:

\[ RR = (d_1 / l_1) / (d_0 / l_0) = q_1 / q_0 \]  

(4.5)

where \( d \) indicates the number of deaths in the population and \( l \) is the total population. In other words, the RR is the probability of loss/death in the affected population \( (q_1) \) divided by the probability of loss/death among those who do not suffer from the adverse pregnancy or birth outcome \( (q_0) \). In general, the relative risk indicates how many times more likely persons with the risk factor will experience the outcome as compared to persons without the risk factor. For example, if the relative risk of lung cancer death for people who smoke is 11, this means that smokers are 11 times as likely as non-smokers to die of lung cancer. The larger the relative risk, the stronger the association between the outcome and the risk factor. A value of 1 indicates no association. The RR thus measures the strength of association, but it is important to remember that it does not prove actual causality.

The RR is conceptually comparable to another commonly used ratio, the odds ratio (OR). The OR compares the odds of exposure between those with, versus those without, the outcome, or compares the odds of the outcome between those exposed and those not exposed.
EARLY LIFE CHANGES

Here, the comparison of the odds of loss/death among those who are affected and those who do not suffer from the adverse pregnancy or birth outcome is reflected in:

$$\text{OR} = \frac{q_1 / (1 - q_1)}{q_0 / (1 - q_0)}$$  \hspace{1cm} (4.6)

The OR approximates the RR when the incidence of the disease or outcome is low. Nevertheless, RR is generally preferred over OR.

The **attributable risk among the exposed, AR(E)**, quantifies the effect or impact that a given risk factor has in the affected or exposed population. More precisely, the AR(E) indicates, for the exposed population, the proportion of cases of the outcome (i.e. loss or death) which is attributable to, or due to, the exposure or risk factor of interest. The measure is calculated as:

$$\text{AR(E)} = \frac{I_1 - I_0}{I_1}$$  \hspace{1cm} (4.7)

or risk difference between the exposed and the non-exposed group divided by incidence in the exposed group. Further,

$$\text{AR(E)} = \frac{\text{RR} - 1}{\text{RR}}$$  \hspace{1cm} (4.8)

which indicates that the measure can also be computed on the basis of the RR (Young 1998).

Whereas AR(E) is concerned with the impact in the exposed population, the **etiologic fraction (EF)** quantifies the effect of the risk factor within the total population. In other words, the EF considers the public health impact. The etiologic fraction is the proportion of all cases of the outcome in the total population that can be attributed to exposure to the risk factor. However, it should be noted that the observed fraction may be polluted by confounding factors (Miettinen 1974). The etiologic fraction is calculated as:

$$\text{EF} = \frac{I - I_0}{I}$$  \hspace{1cm} (4.9)

where $I_0$ is the incidence among the non-exposed and $I$ is the incidence in the total population, i.e. the exposed and the non-exposed. Like the AR(E), the EF can also be computed as a function of the RR:

$$\text{EF} = \frac{P(\text{RR} - 1)}{P(\text{RR} - 1) + 1}$$  \hspace{1cm} (4.10)

where $P$ is the proportion of the population that is exposed to the risk factor, i.e. the prevalence of the risk factor. This equation demonstrates that public health importance is not only affected by the strength of the association between risk factor and outcome, but also by the prevalence of the risk factor in the population. The EF increases when risk factors are more common but rare risk factors have a reduced impact. The etiologic fraction is sometimes also referred to as the population attributable fraction or population attributable risk (PAR).

As described above, the RR, AR(E) and EF are relatively easy to interpret in dichotomous situations such as exposed versus non-exposed. Sometimes, however, the type of variable and the available data offer the possibility of creating polytomous categories. The various categories are compared to one and the same reference category. For example, birth weight categories < 1,500 g, 1,500-1,999 g, 2,000-2,499 g, 3,000-3,499 g, and ≥ 3,500 g could be compared to weight group 2,500-2,999 g. In such as polytomous situation, RR is calculated as:
CHAPTER 4: DATA AND METHODS

\[ RR_i = \frac{I_i}{I_0} \]  

(4.11)

where \( i \) refers to the category \( i \) that is to be compared to the reference category, category \( 0 \). Consequently, \( RR_i \) indicates how much more frequently a given outcome occurs in persons from category \( i \) compared to persons in the reference group. Using equation (4.8), \( AR(E) \) can be computed but its results are not easy to interpret. In general terms, it refers to the proportion of cases of the outcome in category \( i \) that can be attributed to being exposed to this category. Finally, etiologic fractions for polytomous categories are calculated as:

\[ EF_i = \frac{P_i \times (RR_i - 1)}{\sum P_i \times (RR_i - 1) + 1} \]  

(4.12)

following Kramer et al. (2000) and Miettinen (1974). Here, \( P_i \) is the prevalence of the \( i^{th} \) category and \( \sum \) indicates summation over all the \( i \) categories. \( EF_i \) denotes the proportion of all cases of the outcome in the total population that can be attributed to category \( i \).

4.2.3 DATA: SOURCES AND COMPLICATIONS

The situation in the developed world, i.e. in the EME region, is described and analysed in this study on the basis of secondary data. The study aims to expand and complement demographic data sources by using data from sources in the health sciences. Most data were obtained from literature in fields such as obstetrics and gynaecology, neonatology, paediatrics, genetics, cardiology, epidemiology, and public health. Sources included articles in journals, reports, dissertations, and books. A helpful tool in the search for literature was the Medline database of the National Library of Medicine in the United States (see PubMed, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed). Preference was given to the most recent publications and data available. In addition, publications on population-based studies were preferred over studies on selective groups such as hospital populations. Valuable information was also obtained from some studies that included multiplets. Although the focus here is on singletons.

The population-based studies were generally based on vital registration systems, specific registries, and/or surveys. Examples are the Medical Birth Registry of Norway (Bakketeig et al. 1978), the All Wales Perinatal Survey (Carljidge and Stewart 1995), and the 1992 Trent Regional Epidemiological Survey of Perinatal Mortality in the UK (Gardosi et al. 1998). In Victoria, Australia, the Consultive Council on Obstetric and Paediatric Mortality and Morbidity compiled data from various sources, including the Survey of Perinatal Deaths in Victoria, the Newborn Emergency Transport Service, and the Congenital Malformations/Birth Defects Register (CCOPMM 2000). Many regions have population-based surveillance systems for congenital anomalies, such as the Manitoba Congenital Anomalies Registry in Canada (Evans et al. 1992) and the Metropolitan Atlanta Congenital Defects Program (MACDP) in the United States (Roberts et al. 1995). In Europe, EUROCAT (European Registration of Congenital Anomalies) carries out an epidemiologic surveillance of congenital anomalies. EUROCAT is a network of population-based registries for children with congenital anomalies in various countries and regions of Europe (EUROCAT 2002). However, EUROCAT does not register all congenital anomalies but applies an internationally operated “list of cases for exclusion”. Excluded are anomalies that are relatively mild or do
EARLY LIFE CHANGES

not require treatment in general, e.g. spina bifida occulta uncomplicated, glandular hypospadias, and single umbilical artery (Cornel et al. 1993b). Besides the literature, data were obtained from a database of the National Center for Health Statistic (NCHS) in the United States. The NCHS publishes cd-roms presenting linked birth-death data by birth cohort for the USA. For the present study, data from the 1988 birth cohort were available (NCHS 1995). The information in the database includes sex, birth weight, gestational age at birth, age at death, and also cause of death. However, in general, the present study does not consider cause-of-death data. It has been acknowledged that death certificates often contain inaccurate information on causes, especially when the individual making out the certificate did not know the deceased (Rockett 1994). A study in Connecticut (USA) compared autopsy reports and death certificates and found major disagreements that would require reclassification of the cause of death on the death certificates in as many as 29% of 272 cases (Kirchner et al. 1985 cited by Young 1998). Besides diagnostic accuracy, the issue of single vs. multiple causes of death also affects cause-of-death statistics. Traditionally, mortality statistics have been based on a single cause for each death (Young 1998). However, the decision to choose one contributing factor as the single underlying cause over another contributing factor “may be quite arbitrary” (Rockett 1994, p.12). Therefore, the present study focuses on the presence or absence of the selected risk factors rather than on causes of death.

During the study, several drawbacks with the sources and the information acquired from them became apparent. The subsections below discuss the most important observations in this respect.

Fragmentation and disparity

The information required to assess the relative importance of each of the selected risk factors in relation to foetal loss, stillbirth, and neonatal death was not readily available. Data were limited and fragmented and most publications provided none or only part of the data required to calculate the RR, AR(E), and EF. The majority of publications had been written for objectives other than those at the centre of the current study. Moreover, a large number of studies focused on very specific population groups, risk factors and/or outcomes, e.g. children from diabetic mothers, children born after in-vitro fertilisation, very low birth weight (< 1,500 g), spina bifida, perinatal mortality, in-hospital mortality, and/or school-age developmental outcome.

The literature reflects that research interests have changed over the years as new knowledge, technologies, treatments, and therapies have become available and survival improved. In the developing world, the availability, quality, and reliability of data are often less and publications still seem to focus more frequently on general health and survival, and on health care. In the developed world, the medical and epidemiological literature pays most attention to those groups with the highest risks, such as VLBW and extremely preterm infants. Moreover, many studies on risk factors focus only on high-risk groups without comparing them to unaffected persons (i.e. the reference group) or the total population. Furthermore, modern technologies have triggered extensive new discussions and debates. For instance, on induced abortion after prenatal diagnosis of an anomaly, and about choices between life and
death for extremely preterm infants. Much of the literature found focuses on topics related to these issues.

Despite the huge number of publications on foetus and neonate, and on pregnancy and birth outcome in relation to survival, the data of use for the present study turned out to be limited. Part of the reason why some publications could not be used was the manner in which the authors presented their findings and results. For example, some researchers only presented end results such as regression coefficients. Also, perinatal mortality was often not split into a stillbirth and an early-neonatal-death component. Similarly, researchers generally failed to specify ‘in-hospital’ mortality by age. In addition to all of this, some authors lacked clarity when it came to the presentation of figures. For example, it was not always clear whether data referred to all births (both live and stillbirths) or only to live births.

In the present study, the data that were obtained are tabulated to allow comparing them and to reach general assumptions and point estimates. However, comparison is complicated by disparities among the studies. Moreover, results sometimes show considerable differences even though they are measurements of the same (e.g. prevalence in live births, incidence of neonatal death). Some of these differences can be explained by disparities between studies in study design and methodology. These include issues such as data sources, definitions, inclusion and/or exclusion criteria, sample size, diagnostic procedure, and available technology. For example, studies on congenital heart disease (CHD) show large differences in birth prevalence. CHD is difficult to recognise and is often not detected until later in life. As a result, prevalence estimates are affected by duration of follow-up, methods of ascertainment, diagnostic procedures, and degree of paediatric cardiology experience (Buskens 1994; Hoffman 1995a; Shibuya and Murray 1998c). In addition, differences in the definition of CHD are said to play a role. Given these factors, Hoffman (1995a, p.103) believes that data on CHD collected more than 10 to 15 years ago are less accurate and complete than more recently collected data. However, variation in outcomes between studies may also reflect ‘true’ differences that can be explained by the population under study (rather than by disparities in study design and methodology). Factors that may affect results include year of study, (gestational) age, geographic region, race and ethnic background.

The sources, and the information acquired from them, thus had several drawbacks. Some of the problems described above have also been pointed out by other authors such as Kalter (1991). Kalter reviewed published reports on the relationship between perinatal mortality and congenital malformations. Problems that he observed included the following: a multitude of perspectives, inconsistent definitions, fragmentary data, infrequent specification, and variability in age, weight, and size criteria.

Antepartum data

The most problematic data were those on the antepartum period (i.e. gestation). Despite developments in technology, the gestational period is virtually inaccessible to observation. This is especially true for very early gestation, when the woman may not even be aware of her pregnancy. As a result, no vital registration exists of spontaneous abortion, and antepartum data are very limited. Fortunately, the present study only requires antepartum data on congenital anomalies, IUGR, and spontaneous abortion. Data on stillbirth appear to be more
available since late foetal deaths are regarded and registered as vital events (see Van der Veen 2001).

During gestation, the ways to examine a foetus and to detect abnormalities are limited. Therefore, foetal characteristics such as anomalies and growth retardation often remain concealed until pregnancy termination, i.e. spontaneous abortion, induced abortion, stillbirth, or live birth. Nevertheless, over time, modern technologies and prenatal diagnostic tests such as ultrasonography, amniocentesis and chorionic villus sampling, have increased our knowledge about the foetus. However, diagnostic methods are unable to detect all types of anomalies. Their use is mainly restricted to chromosomal anomalies although amniocentesis can also detect neural tube defects. Furthermore, the tests themselves may affect the risk of spontaneous abortion. Studies and publications on prenatal diagnostic tests frequently focus on the diagnostic or predictive ability of the test or method, rather than on the health status of the foetuses in the population. With regard to IUGR, the size of the embryo or foetus may be assessed with the help of ultrasonography but subsequently needs to be compared to a standard. This standard is a prerequisite in distinguishing ‘abnormal’ from ‘normal’. In an ideal situation, it would be based on the individual growth potential of the foetus (also see Chapter 3). The discipline of biometry is in search of models to predict size and to individualise growth assessment. However, publications in this field turned out not to be useful here.

The majority of antepartum data on congenital anomalies and IUGR/SGA are outcome-based prevalence proportions assessed after termination of pregnancy. Incidence of anomalies and growth retardation among embryos or foetuses in utero are beyond observation. Prevalence of anomalies and IUGR in the population in utero may be approximated by prevalence proportions among induced abortions and hysterectomies. However, maternal age distribution is likely to affect the results of these studies. The age distribution of women who undergo induced abortion often differs from the age distribution of women who continue their pregnancy. Among women who opt for induced abortion, the age groups at both extremes (i.e. the youngest as well as the most advanced ages) are likely to be overrepresented. In addition, the risk of congenital anomalies is higher among mothers in the advanced age groups. Furthermore, the existence of selective abortion might corrupt the prevalence estimates. However, prenatal diagnosis and selective abortion are quite recent phenomena, especially on a large scale. In many older studies, selective abortion is likely to play no or only a minor role.

Prevalence of anomalies and IUGR in the population in utero may also be estimated by combining outcome-based prevalence proportions with knowledge about overall foetal survival. In the literature, several authors including Creasy and Alberman (1976), Shiota (1993), and Hoffman (1995b) have reconstructed cohorts of conceptuses by combining outcome-based prevalence figures with incidence figures of foetal loss. Creasy and Alberman (1976) used the foetal life table from French and Bierman (1962) to estimate the prevalence of ASB-type malformations\(^1\) in a population in utero, and to assess the incidence of foetal loss

---

\(^1\) Anencephaly, craniorachischisis, exencephaly, encephalocele, or spina bifida cystica, but no demonstrated chromosome anomaly (Creasy and Alberman 1976).
among malformed cases. Likewise, Hoffman (1995b) combined assumptions on foetal loss with prevalence proportions of congenital heart disease, and Shiota (1993) also reconstructed a cohort of conceptuses to study the effect of malformations on loss. However, foetal survival analysis is sensitive to data complications. Problems include ascertainment of pregnancy, entry to the population under study, missed abortion, and censoring (also see Chapter 5).

4.2.4 METHODS
As discussed in Section 4.2.1, the relative importance of the selected risk factors in relation to survival is to be assessed in this study on the basis of incidence of loss/death, RR, prevalence of the risk factor in the population, AR(E), and EF. However, these measures were not readily available since data were fragmented and limited (see Section 4.2.3). Fortunately, prevalence and incidence proportions were quite often available or could be computed on the basis of published data. Still, prevalence proportions were mainly outcome-based, i.e. assessed at specific outcome points such as spontaneous loss, stillbirth, live birth, and neonatal death. Moreover, most sources of information did not provide RR, AR(E), and EF or sufficient data to estimate them. Consequently, the search for data focused on the information that was available, and alternative methods were designed and applied to reach estimates for the EME region as a whole. Very briefly, the present study combined various assumptions and types of measures derived from secondary data sources with mortality figures from a hypothetical cohort, which were obtained by life table analysis. For more information, see the subsections below that provide a more extensive discussion of the methods used.

Life table analysis and construction of a hypothetical cohort
Various foetal and neonatal life table data from EME countries have been published over the years. The life tables indicate the timing of loss/death during gestation and the neonatal period as observed in the population in general, regardless of the presence or absence of specific risk factors. In the present study, risk-factor-specific prevalence and incidence proportions are combined with these overall mortality figures. Two life table data sets are selected from published materials: one to construct a foetal life table and one to make up a neonatal life table (see Chapter 5).

Some authors, such as Bakketeig et al. (1978) and Kristensen and Mac (1992), have published data for foeto-infant, or foetus-infant, life tables that combine antenatal and postnatal outcomes. However, the data available from the literature to construct such a combined life table are very limited and dated. Moreover, a combined life table would somehow have to indicate the transition from in utero to life outside the womb. Bakketeig et al. (1978) used absorbing states and mentioned age at death after live birth in their designation of the states. Their multiple-decrement life table contains headings such as ‘foetal death prior labour’, ‘live birth, death 1-6 days’, and ‘live birth, survived 1 year’, set against age in completed weeks of gestation. In the present study, the two life tables are combined by taking the total number of live births from the foetal life table as radix for the neonatal life table (see Chapter 5). In this way a hypothetical cohort can be created that is followed from early gestation until the end of the neonatal period. The link between the two tables also emphasises that we should be aware that babies of the same age do not necessarily have the same
**EARLY LIFE CHANGES**

*gestational* age, i.e. they may not have lived the same total amount of time. Indeed, persons with identical age, or at the same point in duration since the last demographic event, may be at significantly different levels of biological and psychological development (Willekens 1997 cited by Mills 2000).

The constructed life tables include the following events: spontaneous foetal loss/stillbirth, live birth, and death during the neonatal period (see Chapter 5). As stated earlier in Section 4.2.1, the current study does not take an interest in induced abortion as an adverse survival outcome and therefore, it was decided to exclude induced abortion from the foetal life table as a competing risk. However, the book does present figures on the prevalence of pregnancy outcomes among induced abortions since they provide an indication of prevalence in the population in utero.

Although, the life tables and the hypothetical cohort are based on only two data sets, they are assumed to represent the general structure of foetal loss and neonatal mortality as observed in the EME region. Both data sets and life tables are discussed extensively in Chapter 5. In addition, Chapter 5 deals with life table methodology and the construction of the life tables and the hypothetical cohort.

**Dichotomous and polytomous variables**

Survival status is clearly a dichotomous variable since a person is either dead or alive. However, the selected risk factors (pregnancy and birth outcomes) can be operationalised as either dichotomous or polytomous variables. For some risk factors, e.g. congenital anomalies, IUGR/SGA, and birth asphyxia, a division into two categories seems most obvious although these variables could also be split into multiple categories, such as severe, moderate, mild, and no anomalies/IUGR/asphyxia. The variables birth weight and gestational age at birth are by nature continuous and provide a more logical opportunity to use polytomous categories. Indeed, many publications split these factors into multiple categories. However, the classification of these categories is not consistent between studies and some publications only focus on certain groups of infants (i.e. the lowest weights or the most preterm cases). Moreover, the use of polytomous variables is likely to limit the information available for calculations and estimations. Dichotomous variables, on the other hand, require less detailed data, which are likely to be more available.

Therefore, the present study focuses on dichotomous variables. The following dichotomous situations are compared: anomalous versus non-anomalous, LBW (< 2,500 g) versus non-LBW (≥ 2,500 g), IUGR/SGA versus non-SGA, preterm (< 37 weeks) versus non-preterm (≥ 37 weeks), and asphyxiated at birth versus not asphyxiated at birth. Heavy weight at birth, postterm birth, and large-for-gestational-age (LGA) are not among the selected risk factors. Some polytomous results are obtained for preterm birth, and to a limited extent also for low birth weight and IUGR.

In the EME region, survival chances of LBW and preterm neonates have improved considerably over time and with that, research interests have shifted to increasingly lower weight and gestational age groups. Moreover, the customary birth weight and gestational age limits of 2,500 g and 37 weeks are arbitrary and their values can be questioned when it comes to distinguishing high-risk from lower-risk neonates. Therefore, dichotomous variables are
also created for low birth weight defined as < 2,000 g, for VLBW defined as < 1,500 g, and for very preterm birth defined as < 32 weeks.

Alternative ways to estimate the relative risk and the attributable risk

Following equation (4.4), the calculation of the relative risk of death requires knowledge about both incidence of death in the group with the risk factor and the incidence among those without the risk factor. Or, as equation (4.5) shows, information is needed on the total numbers in both populations and on the numbers of death in both groups. However, as has already been noted above, data are not readily available and consequently, alternative methods have to be applied to estimate RR, AR(E), and EF.

Suppose, \( B \) is the total number of live births in a population and \( D \) represents the total number of neonatal deaths. Within this population,

\[
B = B_1 + B_0 \tag{4.13}
\]

and

\[
D = D_1 + D_0 \tag{4.14}
\]

where \( B_1 \) and \( D_1 \) refer to the group in the population that is affected by the risk factor while \( B_0 \) and \( D_0 \) reflect numbers among those without the risk factor. Subsequently, equations (4.4) and (4.5) could be rewritten as

\[
RR = \frac{D_1/B_1}{D_0/B_0} \tag{4.15}
\]

indicating the relative risk of neonatal death. If only totals \( B \) and \( D \) are known in a population, the relative risk of neonatal death could be estimated by combining these totals with at least two of the following three measures: (1) prevalence of the risk factor at live birth, \( p^B_1 \), (2) incidence proportion of neonatal death among live births affected by the risk factor, \( I_1 \) (which equals \( D_1 \) divided by \( B_1 \)), or (3) prevalence of the risk factor in neonatal deaths, \( p^D_1 \). Several relationships exist between the variables. For example, \( B \) could be computed as:

\[
B = B_1 + B_0 = [B * p^B_1] + [B * (1 - p^B_1)] \tag{4.16}
\]

where \( B * p^B_1 \) equals \( B_1 \) and \( B * (1 - p^B_1) \) equals \( B_0 \). Similarly,

\[
D = D_1 + D_0 = [D * p^D_1] + [D * (1 - p^D_1)] \tag{4.17}
\]

while \( D_1 \) also equals \( B_1 \) multiplied by \( I_1 \).

The relative risk could be estimated by repeatedly combining two of the three measures given above (i.e. \( p^B_1, I_1 \), or \( p^D_1 \)) with both \( B \) and \( D \). For instance, if only \( B, D, p^B_1 \) and \( I_1 \) are known, the relative risk could be calculated as follows:

\[
RR = I_1/I_0 = I_1/[(D - D_1)/(B - B_1)] \tag{4.18}
\]

where \( D_1 = B * p^B_1 * I_1 \), and \( B_1 = B * p^B_1 \). When only \( B, D, I_1 \) and \( p^D_1 \) are known, equation (4.18) can also be applied but then \( D_1 \) needs to be computed as \( D * p^D_1 \) and \( B_1 \) as \( D_1/I_1 \). When only \( B, D, p^B_1 \) and \( p^D_1 \) are known, the equation to calculate relative risk should be rewritten to exclude reference to \( I_1 \). Then,

\[
RR = [D_1/B_1] / [(D - D_1)/(B - B_1)] \tag{4.19}
\]
EARLY LIFE CHANGES

where \( D_i = D * p^{D}_i \) and \( B_i = B * p^{B}_i \).

All the equations above pertain to the estimation of the relative risk of neonatal death. The relative risks of early neonatal death, spontaneous abortion, and stillbirth can be obtained in a similar way but require a different interpretation of the symbols. For early neonatal death, only a minor adjustment in interpretation is required: \( D \) needs to be regarded to reflect the total number of early neonatal deaths instead of neonatal deaths. The relative risks of spontaneous abortion and of stillbirth require a more extensive redefinition. First, all references to neonatal death (in \( D, I_i \) and \( p^{D}_i \)) should be interpreted as references to either spontaneous abortion or stillbirth. For instance, in the calculation of the RR of spontaneous abortion, \( p^{D}_i \) reflects prevalence of the risk factor among spontaneous abortions (instead of neonatal deaths). Likewise, all references to live births (in \( B, p^{B}_i \) and \( I_i \)) should be replaced by alternative interpretations of the population at risk of experiencing the event. The population at risk of spontaneous abortion consists of all conceptions. Concerning stillbirth, this population is dependent on the definition of stillbirth or late foetal death. When the defining age limit is set at 28 gestational weeks, the population at risk of stillbirth consists of all foetuses in utero at the beginning of gestational week 28, i.e. pregnancies still in progress at 28 weeks. Thus, probabilities here are conditional probabilities. As a result, \( I_i \) of stillbirth should be interpreted as the incidence proportion of stillbirth among exposed or affected foetuses in utero at the beginning of gestational week 28. Similarly, in the case of spontaneous abortion, \( I_i \) indicates incidence proportion of spontaneous abortion among conceptions exposed to or affected by the risk factor.

With regard to attributable risk, equation (4.8) shows that the attributable risk among the exposed, \( AR(E) \), is quite simple to calculate once the relative risk is known. However, the etiologic fraction, \( EF \), requires additional information, besides RR, on the prevalence of the risk factor. Indeed, equation (4.10) could be rewritten as:

\[
EF = \frac{p^{B}_i * (RR - 1)}{p^{B}_i * (RR - 1) + 1}
\] (4.20)

However, calculation of the EF is also possible without \( p^{B}_i \) and \( RR \): on the basis of \( B, D, I_i \) and \( p^{D}_i \). Following equation (4.9),

\[
EF = \frac{I - (D * [1 - p^{D}_i]) / (B - B_i)}{I}
\] (4.21)

where \( I = D / B \) and \( B_i = (D * p^{D}_i) / I \).

From input to output

The quality and reliability of estimations are greatly affected by the quality and reliability of the data that are used to obtain them. In other words, output depends on input. In the present study, \( B \) and \( D \) in the equations in the previous subsection are known from the hypothetical cohort and life tables. Values of the other measures, i.e. \( p^{B}_i, I_i, p^{D}_i \), and when available also \( I_0 \) and/or \( RR \), are derived from various publications.

For each measure, values that could be obtained were tabulated to compare them amongst themselves. As was already discussed in Section 4.2.3, the results show differences that are on some occasions considerable. The study aims to estimate the situation for the EME region as a whole, and not by country, region, or ethnic group. Consequently, general
assumptions have to be made about the levels of the measures in the EME region, without being population- or study-specific. In order to serve as input data, the ranges of data observed in the combined literature are restricted to only one or two estimates or assumptions. These estimates and assumptions are based on mean and median, but information about the studies and their design is also taken into account. Important information in this respect includes definitions, inclusion criteria, type of study (e.g. hospital-based or population-based), and population under study (e.g. gestational age). In addition, importance is attached to previous review publications, including Kalter (1991), Hoffman (1995a, 1995b), Murray and Lopez (1998), and Van der Veen (2001).

Repeatedly, two of the three assumptions for \( p^B, I, \) or \( p^D \) are combined with \( B \) and \( D \) from the hypothetical cohort and entered into equations (4.18) and (4.19) to estimate the RR. Subsequently, equations (4.8), (4.20), and (4.21) are used to estimate the AR(E) and the EF. Since three different combinations of data input are possible\(^2\), the results consist of three sets of relative and attributable risks. In an ideal situation, where the particular combinations of \( B, D, \) and \( p^B, I, \) or \( p^D \) are realistic, these three sets would have equal values. In the present study, they frequently provide a range of values for the RR, AR(E), and EF. The plausibility of the particular data combinations is further checked by calculating the one measure – either \( p^B, I, \) or \( p^D \) – that is missing from the input data from the others. For instance,

\[
p^B = \frac{[p^D \cdot D]}{[I \cdot B]} \tag{4.22}
\]

expresses the calculated prevalence of the risk factor among live births. Similarly, the incidence proportion of loss/death among those affected by the risk factor equals:

\[
I = \frac{[p^D \cdot D]}{[p^B \cdot B]} \tag{4.23}
\]

and

\[
p^D = \frac{[I \cdot p^B \cdot B]}{D} \tag{4.24}
\]

indicates the prevalence of the risk factor among losses/deaths.

In the present study, the values obtained using equations (4.22), (4.23), and (4.24) often do not equal the initial assumptions. This implies that the particular combinations of data in the input assumptions are not possible for the hypothetical cohort. Subsequently, the results consist of an estimated range of values for RR, AR(E), and EF. Overall, it should be noted that the results of the study do not reflect absolute, ‘true’ figures. They consist of ranges of estimates. Nevertheless, they are sufficient for indicating the relative importance of the selected risk factors in the EME region for the processes of spontaneous abortion, stillbirth, and neonatal death.

### 4.3 Part II: Regions in transition

What is the significance of the selected risk factors within the framework of an epidemiologic transition as based on the current situation in regions that are approaching the later stages of

\(^2\) (1) \( B \) and \( D \) in combination with \( p^B \) and \( I \), (2) \( B \) and \( D \) in combination with \( I \) and \( p^D \), and (3) \( B \) and \( D \) in combination with \( p^B \) and \( p^D \).
EARLY LIFE CHANGES

the epidemiologic transition? This second research question also has several subquestions: What is the situation in regions in transition with regard to loss and death during gestation, birth, and the neonatal period and the selected risk factors? What is the relative importance of the selected risk factors measured at the level of the population? And, measured at the level of the population? When put alongside figures from the developed world and figures from regions that are less advanced in the transition, what do these results disclose about epidemiologic transition, or about health transition? Does the relative importance of the selected risk factors change during the epidemiologic transition?

The objectives of the second research question are thus twofold: (1) to assess the situation in regions in transition that are approaching the later stages of the epidemiologic transition and (2) to place the selected risk factors in the framework of the epidemiologic transition. These issues will be addressed in Chapters 8 and 9. The research providing the answers to the questions is based on the data and methods that are described below.

4.3.1 CHOICE FOR A HOSPITAL SURVEY IN A CASE REGION

In this second part of the research, the regions of interest are ones that have not yet reached the final stages of the epidemiologic transition as observed in the EME region and in other developed countries. They are regions that are part of the developing world. However, they are much more advanced in the epidemiologic transition than most of the least developed regions, such as most countries in Sub-Saharan Africa. In other words, the regions of interest are regions in the midst of transition. They are likely to have experienced recent changes in health and survival and will face additional changes within the near future. Their situation may be complicated by the coexistence of different categories of diseases and causes of death since they frequently face a double burden, an overlap of eras, or countertransition (see Chapters 1 and 2). Examples of regions approaching the later stages of the epidemiologic transition include South India and Mexico (Den Draak 1999; also see Chapter 2).

For most less developed countries, data on health and survival are sparse if available at all. Susser et al. (1985 cited by Kline et al. 1989, p.241) note that data and economic development go hand in hand. Indeed, a glance in the Demographic Yearbook by the United Nations shows that data on perinatal and infant mortality are missing for many developing countries (see also Chapter 2). Vital registration data are often missing or are limited. During recent decades, the main sources of information on child survival in developing countries have been large-scale demographic surveys, such as the Demographic and Health Survey (DHS) and the National Family Health Survey (NFHS) in India. However, these surveys have been found to provide insufficient information to explain child survival levels and the mechanisms underlying child health and survival (Boerma 1996; Padmadas 2000). Other, small-scale surveys and studies, if available, could provide more detailed information but only to a limited extent, especially if one is searching for specific data about particular outcomes.

The present study tries to deal with the shortfall in data in an effective way by carrying out a hospital survey in a case region. Although the contribution by case studies is limited when it comes to statements with statistical generalisation, they are able to provide insights that are valuable for the further development of theories (i.e. analytic generalisation) (Yin
1994). The present hospital survey is intended to expand and complement demographic data sources. A large population- or community-based survey that includes health professionals visiting houses, was considered not to be feasible due to time, logistic, and financial limitations. However, the results of the hospital survey are compared to, and complemented by, vital registration data, data from large-scale surveys, and data obtained from medical and epidemiological literature. The choice for Kerala as a case region and illustration of a region in transition, and the choice of the hospital, were based on the data presented in Chapter 2 and on a feasibility study in South India.

4.3.2 The feasibility study

In December 1998, a feasibility study was undertaken in South India, in both Karnataka and Kerala (Den Draak 2000). The objective of the study was to explore the possibility of collecting hospital data in the south of India. During this study, interviews were conducted with several researchers and medical doctors from institutes and hospitals in Dharwad and Hubli (Karnataka) and Thiruvananthapuram (Kerala). Attention was paid to factors such as the position of the region in the epidemiologic transition, the potential size of the survey, the representativeness of the survey, and the availability of data. Ultimately, on the basis of the feasibility study, the staff of Sri Avittom Thirunal (SAT) Hospital in Thiruvananthapuram (Trivandrum), a large hospital in the public sector, were contacted and kindly offered their cooperation.

On the basis of the figures in Chapter 2, it was concluded that the states of Karnataka and Kerala in South India are both regions in transition. However, Kerala seems to be more advanced in terms of the epidemiologic transition than Karnataka. Moreover, the majority of people in Kerala are literate and aware of their rights for health and good-quality health care. Most pregnant women make use of antenatal care and/or deliver in health institutions (see Section 2.4.2). For 1990, the Sample Registration System (SRS) estimated that 89.8% of deliveries were institutional (Ministry of Health and Family Welfare 1994). According to NFHS-1, 88% of deliveries took place in health facilities and institutions (PRC, Thiruvananthapuram and IIPS 1995) while in Karnataka it was only 37.5% of births (PRC, Bangalore and IIPS 1995). Kerala figures for more recent years are even higher. NFHS-2 estimated 93% of births in Kerala to occur in health facilities compared to only 51% in Karnataka (IIPS and ORC Macro 2001a, 2001b). The Rapid Household Survey in Kerala in 1998-1999 even estimated as high as 97% (IIPS 2001 cited by IIPS and ORC Macro 2001b). The high proportion of births occurring in health institutions implies that, in Kerala, data from hospitals and clinics on births represent the state’s population in a rather complete way.

Nevertheless, within Kerala, hospitals in the public sector see different population groups than private hospitals and clinics. Similar to Karnataka, government hospitals in Kerala are more likely to be visited by poorer and less educated people from lower socioeconomic classes. In addition, private hospitals are said to refer many complicated cases to government hospitals in fear of litigation and bad publicity. Therefore, a sample from a Kerala government hospital is unlikely to be representative of the total population of pregnant women in the state.
EARLY LIFE CHANGES

Nevertheless, large teaching hospitals run by the government have several advantages for research purposes over private hospitals and smaller public hospitals. First, academic teaching hospitals have more experience with research. They are also able to provide more complete data records and often have a large number of facilities to do tests, including laboratories. Private hospitals may be reluctant to provide data out of fear for bad publicity. Finally, a large teaching hospital enables the collection of information on a large number of cases within a relatively short time span. Moreover, their increased role in complicated cases should result in a survey containing a larger number of complications and deaths that are of interest for this study. For further information on the feasibility study and its results, please refer to Den Draak (2000).

4.3.3 INTRODUCTION TO SRI AVITTM THIRUNAL HOSPITAL

Sri Avittom Thirunal (SAT) Hospital is a major teaching hospital specialising in Gynaecology, Obstetrics, and Paediatrics and is part of Medical College, Thiruvananthapuram. The mother and child hospital has three major departments: Obstetrics and Gynaecology, Paediatrics, and Paediatric Surgery. It began in 1950 as an 80-bed children's hospital. In 1952, SAT Hospital (SATH) was officially inaugurated and a maternity unit was added during the same year. Since then, the number of services and patients has been steadily rising over five decades. Starting with only 1,000 admissions in 1951, the hospital is nowadays overcrowded and admits and treats 50,000 inpatients and 250,000 outpatients each year. Bed capacity is only 786 while the occupancy ‘rate’ is about 150% (SAT Hospital Website 2001). On average, 45-46 deliveries per day take place in SAT Hospital (based on unpublished hospital statistics). The hospital covers a wide geographic area since it serves as a so-called ‘referral hospital’. However, there are many patients who come directly without being referred.

The subsections below briefly describe the organisation and registration in SAT Hospital concerning antenatal care, care during delivery, and neonatal care. This provides a clear picture of the data that are available in the hospital, their origin, and their reliability.

Antenatal care and check-ups

SAT Hospital provides antenatal check-ups and care. At the outpatient clinic, pregnant women receive their check-ups. As with the rest of the hospital, the outpatient clinic is often overcrowded. Before the actual antenatal check-up, basic information is registered on an OP (outpatient) card and on an antenatal chart. This information includes: name, address, age, income, family history, number of years married, LMP (date of first day of last menstrual period), previous obstetric history, date of antenatal examination, and weight at examination. The data with regard to medical and obstetric history are based on the information that the women give during the interview rather than on hospital records. Expected date of confinement (EDC) is set at LMP plus nine months and seven days. Also, depending on the day of the week that the woman visits the outpatient clinic, she is assigned to one of the six maternity units (O1 through O6) of the hospital.

After this registration, the women see several doctors and hospital staff members for the actual antenatal check-up and care. Blood pressure, weight, fundus height, presentation
and position of the foetus, and foetal heart are preferably checked at each antenatal visit. Ideally performed at least once during the pregnancy are: the assessment of blood group and rhesus type, the measurement of haemoglobin level, an urine test for albumin and sugar, a Glucose Challenge Test (GCT) for gestational diabetes, a Veneral Disease Research Laboratory (VDRL) test for venereal diseases, and a HBsAg (hepatitis-B surface antigen) test for hepatitis B. In addition, the majority of women have at least one ultrasound scan (USS). During the antenatal visits, the women receive one or two tetanus injections. Furthermore, iron and folic acid supplements as well as calcium tablets are prescribed and distributed free of charge.

The antenatal charts, designed in the 1960s, are not kept in SAT Hospital but the women take the charts home with them and are supposed to bring them each antenatal visit. A new antenatal chart is initiated when the former chart is full or when the woman has forgotten to bring it. The doctors advise the women to come regularly for antenatal check-ups according to the following schedule: monthly until 28 weeks of gestation, every fortnight until 36 weeks, and after that weekly. High-risk cases are advised to have check-ups even more frequently: every fortnight. On the basis of the number of antenatal visits, patients are classified into three categories. Women who have come to SAT Hospital at least twice during their pregnancy are regarded as booked cases. Women who have had two or more antenatal visits but in a different hospital or clinic are the so-called booked-outside cases. Finally, women who had less than two antenatal check-ups with a doctor during their pregnancy are called non-booked. However, some high-risk patients who were hospitalised for a period during their pregnancy may have only one antenatal visit written down on their chart but are also regarded as being booked.

The delivery

Women who come to SAT Hospital for the delivery of their baby are admitted into the hospital and become inpatients. All inpatients are awarded an IP (inpatient) number by the hospital administration. In addition, a case record is kept for each inpatient in which health status and some personal information will be entered during the patient’s stay in SAT Hospital.

In general, women who come for the delivery of their baby are initially admitted into the first stage labour room. These women include women with labour pain, women who are bleeding or leaking amniotic fluid, past date cases, and cases referred by the outpatient clinic. During an interview, information about the health of the woman, the obstetric history of the pregnancy, results of previous tests and investigations, and other medically relevant information is registered in the case record. The majority of this information is copied from antenatal charts and notes of test results. If necessary and when possible, additional tests and investigations are performed in SAT Hospital before delivery. Subsequently, a decision is made about the mode of delivery, such as whether an induction of labour or a caesarean

---

3 As a consequence, modern tests and measures, such as ultrasound scans, are not listed on the chart while the form contains some entries that are no longer used.
EARLY LIFE CHANGES

section is necessary. Once the membranes are ruptured and the labour is progressing well, or a caesarean section is needed, the woman is transferred to the second stage labour room.

The second stage labour room has an ‘emergency’ operating theatre attached where caesarean sections can take place. However, the operating theatre is fully occupied, limiting the possibilities in the case of a real emergency. Hence, the doctors and the hospital have to avoid taking risks. The Department of Obstetrics and Gynaecology also has a second operating theatre at its disposal: the ‘major operating theatre’. This operating theatre may be used to perform elective caesarean sections, i.e. predetermined sections. Women who are scheduled for a predetermined caesarean section go directly from their hospital ward to the operating theatre; they are not admitted to the first stage labour room. Finally, women with infections and other complicated cases, deliver their baby in the septic labour room. Cases of postpartum haemorrhage (PPH) are also transferred to this labour room.

There are thus several options as to where the birth can take place. SAT Hospital has no private rooms available for women to deliver their baby. Furthermore, the labour rooms only allow patients and medical staff; bystanders, i.e. family and friends of the women, are not allowed in the labour rooms. Once a month the second stage labour room is thoroughly cleaned and its activities are transferred to the first stage labour room.

In SAT Hospital, births are distinguished from abortions on the basis of the following criteria: birth weight \( \geq 1,000 \text{ g} \) and gestational age \( \geq 28 \text{ weeks} \). All births in SAT Hospital, regardless of where and how they take place, and their outcome, are registered in a register of deliveries, the Obstetric Register (OR), and are awarded a serial or OR number. Abortions are registered in the so-called abortion files and generally not in the OR register. The OR register contains the following data: OR number, IP number, case status (booked, booked outside or non-booked), unit number, date and time of admission, name and address, age, income, summary of past obstetric history, mother’s haemoglobin level and blood group, nature of labour (normal or abnormal), presentation and position of the foetus, gestational age (full term or preterm), outcome (born alive, intrauterine death, or stillbirth), sex, weight, head circumference, placental weight, date and time of birth, name of doctor, mode of delivery (normal vaginal delivery, or caesarean section etc.) and indication, any induction of labour, any complications, Apgar score, any congenital anomalies, and any other remarks. Car deliveries, that is deliveries that took place on the way to SAT Hospital, are also entered in the OR register. The OR register is located in the second stage labour room.

Naturally, notes about the delivery and its progress are also registered in the case record of the patient. After delivery, a so-called mother-baby chart is routinely entered in the case record. This chart provides a sort of summary and contains the following information about the baby: date and time of delivery, sex, weight, Apgar score, whether cried immediately or not, any distress, and any congenital anomalies. With regard to the mother, the data registered in this chart are: age, gravida, LMP, EDC, type of delivery, haemoglobin level, blood group, and any complications.

Care for the newborn and stay in SAT Hospital

After birth, a nurse cleans and weighs the baby. In the second stage labour room and in the operating theatres, the weighing scales are electronic. However, the weighing scale in the
A septic labour room is mechanical. During the first half-hour, the newborn is kept under a lamp to keep warm before going to its mother.

A doctor checks the baby after birth to see whether it has any health problems or anomalies. For a healthy baby without problems, the Apgar score at 1 minute is set at 9. The maximum score of 10 is no longer awarded in SAT Hospital. For babies with a poor Apgar score at 1 minute, the score is assessed again at 5 minutes and if necessary, also at 10 minutes.

In the event of any problems or complications, the paediatrician is called in. A paediatrician is always present in the second stage labour room. This paediatrician does not see all babies born in SAT Hospital, but does routinely see all babies born by caesarean section and all high-risk cases. In SAT Hospital, babies with a birth weight of less than 2.5 kg are regarded as ‘small-for-gestational-age’ (SGA). Babies weighing less than 1.8 kg are considered to be ‘extremely SGA’. For preterm babies, the paediatrician applies a growth curve ‘roughly from the mind’ to assess the relationship between size and gestational age. The gestational age of preterm babies is approximated by the paediatrician on the basis of a scoring system, such as the New Ballard Score (Ballard et al. 1991 cited by Cloherty and Stark 1998) or the scoring system by Singh et al. 1975 (cited by Singh 1993). These scoring systems base their estimation of gestational age on the external appearance of the baby and on neurological assessment.

Unhealthy babies, and babies suspected of having health problems, are transferred to either the In-Born Nursery or the Sick Nursery (see below). Healthy babies stay with their mother in one of the maternity wards. Usually, the grandmothers and other family members help out by taking care of both mother and baby. For example, they provide the new mother with food or take the baby to the paediatrician for a check-up.

In SAT Hospital, each of the six maternity units has its own ward. In addition, separate wards exist for cases with an infection, for post-caesarean cases, and for post-sterilisation cases. Women who are able and willing to spend money in return for some privacy and a more comfortable room, can stay in one of the pay wards. In a normal situation without major complications, mothers and babies are kept in hospital for one to three days after delivery. During their stay, the hospital strongly promotes breastfeeding; bottles are not allowed in the wards unless medically indicated. In addition, SAT Hospital promotes and provides immunisations for the newborn against tuberculosis and polio. Every morning, paediatricians make a round visiting all wards to check the health of the babies.

Discharges from hospital are entered in a discharge register, or paying book in the case of a pay ward. In this way, each ward daily records the IP numbers, names, and the dates of admission of those patients who are being discharged on that day. Before going home, the discharged patients also have to go to the office of the hospital security service where their IP number is recorded in another register. At discharge, the patients receive a summary report that describes their own and their babies’ health status and the events of the past days. Especially for this, a small booklet, the so-called health record, can be bought at the hospital’s coffee shop. In addition to the summary, the booklet also contains information on the office hours of the outpatient clinic as well as an immunisation schedule for the baby. The health record is useful for future check-ups.
EARLY LIFE CHANGES

Due to its policies, SAT Hospital has received the designation ‘baby-friendly hospital’. The government awarded this designation because the hospital meets several criteria such as the promotion and provision of immunisations. The hospital also promotes breastfeeding and tries to strengthen the bond between mother and child by letting the baby stay with its mother from half an hour after birth. Furthermore, the hospital has several nurseries where it provides neonatal care.

The nurseries: care for the unhealthy newborn
SAT Hospital has three nurseries: the Out-Born Nursery (OBN), the In-Born Nursery (IBN), and the Sick Nursery (SN). Newborn babies that have health problems, or are suspected of having health problems are admitted to one of these nurseries. The OBN provides care for babies that were born outside SAT Hospital. The IBN only admits so-called ‘clean’ newborns – babies born in SAT Hospital and transferred directly from the labour room or operating theatre to the nursery. This is to avoid infections. Babies who were born in SAT Hospital but have been in one of the wards after birth, or who were born in the septic labour room, are transferred to the SN in case of health problems. Newborns who need to undergo surgery stay in the Department of Paediatric Surgery. Finally, since the nurseries do not have ventilators, newborns who are in need of artificial respiration are transferred to the intensive care unit (ICU) at the Department of Paediatric Surgery.

To avoid the spread of infections, access to the nurseries is limited to nursery staff members. However, the IBN has a small office attached where doctors and nurses can be addressed and where the combined administrations of both IBN and SN are kept. When a newborn is admitted to one of the nurseries, the baby becomes an inpatient with its own IP number and case record. The information on the mother-baby chart in the case record of the mother is copied into the babies’ case record. The IBN/SN administration includes an admission register, a death register, a discharge register, and the case records of the individual patients. The admission register, death register, and discharge register contain information about the mother’s name, address, and income, and the baby’s date of birth, sex, birth weight, date of admission, date of death or discharge, indication for admission, and cause of death.

After discharge from the IBN/SN, the baby usually stays in hospital for a couple of days with its mother. Mother and child then stay in one of the maternity wards. If necessary, the baby is readmitted to the sick nursery. To be sure that the mother is present every day in the hospital and breastfeeds her baby when possible, SAT Hospital generally does not discharge mothers before their newborn babies. At discharge from the IBN/SN, the baby receives a discharge card summarising all the relevant information about its health. This discharge card is to be brought along to check-ups.

Follow-up and check-ups after discharge
Mothers and their newborn babies are advised to come to the outpatient clinic 42 days after delivery for a postnatal check-up. The check-ups take place in the office of the Family Welfare Department, which also provides services related to contraception and MTP (Medical Termination of Pregnancy). Hours for the postnatal check-up are Tuesdays and Fridays from 10.00 to 12.00 hours. However, it is possible for the women themselves to have a postnatal
CHAPTER 4: DATA AND METHODS

check-up on any day of the week (except Sundays), but the paediatrician who checks the babies is only present on Tuesday and Friday mornings. Check-ups are advised but not compulsory. In addition, it is advised that all babies have a second set of immunisations around 42 days after birth, including DTP (difteria, pertusis, and tetanus) and OPV (Oral Polio Vaccine). However, many women do not come back to SAT Hospital for their postnatal check-up but visit a doctor or clinic closer to home.

The Family Welfare staff enter some general data on the woman and the delivery in the Postnatal Register, such as name, IP number, and mode of delivery. In addition, they weigh the baby and measure its head circumference. This information is recorded in the Baby-Weight Register next to mother’s name and weight at birth. However, the IP number is not entered in the register for babies. The paediatrician comes round for about half an hour at the end of the morning. The paediatrician checks all babies individually and writes down some additional information in the Baby-Weight Register. Babies that come after the paediatrician has left are not entered in the register.

For babies who have been discharged from the IBN/SN, the Department of Paediatrics provides extra consultations on Fridays between about 11.00 and 13.00 hours. The first check-up usually takes place around two weeks after discharge. During the check-up, head circumference and weight are measured by a nurse and entered in the clinic’s register. Except for the mother’s name, no additional information is recorded in this register. A paediatrician examines the baby. The findings of the paediatrician are entered on the discharge papers that the mothers again take home.

4.3.4 DATA COLLECTION IN SAT HOSPITAL

In the period October-December 2000, data were collected in SAT Hospital for this study on the following topics: general characteristics of the woman, delivery in SAT Hospital and the pregnancy preceding the delivery, pregnancy and birth outcome, and follow-up of the neonate. The data were a combination of prospective and retrospective data. Most of the data were obtained from hospital registers and patient files or case records. However, a few questions were directly posed to the women. To allow for a simple and systematic collection and registration of the data, the information from registers and records was copied on to question forms or so-called pro formas. Furthermore, the mothers in the survey received a short questionnaire, the postnatal card, that they were asked to complete and send back once their baby had reached the age of 28 days. Additional information on the Indian context and the situation in SAT Hospital came from local informants. These informants are listed in Appendix D.

General design

The intention of the survey at SAT Hospital was to collect data on pregnancies, deliveries, and the health and survival of newborns in the neonatal period. The data that were collected in SAT Hospital at the end of 2000 included the following:

- general background information,
  e.g. maternal age, income, religion, and level of education,
EARLY LIFE CHANGES

- health status of the woman,
  e.g. height, pre-existing health problems, and use of medication,
- previous obstetric history,
  e.g. gravida, outcome of previous pregnancies, and months since last day of previous pregnancy,
- present pregnancy,
  e.g. LMP, maternal health problems and complications of pregnancy, and use of antenatal care,
- delivery,
  e.g. complications during delivery, induction of labour, and mode of delivery,
- pregnancy and birth outcome,
  e.g. sex, birth weight, gestational age at birth, and congenital anomalies,
- neonatal health and survival status,
  e.g. admission to hospital immediately after birth, indication for admission, survival, and cause of death.

The resultant data set thus contained information on both mother and child.

The target was to have a data set that would include approximately 1,000 singleton births. On the basis of the secondary data for the EME region discussed in Chapters 6 and 7, a data set of this size could be expected to include the following numbers of adverse outcomes: approximately 30 cases with congenital anomalies (including 6 chromosomal aberrations, 1-2 NTDs, and 8-10 CHD), 50-100 low birth weight cases, 60-100 preterm births, 30-100 IUGR/SGA cases, and 3-10 cases of birth asphyxia. For some of the adverse outcomes, the number of cases in the sample might be somewhat small for statistical analysis. The target of 1,000 singleton births was achieved in less than 22 days. The total survey consisted of all deliveries that took place between Monday October 9, 2000, 00.00 hours and Monday October 30, 2000, 09.00 hours. The information on the pregnancies preceding the births was collected retrospectively. The OR register served as a checkpoint to make sure that indeed all births were included in the survey. In total, the data set consisted of 1,022 births of whom 1,004 were singletons and 9 pairs of twins.

The data were obtained from several registers and records: antenatal charts, case records of the women, the OR register, IBN/SN registers, case records of the babies (if applicable), discharge registers and paying books in the maternity wards, Postnatal Register, Baby-Weight Register, and IBN/SN outpatient clinic register. The information in these data sources were linked to each other by help of the IP numbers of mother and child, woman’s name in combination with husband’s name, OR number and unit number, and birth weight, depending on what was available and what seemed to be most reliable. In the case of the IBN/SN outpatient clinic check-ups, the nurses were specially requested to add the IP number of the baby to the register. No information was collected from the abortion files on spontaneous abortion. These abortion files were not linked to the antenatal charts. Furthermore, the data on abortion would have been limited to only those women who actually had an abortion while the total population at risk of spontaneous abortion (i.e. all pregnant women) would not have been known and remain unobserved.
CHAPTER 4: DATA AND METHODS

A team of three people, including the author, collected the data. Given that most of the data were obtained from medical records the help of a medically trained person was essential. Support came from a postgraduate student in Obstetrics & Gynaecology, Dr. Sree Kala Devi S., studying at SAT Hospital/Medical College Thiruvanthapuram. The project became a collaboration as data were also collected for her thesis. As a result, some extra questions were added to the question form, such as whether or not the woman underwent a postpartum sterilisation (PPS) during her stay in SAT Hospital. Additional support came from a former student of the University of Kerala, Dhanya Padmakumary, who had recently completed her Master’s studies in Demography. Dhanya addressed the women during their stay in hospital after the delivery in the local language Malayalam to explain the short questionnaire (or postnatal card) and to ask some additional questions.

Pro Forma part A: data based on hospital records

Question form Pro Forma part A (Appendix E) was designed to collect data on general characteristics, the pregnancy, the delivery, and pregnancy and birth outcome. The form contained the following sections: identification data, general information on the woman, obstetric history and previous obstetric history, general health of the woman, antenatal care, complications in the present pregnancy, the delivery, information on the newborn, and a section providing some, limited, postnatal information. The majority of this information was obtained from the antenatal charts, the case records of the patients, and/or the OR register.

The initial plan was to collect the data on the pregnancies in the first stage labour room because quite a few antenatal charts get lost or become dirty and are therefore thrown away during the process of labour. However, collecting antenatal data in the first stage labour room, before the delivery had taken place, turned out to be too time consuming. In addition, some of the women who were registered in the first stage labour room did not deliver their baby during our survey period. Therefore, it was decided to try to collect most antenatal charts and case records in the second stage labour room, soon after the delivery. Records that were missed in this way, e.g. of women who gave birth during the night, were found on the maternity wards.

The majority of the information on Pro Forma part A was obtained from hospital records. However, some of these data had been provided to the hospital directly by the women themselves, including age, income, previous obstetric history and LMP. This has some consequences for the reliability of the information. For example, age was not always stated in completed years, but was probably in some cases rounded up. The data on monthly income are clearly unreliable. The government hospital applies income limits to assess the prices that patients need to pay for the care provided. Below an income level of Rs 1600.-- per month the care is free of charge. In addition, bureaucratic and time-consuming regulations may have prevented patients from stating their actual monthly income. The hospital generally does not check to see whether the stated income is correct.

Information in the files on previous obstetric history and LMP has also been provided by the women themselves. The majority of women knew the date of their last menstrual period and were able to produce this date without needing to give it much thought. Local informants and women in the waiting room of the outpatient clinic gave several reasons for
EARLY LIFE CHANGES

this. The following seemed to be most important. First of all, fertility and having children are very important for Indian women, and in Indian society in general. After marriage, women and/or their family members closely follow the menstrual cycle to detect possible pregnancies. In addition, based on traditions, Hindu women have certain customs and restrictions related to their menstrual period, e.g. with regard to visiting temples, doing poojas, and having sexual intercourse. This makes them aware of their menstrual periods. Further, health education programmes in India teach women that it is important to know their LMP and to watch the menstrual cycle.

Pro Forma part A also contained a limited section on postnatal information. If the baby was transferred to the IBN/SN or had died before being admitted to the nursery, this was registered on the form. In addition, we recorded discharge date of the mother (and newborn), if possible. These dates came from the discharge registers and paying books in the maternity and pay wards. Unfortunately, the discharge register maintained by the security office could not be used for this purpose since it would have been too time-consuming to find out which patients in this register were part of our survey. Use of the other discharge registers and paying books also posed some difficulties. The design of these registers complicated the search for women in our survey, and the registers contained quite a few typos and inaccuracies, especially with regard to IP number. To complicate matters even further, some mistakes were made in the allocation of the IP numbers during the second half of the survey period and, therefore, several IP numbers were not unique. It is also believed that some of the women checked out without being properly discharged. Overall, discharge date could not be determined for 13% of the women in the survey.

Pro Forma part A: direct data

Some of the data on Pro Forma part A were not obtained from hospital records but were directly obtained from the women themselves while they were still staying in one of the maternity or pay wards. The women were located and identified in the hospital with help of their name, their husband’s name, unit number, and IP number. Unfortunately, however, we were not able to find all the women before their discharge. In total, about 86% of mothers that had given birth to singletons and all the mothers of multiplets were found in the hospital wards.

The information that was asked of the women included level of education, paid employment, religion, and place of residence. Although some of the questions were not included on the pre-printed Pro Forma part A, they were asked routinely during our survey. Data on these topics were not included in the hospital records and could only be obtained in this way, by directly asking the women themselves. Postnatal information was asked on postpartum sterilisation (PPS) and immunisation of the baby. The questions were preferably posed to the women themselves, but in those cases where this was not possible family members answered them. It should be noted that this is likely to have consequences for the reliability of the data. In addition, some women may have had the desire to provide ‘socially desirable’ answers, for example with regard to immunisation of the baby and educational level. The information on PPS was later checked with the PPS registers of the hospital.
In addition to the questions mentioned above, we measured the height of the woman standing up, or sometimes lying down since for some women it was still too painful, tiring or difficult to stand up so soon after the delivery. The measured length should therefore be regarded as only an approximation of the woman’s height.

Pro Forma part B: unhealthy newborns

The admission register in the IBN/SN was frequently checked to see whether any of the neonates in our survey were admitted to one of these nurseries. Date and time of birth, sex, birth weight, names of the mother and father, and level of income helped to identify these babies and to link them to their mother.

For babies that were admitted to the IBN/SN, the author completed question form Pro Forma part B (Appendix F). This Pro Forma contained sections on pregnancy and birth outcome (e.g. date of birth, sex, birth weight, gestational age, and Apgar score), indication for admission, additional health problems, congenital anomalies, date of discharge or death, and cause of death. The information was based on the various registers (admission register, discharge register, death register), the case record of the baby, and the possible death certificate. However, unfortunately, it was not possible to obtain and view all case records and death certificates.

The registers, especially the admission register, were found to contain quite a lot of typos and inaccuracies, for example with regard to date of birth and birth weight. At least one neonate who had been admitted to the nursery was not listed in the admission register. In some cases, the information in the IBN/SN registers and case records differed from the information in Pro Forma part A, the OR register, and/or the mother’s case record. The differences most frequently concerned time of birth, birth weight, and Apgar score.

The postnatal cards

When we visited the women on the wards, we gave them a short questionnaire – the so-called postnatal card (Appendix G). In addition, they received a preprinted and prestamped envelope. The questions on the card were written in the local language Malayalam. We asked them to fill out the card and send it back in the envelope once their baby reached the age of 28 days. The questions on the card pertained to possible health problems, hospitalisation, and death during the neonatal period. In those cases where we were unable to give the card in person, we sent the card by mail. Although initially the cards were handed out or sent without an accompanying note, later a small letter was added to make the distinction between SAT Hospital and the researchers clearer (see also Section 4.3.5).

If we did not receive the postnatal card shortly after the end of the neonatal period, a second postnatal card with accompanying note and envelope was sent to the mother as a reminder. Unfortunately, the Indian postal services were on strike for about two weeks during the period in which most of the cards were supposed to be sent back. However, 596 postnatal cards were returned before January 11, 2001. Cards received later were excluded from further analysis.

Regrettably, some questions on the postnatal card (although written in Malayalam) turned out to be unclear to some women, especially the questions on hospitalisation during the
EARLY LIFE CHANGES

neonatal period after discharge from SAT Hospital (questions 4 through 7). In the answers to these questions, various women included visits to doctors during surgery hours without actual hospitalisation. Also, the date of discharge from SAT Hospital after the delivery was frequently entered in question 7. Naturally, the answers on the postnatal cards reflect the mother’s impression and her diagnosis of the baby’s health status. This means that we have to interpret the results with some caution. For instance, the relatively large group of neonates that had ‘measles’ (17 cases) more likely indicates a high incidence of rashes rather than an epidemic of measles. In addition, we suspect that some women have included health problems in their answers that occurred only after the neonatal period. They answered that their baby had not experienced any disease during the first 28 days of life but then entered health problems in question 3, such as fever, cough, and chest infection. There was even one case in which the baby was known to have died during its stay in SAT Hospital but the postnatal card did not mention this death. Conversely, two mothers had filled out ages at death and causes of death on the card in question 8, but on inquiry it turned out that their babies were still alive.

Postnatal check-ups and IBN/SN check-ups

Besides the postnatal cards, additional data on neonatal survival came from the postnatal check-ups and the check-ups for babies from the IBN/SN. Information was copied from the registers (the Postnatal Register and the Baby Weight Register) in the Family Welfare Department for those women and children in the survey who came for a postnatal check-up between November 17 and December 15, 2000. These data were later linked to the data that had been collected earlier. Linkage to the correct pro forma number took place on the basis of a combination of the following variables, when available: IP number, OR number, name of mother, date of birth, sex of the baby, and birth weight. However, it was not possible to find all the cases; a few cases could not be linked back. In total, 18% of the women in the survey came to the outpatient clinic for a postnatal check-up during the above mentioned time period and could be linked back to previously collected data. It is likely that many of the other women went to clinics and/or doctors closer to their homes. For 68% of the women who did come back reliable information about the survival of their baby was available. In the other cases, the registers did not contain any specific information on the child.

The information in the register of the outpatient clinic for babies discharged from the IBN/SN was very limited. Nurses were therefore requested to additionally register the IP numbers that the newborns had during their stay in IBN/SN. However, in quite a few cases this number was unknown and therefore it was impossible to link these babies to the previously collected data.

The resultant data set

The overall survey consisted of 1,022 consecutive births between Monday October 9, 2000 and Monday October 30, 2000. These births included 1,004 singletons and 9 pairs of twins. However, only singleton births were included in the analysis in the present study. Additional inclusion criteria were: a birth weight ≥ 1,000 g and a gestational age ≥ 28 completed weeks. These additional criteria follow guidelines from the WHO (1993) for the international comparison of perinatal statistics (see Chapter 3). Moreover, these are the same criteria that
SAT Hospital applies to distinguish spontaneous abortions from stillbirths. Two singleton cases in the data set weighed less than 1,000 g at birth and there was one case in which birth weight was unknown but the gestational age was less than 28 weeks. All three cases were antepartum stillbirths and, since they did not correspond with the criteria, they were excluded from further analysis. As a result, the basic SAT Hospital (SATH) data set that was used for the analyses in Chapters 8 and 9 consisted of 1,001 singleton births.

The majority of the 1,001 births were male: 523 males compared to 478 female babies, a sex ratio of 109.4 males to 100 females. This sex ratio is somewhat higher than the generally accepted sex ratio of viable human foetuses of 106 males to 100 females (Cunningham et al. 1993). The majority of cases had received antenatal care. Overall, 987 cases (98.6%) were booked in either SAT Hospital or another hospital, while only 12 cases (1.2%) were non-booked. In two cases antenatal care and referral status were unknown. Of the booked cases, 787 (79.7%) were booked in SAT Hospital while 200 (20.3%) had antenatal visits in another hospital and had been referred to SAT Hospital.

The hospital survey was designed to collect follow-up data on liveborn children until the end of the neonatal period. The SATH survey contained in total 988 live births but, unfortunately, quite a few cases were lost to follow-up (see below). Within the hospital, 116 neonates (11.7%) out of the 988 singleton live births were admitted to the In-Born Nursery (IBN) or Sick Nursery (SN), although not all immediately after birth. Eleven of these neonates (9.5%) died during their first stay at one of the two nurseries. The other 105 neonates were discharged but three of them (2.9%) had to be readmitted to the IBN/SN during their stay in SAT Hospital. Of these three, one neonate died before discharge. Therefore, to summarise, 12 out of 116 or 10.3% of neonates that were admitted to the IBN/SN died during their stay in hospital. In addition, one baby died on the 4th day of life before being admitted to the IBN/SN.

In total, 975 (988 minus 13) neonates left SAT Hospital alive. In 583 of these cases (59.8%), the mothers completed our postnatal card and we received them before January 10, 2001. None of these babies appeared to have died during the neonatal period. 177 (18.2%) of the 975 mothers that had left SAT Hospital with a live neonate came back for a postnatal check-up between November 17 and December 15, 2000 and could be identified. The baby was still alive in 120 cases but in one case the child had died during the neonatal period. In the other 56 cases, no information could be obtained on baby’s survival status. Further, 24 babies in the survey came to the IBN/SN for check-ups after discharge and could be identified.

Overall, 359 neonates (36.3% of live births) are unknown to have survived their neonatal period or not, i.e. they were censored or lost to follow-up. This group includes 308 children (31.2% of live births) who were even lost to follow-up before the end of their early neonatal period. For all of these babies, no information about neonatal survival status could be obtained through the discharge registers, the postnatal card, the postnatal check-up, and/or the IBN/SN check-up. In the life table analysis of the SATH survey in Chapter 8 (Section 8.2.2), censored cases are included up to the latest age (in days) that they were known to be alive. Some were already lost to observation during their stay in hospital (i.e. no discharge date
known or any other information) while others were lost after discharge (e.g. no postnatal card returned, did not go to postnatal check-ups). Because high-risk cases generally stay longer in hospital and are more likely to return for check-ups, most of the censored cases probably have survived the neonatal period. Hence, Chapter 8 analyses the SATH data in two different manners: (1) including censoring, and (2) under the assumption that all censored cases survived the neonatal period. An overview of the total sample or data set and the neonatal follow-up data is provided in Table 4.1.

### 4.3.5 Hospital Surveys as Data Sources

Hospital records and hospital surveys may provide an additional data source for demographers. However, as with any data source, the use of hospital data has its pros as well as its cons. Below, some difficulties and challenges encountered in relation to the survey in SAT Hospital are reflected upon.

A recognised and important drawback of using any hospital data is *selection bias*. Not all pregnant women seek the assistance of medical services in hospital. Women who are more likely to visit hospitals include women who suffer from symptoms and complications, women with high-risk pregnancies, women with previous obstetric problems, better educated women, higher socioeconomic classes, and women able to reach the hospital on time. Furthermore, the
selection and the characteristics of the women in a survey will depend on the type of hospital. Is the hospital a large, academic, referral hospital in a city, or a small rural clinic? Also, patients of public hospitals differ, at least with regard to certain characteristics, from patients of hospitals in the private sector. SAT Hospital in Thiruvananthapuram is a large hospital in the public sector that attracts people from a wide geographic area, including many complicated cases. A large group of patients come directly without referral. In Chapter 8, results from the survey in SAT Hospital are compared to results from other studies in Kerala to assess the possibility of selection bias.

Besides the risk of selection bias, the use of hospital records makes a researcher dependent on the availability of data in the hospital records. However, additional information can be obtained by conducting an extended survey that includes supplementary questionnaires and interviews with patients. In the present study, additional data on education, paid employment, religion, place of residence, maternal height, and neonatal health and survival were obtained directly from the mothers themselves.

Also when using hospital records, the reliability of the data is dependent on the quality of the hospital records. This means that in order to assess reliability of the data and possible problems in the data set, the researcher needs to know where the data have come from and how they were obtained or measured. For instance, it is important to know whether data are based on medical tests and previous hospital records or whether the information comes directly from the woman herself. In the present SATH survey, the data on income were provided by the women and were not checked by the hospital even though for some patients it may have been favourable to state a lower income than their actual income.

Related to the valid measurement of the variables are the diagnostic criteria and definitions that are applied by the hospital. Information about these is of great value during further interpretation and analysis of collected data. For example, it is important to know that SAT Hospital defines ‘SGA’ in term newborns as a birth weight of less than 2,500 g rather than basing it on a reference growth curve. Within the present survey, some diagnoses and diagnostic criteria (e.g. with regard to anaemia and pre-eclampsia) could be adapted retrospectively on the basis of recorded test results and measurements.

In addition to the above, all hospital staff members have to be familiar with the definitions and policies of the hospital, and apply them accordingly. Moreover, data in general should be measured and written down in a consistent manner, for all women and newborns in the same way (i.e. reliability). In a large, busy, and overcrowded hospital such as SAT Hospital with a sizeable staff and a large number of students in training, this may not always be the case. During the present survey, there was a mix-up in the assignment and registration of the supposedly unique IP numbers, which are at the basis of the hospital administration and which were used to identify patients in the various registers and records. Furthermore, measurements were not always standardised and assessed in the same manner or at the same time in pregnancy. For instance, different weighing scales (electronic and mechanical) were used to weigh the newborns. Another example is the timing of the first measurement of haemoglobin level in the case records, which varied from gestational week 7 to gestational week 41.
EARLY LIFE CHANGES

Of course, the prime concern of doctors, medical students, and hospital staff is providing proper care to individual patients. They have different priorities and perspectives than demographers, statisticians, and epidemiologists. Still, good record keeping is essential for providing proper care. Nevertheless, local informants think that record keeping in Kerala is sometimes poor. For example, many infections in pregnancy are believed to go unregistered, and are not mentioned in the case records. It is unknown whether this is due to poor record keeping or whether women fail to report these infections. In general, when a case record does not contain information about a specific health problem, there are several possible reasons. These include: the patient did not really experience the problem, the patient did not report the problem, the patient was not tested for the problem, the test results got lost, or the doctor failed to write down the diagnosis and test results.

Dissimilarities between disciplines also reveal themselves in the interpretation of medical files and medical jargon. In order to convert hospital records into a data set that can be used for further analysis, a researcher has to be able to interpret medical test results, to read medical files, and to understand the applied medical jargon and abbreviations. However, a researcher may decide to work with a medically trained person who has these skills. In this respect, the amount of work and effort that was put in by the postgraduate student from SAT Hospital during the present survey is highly appreciated.

The administrative practice or the organisation of administration and registration in a hospital are also likely to affect the process of data collection. In developing countries, collection of hospital data may be complicated by the absence of computerised systems for administration and registration. In SAT Hospital, registration is not computerised and consists of a large variety of records and registers (e.g. case records, admission registers, register of deliveries, death register, paying books). To reconstruct a patient’s hospital history, these registers and records had to be linked to each other. This was time-consuming, difficult and often impossible, partly because many of the registers were not designed for this purpose. For example, the discharge registers contained discharge dates, IP numbers, and names of patients in order of their date of discharge not in order of their IP number. In some cases, information vital to link files was absent. The register of the outpatient clinic of the IBN/SN initially only contained the first names of the mothers (and many mothers have the same name!), and the baby’s present weight and present head circumference while the IP number and date of birth were not included. Due to time constraints and the lack of links between hospital registers, we did not even try to find out whether any of the babies in the survey visited the general outpatient clinic or were readmitted after discharge from the hospital.

Finally, a few words on ethical concerns. Behind the figures and statistics are living people who experience what may be the most important and emotional events in their lives. They have emotions and feelings that should be respected. What is an increase in the number of stillbirths or simply another admission to the IBN/SN for researchers means grief, pain, fear or uncertainty for the parents of the child. Naturally, in the present survey, we considered the feelings of such parents as far as possible, but I regret that we have been unable to avoid some things from slipping through. For instance, the parents of a neonate who died in IBN were sent a postnatal card by mail. In addition, for some parents in the present survey, it was difficult to make a distinction between the researchers and the hospital. For example, at the
beginning of the data collection period the sex of one of the babies was incorrectly written on its postnatal card by mistake. The card was sent to the parents by mail and the father of the child (a girl) became worried and accused the hospital of switching his baby. Ultimately, the head of the maternity unit involved was able to resolve the matter. After this incident, the postnatal cards were accompanied by an additional note from the survey team, which was meant to make the distinction between researchers and the hospital clearer. Nevertheless, it appears that many women answered our questions and filled out our postnatal cards because they felt obliged to the hospital. This could explain the rather high response to the postnatal cards. Indeed, some women even wrote of their gratitude on the postnatal cards while the researchers are the ones who owe them their thankfulness for co-operating and completing the card.

4.3.6 Analysis

Data preparation
The inclusion criteria for the analysis of the data set (i.e. birth weight \( \geq 1,000 \) g and gestational age at birth \( \geq 28 \) weeks) were applied only after gestational age at birth and birth weight were checked and, if necessary, corrected. The original data file in SPSS contained dates in the format ‘day-month-year’ that had to be combined and recalculated into duration times. The duration between two dates (i.e. date I and date II) was calculated in the following manner. First, both dates were split into three different variables indicating day, month, and year respectively (based on Mamun 2001). This was done with help of the SUBSTR function in SPSS. Then, for each of the three variables, date I was subtracted from date II. The thus created variables – let us call them ‘DAY’, ‘MONTH’ and ‘YEAR’ – were subsequently combined and transformed into weeks or days. For the calculation of gestational age in weeks at birth, the following formula was used:

\[
\text{GESTATIONAL AGE} = \left[ \text{DAY} + (\text{MONTH} \times 30.5) + (\text{YEAR} \times 366) \right] / 7 \tag{4.25}
\]

where 30.5 is the average number of days per month between November 1999 through October 2000 and 366 is the total number of days in the leap year 2000 (based on Mamun 2001). Likewise, duration of neonatal survival in the follow-up period after birth was calculated in days by applying the following formula:

\[
\text{NEONATAL SURVIVAL} = \text{DAY} + (\text{MONTH} \times 30.67) + (\text{YEAR} \times 365) \tag{4.26}
\]

where 30.67 is the average number of days per month between October 2000 through December 2000 and 365 is the number of days in a non-leap year.

Gestational age at birth (in weeks) was calculated in two ways: on the basis of LMP and on the first ultrasound scan (USS). Differences between the two estimates of three weeks or more, as well as outliers, were checked and compared manually. For some preterm babies, gestational age on the basis of a paediatric scoring system was also known. In general, preference was given in the following order: (1) gestational age according to the scoring system, (2) age on the basis of LMP, and (3) age on the basis of age estimation in the first USS, depending on their availability. Paediatric scoring systems seem most reliable as they
EARLY LIFE CHANGES

assess gestational age on the basis of visible physical maturity. However, scoring systems may still have a predictive error of about two weeks. In the case of LMP, measurement is based on memory. Furthermore, it cannot be stated with certainty that a woman who misses a menstrual period is indeed pregnant and that a woman who appears to menstruate is not pregnant. Estimation on the basis of LMP can thus result in an error of four weeks or even more. Lastly, estimation on the basis of USS is sensitive to distortion by growth retardation, especially towards the end of the pregnancy. Moreover, it has often been suggested that Indian foetuses and newborns are constitutionally smaller than their Western counterparts, which is also likely to affect estimation of gestational age. In the process of checking and correcting gestational age at birth in the SATH data, possible distortions due to known irregular menstrual cycles (affecting age estimation based on LMP) and/or by known IUGR (affecting age estimation in USS) were taken into consideration. When no reliable age estimate could be made, gestational age at birth was considered to be unknown or ‘missing’. With regard to neonatal follow-up, equation (4.26) was applied to calculate age in days at neonatal death. In addition, if known, age in hours at neonatal death was calculated manually by subtracting date and time of death from date and time of birth. The two results were compared and if necessary, age at death in days was corrected.

Birth weight data from Pro Forma A were compared to Pro Forma B and the postnatal cards. In the event of any differences between the recorded weights, the most likely weight was chosen with a strong preference for Pro Forma A. When serious doubts existed, weight at birth was considered to be unknown or missing. Hospital definitions of small-for-gestational-age (SGA) were not adopted in the analysis. Instead several standard growth curves from the literature were applied to the data set (see Chapter 8) to see which newborns were to be classified SGA.

The presence or absence of congenital anomalies was assessed on the basis of ultrasound scans during pregnancy and/or observation at birth and during the stay in hospital after birth. The information was based on the hospital records and, again, the data from Pro Forma A (based on antenatal charts and labour records) were compared with the data on Pro Forma B (based on IBN/SN patient files). Suspected cases of congenital anomalies, including neonates who were admitted to the IBN/SN to rule out the presence of any anomalies, were also included in the analysis. This was because, in many cases, the options for diagnostic tests were limited. In addition, it is possible that not all congenital anomalies present among the survey population were detected before discharge. Moreover, results of autopsies have not been included since they would require a long waiting time.

The presence or absence of birth asphyxia (BA) was established on the basis of the hospital records (labour records and IBN/SN patient files). Unfortunately, the definitions and criteria that were applied by SAT Hospital to diagnose BA have not been established. Two cases that were admitted to the IBN/SN partly ‘to rule out’ birth asphyxia were included in the BA group as suspected cases, even though their Apgar scores at one minute were 9. Since all cases of BA were live births (see Chapter 8), the reliability and completeness of the hospital registration system (at least with regard to birth asphyxia) can be called into question. In addition to the above, the present study also analysed ‘birth asphyxia’ on the basis of Apgar score. First, Apgar scores from Pro Forma A were compared to the scores on Pro Forma B.
When differences were observed, preference was given to the Apgar scores on Pro Forma B, since these were based on the IBN/SN patient files and were often more complete. Low Apgar scores (≤ 3 at one minute and ≤ 6 at five minutes) were labelled as BA cases. It should be noted that preterm cases with low Apgar scores were included in these BA groups.

Methods of analysis
On the basis of the SATH data, mortality rates and life tables were calculated and the results are presented in Chapter 8. Additional analysis of the SATH data was performed using SPSS and these results are presented in Chapters 8 and 9. The first chapter, Chapter 8, deals with the associations between the selected, adverse pregnancy/birth outcomes on one hand and stillbirth and neonatal death on the other. Chapter 9 discusses additional risk factors and their associations with adverse pregnancy/birth outcome, stillbirth, and neonatal death. All the factors of interest were transformed into binary variables.

The analysis of the data includes the estimation of relative risks. This was performed on the basis of crosstabulations and by using SPSS to calculate the RR and its accompanying 95% confidence interval. If the 95% confidence interval (95% CI) included the value of 1.0, then the value of the RR was regarded as not significantly different from 1.0. In other words, the probability that there is no increased or decreased risk (i.e. incidence in the exposed group equals incidence in the non-exposed group) was considered as too high. It is important to note that the confidence interval is in part affected by sample size. In addition to the RR, Chapters 8 and 9 also present estimates of AR(E) and EF that are based on equations (4.8) and (4.10).

The data were also analysed for associations on the basis of Pearson’s Chi-square and its p-value. In Pearson’s Chi-square test, observed data are compared to a situation in which there is no association between the two variables. The null hypothesis is that there is no association between the two variables in question. The p-value indicates the probability that the model of the null hypothesis (i.e. no association or interaction) predicts the data in the present population. In the present analysis, the null hypothesis was rejected only when this p-value, or probability, was less than 5%. In addition, Cramer’s V was calculated to gain an impression of the strength of the observed associations.

Apart from the results from the present SATH survey, Chapters 8 and 9 also present other data from Kerala on mortality figures and prevalence of pregnancy/birth outcomes. These data were obtained from other, secondary, sources and include both community-based data and hospital data. Major additional sources of data were the Sample Registration System (SRS) and both National Family Health Surveys (NFHS-1 and NFHS-2). For a brief discussion of SRS and NFHS, please refer to Chapter 2. In addition, various articles and books provided information on the situation in Kerala. For example, Padmadas (2000) analysed data from NFHS-1 on child survival. The results from the SATH survey were compared to the other data from Kerala to look for possible effects of selection bias.

Subsequently, the data from SAT Hospital and Kerala were compared with the data for the developed world (see Chapters 5, 6, 7, and 9) and secondary data from India (see Chapters 2, 8, and 9). This made it possible to place the risk factors under study within the framework of an epidemiologic transition and to study the changes in this transition over time.
4.4 Summary and conclusions

The present chapter has discussed the data and methods that were applied in the current study to answer the research questions. The study tries to expand and complement demographic data sources using data from medical and epidemiological sources. Moreover, it adopts several measures for frequency and risk that are common in epidemiology. These include incidence, which refers to change and the number of new cases occurring in a population, and prevalence, which focuses on status and existing cases. Although the distinction between the two concepts seems rather straightforward, it can cause confusion in the period around birth.

Essentially, the study can be divided into two parts, each based on a different approach. The first part focuses on the situation in the developed world (i.e. regions in the later stages of the epidemiologic transition), which is operationalised as the EME region or the established market economies (see also Chapter 2). One of the main objectives is to assess the relative importance of the selected risk factors (see Chapter 3), both at the individual level and at the population level. The relative importance at the level of the individual is established in terms of risk (cf. incidence) of loss or death for persons affected by adverse pregnancy/birth outcome, and in the relative risk of loss or death. The relative risk (RR) indicates how many times more likely persons with the risk factor will experience a particular outcome compared to persons without the risk factor. The relative importance at the level of the population is established in terms of frequency (cf. prevalence) of adverse pregnancy/birth outcome within the total population, and in the proportion of losses or deaths in a population that are due to, or can be attributed to, the risk factor. In this, the attributable risk among the exposed, AR(E), indicates the contribution of the risk factor to losses or deaths in the exposed population, whereas the etiologic fraction (EF) refers to the total population, whether exposed or not to the risk factor. The distinction between an individual level and a population level implies two types of interventions in the underlying processes: (1) improvement of the survival chances of individuals who are affected by the risk factor, and (2) lowering the prevalence of the risk factor in the population.

The situation in the EME region is described and analysed on the basis of secondary data. Most data were obtained from literature in fields such as obstetrics and gynaecology, neonatology, paediatrics, genetics, cardiology, epidemiology, and public health. However, the desired information was not readily available. Data were limited and fragmented and most publications provided none or only part of the data required to make estimations. In addition, results from studies were difficult to compare due to differences among studies (e.g. study design, definitions applied) and in the presentation of data. The present chapter has discussed the alternative methods that were applied to reach estimates of the various measures of interest. These methods combine assumptions derived from secondary data sources with mortality figures obtained using a hypothetical cohort. For a discussion on the construction of the hypothetical cohort and the accompanying life table analysis, please refer to Chapter 5.

The second part of the study focuses on the situation in regions that are approaching the later stages of the epidemiologic transition. For most less developed countries, data on health and survival are sparse, if available at all. During recent decades, the main sources of information on child survival in developing countries have been large-scale demographic
surveys, such as the Demographic Health Survey (DHS) and the National Family Health Survey (NFHS) in India. However, these surveys have been found to provide insufficient information to explain child survival levels and the mechanisms underlying child health and survival. The present study tries to deal with the shortfall in data by implementing a hospital survey in a case region. The hospital survey was conducted in Sri Avittom Thirunal (SAT) Hospital in Thiruvananthapuram, Kerala. The choice for Kerala as a case region and as an illustration of a region in transition, and the choice of the hospital, a large ‘referral’ hospital in the public sector, were based on the data presented in Chapter 2 and on a feasibility study in South India (i.e. Karnataka and Kerala). The results of the hospital survey were compared to, and complemented by, data from the Sample Registration System (SRS), data from NFHS-1 and NFHS-2, and data obtained from medical and epidemiological literature.

The present chapter has attempted to provide a comprehensive and detailed description of the organisation and registration procedures in SAT Hospital concerning antenatal care, care during delivery, and neonatal care. Subsequently, the chapter discussed the design of the survey, its data sources, the collection of the data, the preparation of the data for analysis, and the methods of analysis. Information was collected on general characteristics and the health status of the woman, pregnancy, delivery, pregnancy and birth outcome, and neonatal follow-up. A large proportion of the data was obtained from patient files, or case records, and a variety of hospital registers including those on follow-up check-ups. In addition, the mothers were asked to fill out a short questionnaire on neonatal health and survival. Overall, data were collected on 1,022 consecutive births between Monday October 9, 2000 and Monday 30, 2000. Multiplets (all twins), births < 1,000 g, and births < 28 completed gestational weeks were excluded from further analysis and, as a result, the eventual data set for analysis consisted of 1,001 births. Complete neonatal follow-up data were available for 629 out of the 988 live births while 359 (36.3% of live births) were lost to follow-up (i.e. censored) sometime during the neonatal period. Mortality rates and life tables were calculated and additional analysis of the SATH data was performed using SPSS.

Hospital records and hospital surveys may prove to be an additional data source for demographers. The current chapter reflected on some of the issues and difficulties that were encountered in relation to the survey in SAT Hospital. These include:

- selection bias,
- the availability, reliability, and validity of hospital records,
  e.g. the records held by SAT Hospital did not include information on education, paid employment, religion, and maternal height, and data on income were unreliable,
- differences between the disciplines of medicine and demography/statistics,
  e.g. in priorities, perspectives, and jargon,
- administrative practice in the hospital,
  e.g. the existence of a large range of registers and records, and problems in linking them,
- and ethical concerns.