in Section 7.2. However, the main purpose of this chapter is to show how the linear logistic model can be used to analyse data from aetiological studies.

7.1. BASIC DESIGNS FOR AETIOLOGICAL STUDIES

Two basic designs used in epidemiological surveys into the aetiology of a disease are the **cohort study** and the **case-control study**. Each of these designs is described in the following sub-sections. Fuller details on the design of aetiological studies can be found in Breslow and Day (1980, 1987).

7.1.1 The cohort study

In a cohort study to determine if there is an association between certain exposure factors and the occurrence of a particular disease, a sample consisting of individuals who are free of the disease in question is selected. The individuals in this sample are then stratified according to the exposure factors of interest, and followed-up for a given period of time. Each individual is then classified according to whether or not he or she has developed the disease that is being studied. The relationship between the probability of disease occurrence and the exposure factors can then be investigated.

Example 7.1 The Framingham study
Framingham is an industrial town located some twenty miles west of Boston, Massachusetts, USA. In 1948, a cohort study was begun with the broad aim of determining which of a number of potential risk factors are related to the occurrence of coronary heart disease (CHD). At the start of the study, a large proportion of the town's inhabitants were examined for the presence of CHD. Measurements were also made on a number of other variables, including age, serum cholesterol level, systolic blood pressure, smoking history and the result of an electrocardiogram. Those individuals found to be free of CHD at that time were followed-up for twelve years and those who developed CHD during that period were identified. The resulting data set consisted of this binary response variable and information on the risk factors for 2187 men and 2669 women aged between 30 and 62. The summary data in Table 7.1 are adapted from Truett, Cornfield and

Table 7.1 Proportions of cases of CHD, cross-classified by age and sex, and initial serum cholesterol level

Sex	Age group	Serum Cholesterol level			
		< 190	190–219	220–249	≥ 250
Male	30–49	13/340	18/408	40/421	57/362
	50–62	13/123	33/176	35/174	49/183
Female	30–49	6/542	5/552	10/412	18/357
	50–62	9/58	12/135	21/218	48/395

Kannel (1967) and relate to the initial serum cholesterol level (in units of mg/100 ml) of these individuals, cross-classified according to their age and sex. In analysing these data, the epidemiologist would be interested in the extent to which the occurrence of CHD is associated with initial serum cholesterol level, after age and sex effects have been allowed for, and whether or not the degree of association is similar for each sex and each age group.

A cohort study is often termed a **prospective study**, since the individuals are followed prospectively in time. However, a modification of the design that is useful in studies of occupational health would involve workers who have been in that occupation for a relatively long period of time, ten or twenty years, for example. The current health status of these employees would then be compared with historical records, and comparisons made between different occupational groups on the basis of the probability of disease occurrence. Investigations of this type, where the outcome of interest has occurred before the study has begun, are known as historical cohort studies.

If the follow-up time differed for each individual, if they had been at risk for different periods before the study began, or at risk intermittently through the duration of the study, it would be sensible to take account of this in the analysis. One possibility is to compute the number of **person-years** of exposure for individuals who have been cross-classified according to factors such as age group. The number of individuals who develop the disease in a particular group is then expressed as a proportion of the person-years of exposure. This rate of occurrence of the disease can then be modelled using the **Poisson regression model**, described briefly in Section 8.2. When the time from entry into the study until the occurrence of a particular disease is of interest, models developed for the analysis of survival data, particularly the **proportional hazards model** due to Cox (1972), become appropriate. These techniques for analysing data from occupational health studies can similarly be used in the analysis of occupational mortality data. The methods will not be considered in this book, but are described by Breslow and Day (1987), for example.

7.1.2 The case-control study

A case-control study is an aetiological study in which comparisons are made between individuals who have a particular disease or condition, known as **cases**, and individuals who do not have the disease, known as **controls**. A sample of cases is selected from a population of individuals who have the disease being studied and a sample of controls is selected from individuals that do not have the disease. Information about factors which might be associated with the disease is then obtained retrospectively for each person in the study. For this reason, a case-control study is also known as a **retrospective study**.

Example 7.2 Age at first coitus and cervical cancer
Adelusi (1977) describes a case-control study to investigate whether coital characteristics are associated with the subsequent development of cervical cancer. In this