CHAPTER 2

METHOLODOGY

This chapter includes a description of the methods used in the study. These methods include the following components.

(a) Study Design

(b) Method of Data Selection

(c) Data Management

(d) Method of Data Analysis

1. Study Design

A cross-sectional study design is used, based on (a) cases of HIV/AIDS obtained from official surveillance reports, and (b) estimated populations in 5-years age groups for males and females in each province in each year.

The source population: the entire population of Southern Thailand

The study population: data obtained from HIV/AIDS official surveillance reports, consisting of the numbers of HIV /AIDS cases in the five provinces including Songkhla, Satun, Pattani, Yala and Narathiwat, Thailand, together with the corresponding populations classified by age, gender and year.

2. Method of Data Selection

No selection method was required as the entire population was available and able to be utilized for all analyses. The 11 variables selected were considered to be important risk factors for HIV/AIDS. For example, sexual activity and variable have been identified as determinants for the increased risk of HIV/AIDS in Southern Thailand (Ministry of Public Health, 1997: 163). Furthermore, Southern Thailand has a particular set of issues to consider when studying HIV-infection especially.

The main reason for data being analyses from the specified provinces was that four of these provinces share a physical border with Malaysia, and the fifth (Pattani) is part of the same area, and these areas are currently experiencing a rapid spread of HIV/AIDS.

Furthermore, the South receives workers not only from other parts of Thailand, but also from neighboring Myanmar, Kampucha and Laos. These workers are in pursuit of improving their economic means. In addition, separation from their family places migratory workers in a vulnerable position in relation to risk taking behaviour. Therefore, migratory workers including fishermen in Pattani, Songkhla and Satul provinces are a special risk group. Furthermore commercial sex workers have been found in each province.

Finally, in Hat Yai, Betong and Sungaigolok, there are many opportunities for men to indulge themselves in sexual night life. Tourists flock from Malaysia and Singapore to this booming city to enjoy night clubs, discos and bars. Clearly the sex trade influences the extent of HIV/AIDS transmission in Southern Thailand.

2.1 Case data

The data collection period was between 1 January 1991 until 31 December 1996. Reports were collected from government organizations including general hospitals, rural hospitals, and health offices as well as non government organizations in each province. These reports were confirmed by the records of the Division of Epidemiology, Ministry of Public Health. Data collected provided information on persons testing, and number seropositive (HIV+). The original data set consisted of diagnosis, sex, age, marital status, race, occupational, province, date of diagnosis, status, behaviour, and method of transmission.

The objectives of this study are to identify and describe using statistical methods, trends in the incidence of HIV/AIDS. However, in this study the variables selected to be used for analysis comprise only disease status, sex, age, date of diagnosis and province. The reason for focusing on these variables is to determine the main factors related to the incidence of the disease.

2.2 Population data

The population data information is obtained from the Human Resources Planning Division National Economic and Social Development Board, and comprises the count for each year, gender, and age group.

2.3 Inclusion criteria

Male cases include those who had the result for the HIV antibody test and the ELISA Western blot test between 1 January 1991 and 31 December 1996 in the five provinces under study. Female cases included those who had similar results between 1 January 1993 and 31 December 1996

All cases whose transmission method was sexual intercourse or IVDU were included.

2.4 Exclusion criteria

All male cases with the onset of HIV/AIDS infection before 31 December 1991, and female cases occurring before 31 December 1993 were excluded, as there were very few cases before these times.

All cases whose transmission method was by blood donor or mother to child, were also excluded because these case numbers were very small.

3. Data Management

Study factors :	Demographic (age and gender)
	Geographic (province)
	Method of transmission (sexual intercourse or IVDU)
Outcome factors:	HIV infection
	AIDS

The population data were structured as grouped data records, one for each combination of province, year, gender and age group. The case data were structured as case-by-case records. These files were jointed using a computer program written in the SPIDA language (Gebski et al, 1992). This program is listed in the appendix.

4. Methods of Data Analysis

4.1 Preliminary analysis

A) The incidence rates are computed and used to assess the pattern of the epidemic by plotting the incidence of HIV/AIDS cases stratified by age group, province, gender separately for sexual intercourse and IVDU to see the trend. The incidence rates is defined as follows:

$$\text{Incidence rate per 1000} = \frac{\textit{number of new cases during the year}}{\textit{estimated population on July 1}} * 1000$$

B) The crude odds ratios are computed an estimate of the risk or probability of the outcome. Assuming that the data in the contingency table have the counts *a, b, c,* and *d,* the estimated odds ratios is thus

$$\text{OR} = \frac{ad}{bc}$$

In this formula, a is the number of cases in the group of interest (e.g., males in Pattani in 1994, b is the number of cases in the referent group (taken to be Songkhla 1996), c is the number of noncases in the group of interrest and d is the number of noncases in the reference group.

The crude odds ratios for males and females in each province comparing the pattern of incidence in different years, and the crude odds ratios in each year comparing different province were plotted.

C) Odds Ratio of adjusted computed for remain confounder including age are estimated using Mantel-Haenszel (1959) Adjustment. In this study of we want to compare risks of disease in difference between the province under study with respect to age, gender. An estimates of a common effect are still robust, we need to check that we can justifiably combine the components from different strata. It is reasonable to do this if the stratification variable is simply a confounder, but not if the effect varies across strata.

Thus analyzing methods essentially involve subdividing the data into sets of similar components and then combining the component sets by a kind of averaging process. Confounding may be arise when two or more risk factors are mutually associated. For this reason it is important to give measures of the associations between the various risk factors of interest.

Correlation between two determinants is a necessary condition for confounding, but it is not sufficient. A covariate must also be an independent risk factor for the outcome before it can be a confounder.

Ignoring a confounder gives a single crude estimate of an odds ratio. To obtain a single estimate of an odds ratio adjusted for a confounding variable, the individual estimates obtained from each strat need to be combined in some way. It is important to check that individual estimates are compatible before combining them and testing the homogeneity of odds ratios. A measure of the overall association may be obtained by combining the statistics obtained from the component tables. Since odds ratios have skewed distributions, it is better to work with their logarithms. If the counts are very small, in which case the component odds ratios and standard errors could vary considerably, giving an inaccurate result. Mantel and Haenszel (1959) suggested odds ratios estimation using the following equation.

$$OR_{MH} = \frac{\sum a_g d_g / n_g}{\sum b_g c_g / n_g}$$

where g specifies the stratum. Accordingly this estimate is known as the *Mantel-Haenszel estimate* of the adjusted odds ratio.

4.2 Logistic regression

Logistic regression to estimate models required that the response variable be dichotomous. We focused on the odds ratios as the preferred measure of association. The outcome variable HIV/AIDS status for logistic regression is dichotomous that the logistic model takes the form

$$\ln\left(\frac{p}{1-p}\right) = a + b_1 x_1 + b_2 x_2$$

The main interest is whether the rate parameter varies with exposure. The data may be set up as counts, as in the case of disease status (HIV/AIDS). Concerning the effect of province, year and age group related to patients suffering from the HIV/AIDS virus, the outcome of interest was the number of cases of disease status.

The logistic regression obtained as a result of fitting model to these data is given, entitled *Odds* and *95% CI*. It is useful to examine the p-value for the exposure variable. This is based on using the odds ratio as a test statistic. The numbers labeled *Coeff* and *StErr* are the estimate of the parameter, and its standard error, respectively.

The numbers in the bottom line are as follows. The first *(df)* is the number of degrees of freedom remaining in the data table after fitting the logistic model, and is always zero for a two-by-two table, since there are two independent pieces of data to be fitted and there are two parameters in the model. The second number *(Dev)* is the *deviance,* a measure of the error after fitting the model. Again this must be zero for a two-by-two table since the model is a perfect fit and there is thus no error associated with it.

The logistic model may be used to fit data from any two-by-two contingency table, the data arise from a cross-sectional study. Of course if the data are simple, there is no benefit in fitting the model, since all the relevant statistics are given by simple formulas. The usefulness of the logistic model arises from its ability to handle covariates.