

## Chapter 2

### Methodology

This chapter describes the research methodology used for this research. Study design, study population, variables, data collection, data management, and statistical methods are described.

#### 2.1 Population and Sample

The target population comprised the liver cancer patients in Songkhla province from 1989 to 2007. The sample is patients with liver cancer 1,254 people in Songkhla province during 1989-2007.

#### 2.2 Study Design and Data Source

A retrospective cross-sectional data on liver cancer in Songkhla from 1989 to 2007 were obtained from Songkhla Cancer Registry. These data were collected from diagnostic questionnaire. The causes of illness are diagnosed by physician when illness occurs in Songklanagarind hospital or assessed by team of experts in each village when illness occurs outside hospital.

#### 2.3 Variables and Conceptual Framework

The interest variables of this study comprised six determinants and one outcome variable.

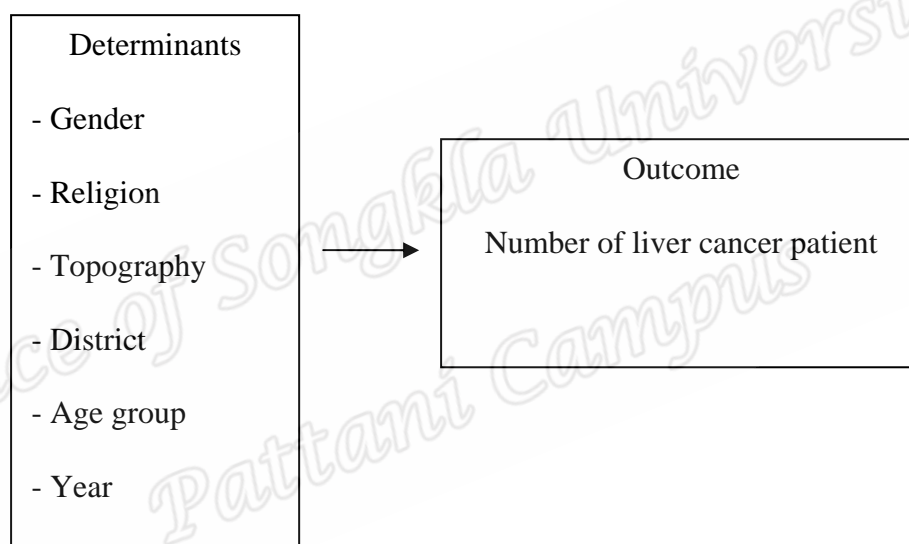
##### *Determinants*

The six determinants variables are gender, religion, topography (HHC and CCA), district, age group and years. The *district* variables comprised 11 districts, there are

Muang (M), Chana and Na Mom (C+Nm), Na Thawi (Nm), Thepha (Ta), Saba Yoi (Sy), Ranod, Krsea Sinthu, Sathing Phra and Khuan Niang (Ot), Rattaphum (Rt), Sadao and Khlong Hoi Khong (Sd+Kh), Hat Yai (Hy), Bang Klam (Bk), and Singha Nakhon (Sn). There are four *age-groups* comprising <30, 30-44, 45-59 and 60+.

### *Outcome*

The outcome variable was number of liver cancer patient in Songkhla province from 1989 to 2007. Figure 2.1 shows a schematic diagram for the variables of interest.



*Figure 2.1: Schematic diagram of variables of interest*

## **2.4 Data Management**

The following computer programs were used for data analysis and thesis preparation.

### *Microsoft Office*

*Excel* was mainly used to manage the data for this study. It used for data cleaning including data coding and fixing the data error.

### *R Program*

*R* was mainly used to analyze the preliminary results and to fit the Poisson regression model for the liver cancer incidence.

## **2.5 Statistical Methods**

Since all determinants variables are categorical the descriptive statistics are described by percentage.

### ***Incidence rate***

The incidence rate form liver cancer was computed using equation below

$$Y_{ghijkl} = K \frac{N_{ghijkl}}{P_{ghijkl}} \quad (2.1)$$

Where  $Y_{ghijkl}$  is incidence rate classified by gender  $g$ , religion  $h$ , topography  $i$ , district  $j$ , age group  $k$ , year  $l$ ,  $N_{ghijkl}$  is the number of liver cancer patient and  $P_{ghijkl}$  is population at risk used housing census in year 2000 and  $K$  is scaling constant such as 1,000, 10,000 or 100,000 population.

### ***Poisson Regression***

Poisson regression is appropriate for fitting model with count data. The number of liver cancer patient which is count data with non-negative integer. The probability of Poisson distribution with observed counts of  $y$  is given by:

$$\text{Prob}(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!} \quad (2.2)$$

We can calculate the corresponding probability that a count equal to  $y$  is observed in a Poisson distribution with mean  $\lambda$ . Since  $\lambda$  is the adjustable parameter in this

Poisson model for the variation in observed count, it is natural to link  $\lambda$  to the values of explanatory variable of interest. Because the mean  $\lambda$  of a Poisson distribution must be greater than zero it would be unsuitable simply to assume that:

$$\lambda = a + b_1x_1 + \dots + b_kx_k \quad (2.3)$$

The restriction of Poisson distribution is the mean must equal with variance. Poisson regression model can be fitted by using the generalized linear models (GLMs) equation with the log link function. Thus the poisson regression model is taken form:

$$\ln(\lambda_{ghijkl}) = \ln(p_{ghijkl}) + \mu + \chi_g + \omega_h + \alpha_i + \beta_j + \kappa_k + \eta_l \quad (2.4)$$

Where  $\lambda$ , is the mean of  $Y_{ghijkl}$ ,  $p_{ghijkl}$  is number of population in gender  $g$ , religion  $h$ , topography  $i$ , district  $j$ , Age group  $k$ , year  $l$ ,  $\chi$  is the effect of gender,  $\omega$  is the effect of religion,  $\alpha$  is the effect of topography,  $\beta$  is the effect of district,  $\kappa$  is the effect of age group and  $\eta$  is the effect of year. We will suppose that the effect of variables

$$\chi_1 = \omega_1 = \alpha_1 = \beta_1 = \kappa_1 = \eta_1 = 0.$$