CHAPTER 2

METHODOLOGY

This chapter describes the methods used in the study. The methodology comprises the following components.

1. Study Design
2. Variables
3. Data Collection and Data Management
4. Statistical and Graphical Methods

Study Design

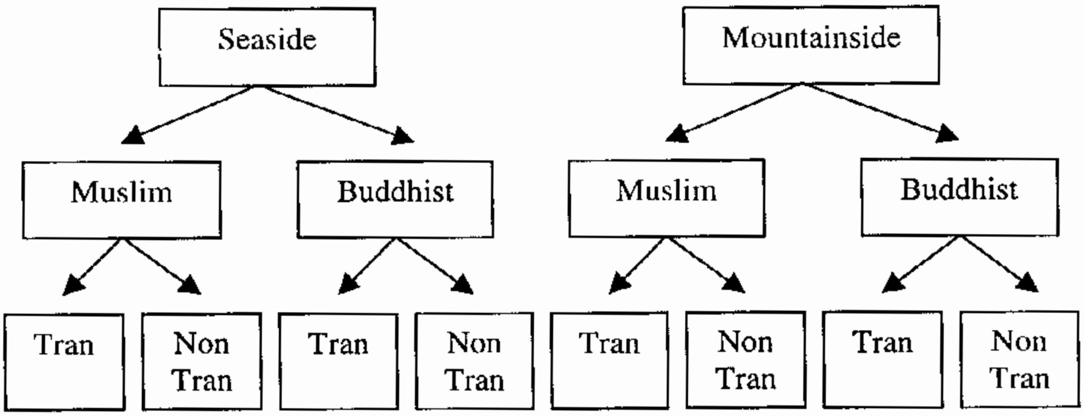
A stratified cross-sectional survey was used for this study because it is the appropriate method for collecting data in order to assess the risk of DHF outbreak at each location.

1. Sampling frame

Random sampling was used for selecting houses in the sample for this study, described as follows.

Pattani Province is divided into two parts, seaside (Nong Jig, Muang, Yaring, Panarehk, Sai Buri and Mai Kaen District) and mountainside (Kok Pho, Mae Lan, Yarang, Mayo, Tung Yangdaeng, and Kar Por Districts) areas. There were selected only districts of each area by simple random method. Religion and transmission/non-transmission areas stratified the target villages, and there was selected only village for each area by simple random method. The stratification is shown on Figure 2.1.

Figure 2.1 Schematic Diagram of stratify random method



2. Sample size

The sample size needed to obtain a specified precision can be estimated using the following equation (McNeil, 1996)

$$n = Z^2_{\alpha/2} \frac{P(1-P)}{d^2} \qquad (2.1)$$

Where $Z_{\alpha/2}$ is the critical value for standardized normal distribution corresponding to a two-tail probability α. The parameter P is the probability of an advise outcome. The value d is half of the width of the $100(1-\alpha)\%$ confidence interval.

An estimate of P is 0.5, the outcome being defined as a water container having mosquito larvae in a specified location in Pattani Province. α is 0.05 and d is 0.05. The number of containers to be sampled in of the study should thus be

$$n = 1.96^2 \times \frac{0.5(1-0.5)}{0.05^2} = 384$$

The containers were sampled from different households, this means that the sample size should be 384 households. But if the sampling unit is taken for convenience, to be a household, and we assume 7 containers per household, only 55 households are needed. But we should also allow for the correlation of outcomes from containers in the same household. This is not expected to exceed 3, so the number of households required is at most $3 \times 55 = 165$.

Table 2.1 shows the distribution of subjects selected using stratification by three covariates (location, religion, and DHF transmission)

Table 2.1 Distribution of subjects selected for this study.

Determinant	Panarehk		Kok Pho		Total
	Muslim	Buddhist	Muslim	Buddhist	
Non Transmission	20	20	20	20	80
Transmission	20	20	20	20	80

Variables

1. Outcome variable in the study

- Container having larvae of Dengue vector

2. Determinant variables in the study

-Water consumption characteristics in the villages: drinking water source, washing water source, renewal of drinking water container, and renewal of washing water container.

-Characteristics of container: container type, lid, material, place, and size.

3. Intervening variables in the study

-Other containers which are not used for storage, drinking and washing

4. Stratification variables

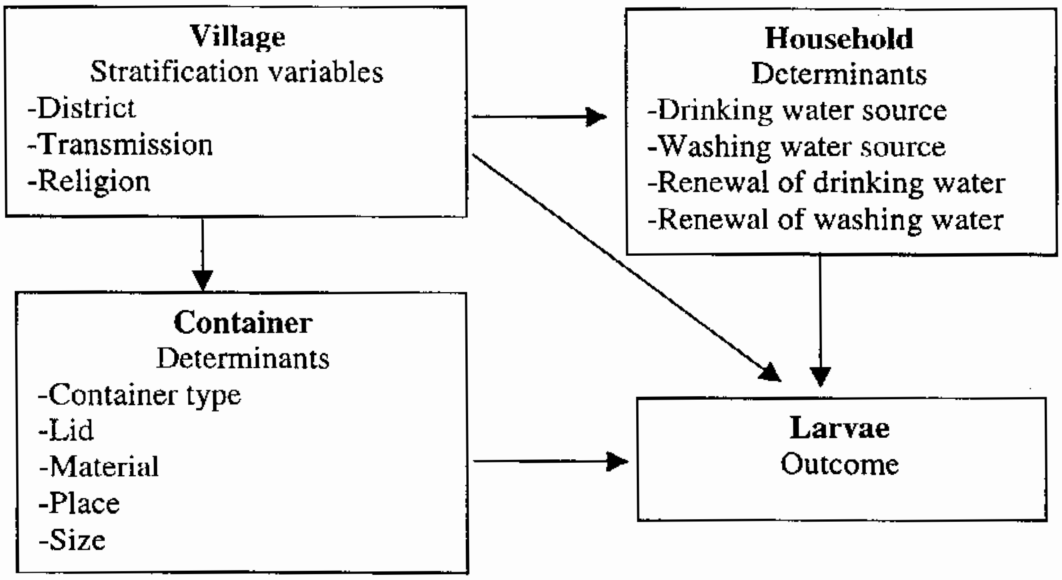
-Religion (Islamic/Buddhist), transmission or non-transmission village, location (seaside, Panarehk district; or mountainside, Kok Pho district)

The categories of each variable are as follows.

Districts	1 = Panarehk, 2 = Kok Pho
Religion	1 = Islamic, 2 = Buddhist
Transmission	0 = Non-transmission, 1 = Transmission
Drinking water source	1 = Well, 2 = Tap water, 3 = Rain, 4 = Carafe, 5 = Others
Washing water source	1 = Well, 2 = Tap water, 3 = Rain, 4 = Carafe, 5 = Others
Drinking water renewal	1 = Every day, 2 = 2-3 days, 3 = 4-6 days, 4 = Every week, 5 = Others
Washing water renewal	1 = Every day, 2 = 2-3 days, 3 = 4-6 days, 4 = Every week, 5 = Others
Container type	1 = Drink, 2 = Wash, 3 = Ant Trap, 4 = Plant, 5 = Flowerpot, 6 = Others, 7 = Unused
Larvae	0 = No, 1= Yes
Lid	0 = No, 1= Yes
Material	1 = Clay, 2 = Cement, 3 = Plastic, 4 = Aluminum, 5 = Others
Place	1= Inside, 2 = Under eaves, 3 = Outdoor
Container size	1 = <50 liter, 2 = 51-100 liter, 3 = 101-150 liter, 4 = 151-200 liter, 5 = >200 liter

The schematic diagram for this study is shown in Figure 2.2.

Figure 2.2 Schematic diagram of variables of interest



Data Collection and Data Management

The instruments of data collection were constructed based on the forms used for larvae identification by the Ministry of Public Health. The household information in Bangklang, Tarnum, Don, Kork Grabue, Na Kate, Sai Khao, and Napradu village was collected between October and November 1998. Information was composed of two parts. The first part was general information of the household, including drinking water source, washing water source, renewal of drinking water in containers, and renewal of washing water in containers. The second part emphasized information concerning the containers, such as container type, place, material, lid, and size of container in each household. The data were gathered and entered into Microsoft Access.

Statistical and Graphical Methods

Matlab version 5 (Hanselman and Littlefield, 1997) and Asp (McNeil, 1998) were used for graphical presentation and statistical analysis.

Percentages were used to describe Dengue vector indices. They are computed to assess the pattern of larval distribution of Dengue factor calculation of House Index, Container Index, Breteau Index, and Stegomyia Index (Pant and Self, 1993).

House index

This is the percentage of houses with one or more habitats positive for *Aedes aegypti* or related species. It is calculated as follows:

$$House\ Index = \frac{No.\ of\ infected\ house}{No.\ of\ inspected\ house} \times 100 \qquad (2.2)$$

Container index

This is the percentage of infected containers, calculated as follows:

$$Container\ Index = \frac{No.\ of\ infected\ containers}{No.\ of\ inspected\ containers} \times 100 \qquad (2.3)$$

Breteau index

This is the percentage of infected containers among inspected houses, and is calculated as follows:

$$Breteau\ Index = \frac{No.\ of\ infected\ containers}{No.\ of\ inspected\ house} \times 100 \qquad (2.4)$$

Stegomyia index

This is related to the number of infected containers to the human population. So the number of larvae per 1,000 persons is calculated as follows:

$$Stegomyia\ Index = \frac{No.\ of\ positive\ containers}{No.\ of\ people\ living\ in\ the\ premises\ surveyed} \times 1{,}000 \qquad (2.5)$$

1. Statistical Methods

1.1 Odds ratio

The crude odds ratio is a measure of the strength of an association between two binary variables (i.e., in which both the outcome and the determinant are dichotomous) (McNeil, 1996, 1998a, 1998b) (McNeil, et al, 1994).

2 × 2 contingency tables method

Assuming that the data in the contingency table have the counts *a, b, c,* and *d*, the estimated odds ratio is given as follows

X is the determinant and Y is the outcome. Each variable is binary (0 or 1)

	Y = 1	*Y = 0*
X=1	*A*	*b*
X=0	*C*	*d*

Let, $n = a + b + c + d$

The odds ratio (OR) is

$$OR = \frac{ad}{bc} \qquad (2.6)$$

The standard error is given by

$$SE(\ln OR) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \qquad (2.7)$$

The 95% confident interval is

$$95\%\ CI = OR \times \exp(\pm 1.96 SE[\ln OR]) \qquad (2.8)$$

The (Mantel-Haenszel) chi-squared statistic is

$$\chi^2_{MH} = \frac{(ad - bc)^2 (n-1)}{(a+c)(b+d)(a+b)(c+d)} \qquad (2.9)$$

Non-stratified r × c tables

In this study, some of variables are multicategorical. We use non-stratified r×c tables to compare them. For example, *X* is category of village such as transmission/non-transmission or Buddhist/Muslim, and *Y* is category of water consumption characteristics such as type of drinking water source or type of container etc.

Assume X is nominal (1, 2, …, r), Y is nominal (1, 2, …, c).

	$Y = 1$	$Y = 2$	…	$Y = c$
$X = 1$	a_{11}	a_{12}		a_{1c}
$X = 2$	a_{21}	a_{22}		a_{2c}
:				
$X = r$	A_{1r}	a_{2r}		a_{rc}

The odds ratio is

$$OR_{ij} = \frac{a_{ij} d_{ij}}{b_{ij} c_{ij}} \tag{2.10}$$

where, $b_{ij} = \sum_{j=1}^{c} a_{ij} - a_{ij}$, $c_{ij} = \sum_{i=1}^{r} a_{ij} - a_{ij}$, $d_{ij} = n - a_{ij} - b_{ij} - c_{ij}$, $n = \sum_{i=1}^{r} \sum_{j=1}^{c} a_{ij}$

The standard error is given by

$$SE(\ln OR_{ij}) = \sqrt{\frac{1}{a_{ij}} + \frac{1}{b_{ij}} + \frac{1}{c_{ij}} + \frac{1}{d_{ij}}} \tag{2.11}$$

The 95% confidence interval is

$$95\% \ CI = OR \times \exp(\pm 1.96 SE[\ln OR]) \tag{2.12}$$

The (Pearson) chi-squared statistic is

$$\chi^2_{(r-1)(c-1)} = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(a_{ij} - \hat{a}_{ij}\right)^2}{\hat{a}_{ij}} \tag{2.13}$$

Stratified r × c tables

The odds ratios may be of adjusted using the stratified r × c tables method. In this study, these are used to compare larvae in containers with respect to differences between type, place, and size of containers with respect to lid.

Suppose X is nominal (1, 2, ..., r), Y is nominal (1, 2, ..., c), and Z has s levels.

	$Y = 1$	$Y = 2$	...	$Y = c$
$X = 1$	a_{11k}	a_{12k}		a_{1ck}
$X = 2$	a_{21k}	a_{22k}		a_{2ck}
$\vdots$				
$X = r$	a_{1rk}	a_{2rk}		a_{rck}

Define, as before $b_{ijk} = \sum_{j=1}^{c} a_{ijk} - a_{ijk}$, $c_{ijk} = \sum_{i=1}^{r} a_{ijk} - a_{ijk}$, $d_{ijk} = n_k - a_{ijk} - b_{ijk} - c_{ijk}$,

$$n_k = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{s} a_{ijk}$$

Than the odds ratio is given by

$$OR_{ij} = \frac{\sum a_{ijk} d_{ijk} / n_k}{\sum b_{ijk} c_{ijk} / n_k} \tag{2.14}$$

The standard error is given by

$$SE(\ln OR_{ij}) = \sqrt{\frac{1}{a_{ijk}} + \frac{1}{b_{ijk}} + \frac{1}{c_{ijk}} + \frac{1}{d_{ijk}}} \tag{2.15}$$

The 95% confident interval is

$$95\%\ CI = OR \times \exp(\pm 1.96 SE[\ln OR]) \tag{2.16}$$

The (Mantel and Haenszel) chi-squared statistic, for testing, for independent, is

$$\chi^2_{MHS} = \frac{\left(\sum_{k=1}^{s} \frac{a_k d_k - b_k c_k}{n_k}\right)^2}{\sum_{k=1}^{s} \frac{(a_k + c_k)(b_k + d_k)(a_k + b_k)(c_k + d_k)}{(n_k - 1)n_k^2}} \tag{2.17}$$

The (Mantel and Haenszel) chi-squared test for homogeneity,

$$\chi^2_{NHS} = \frac{\left(\sum \frac{(a_{ijk}d_{ijk} - b_{ijk}c_{ijk})}{n_{ijk}}\right)^2}{\sum \frac{(a_{ijk}+c_{ijk})(b_{ijk}+d_{ijk})(a_{ijk}+b_{ijk})(b_{ijk}+c_{ijk})}{(n_{ijk}-1)n^2_{ijk}}} \qquad (2.18)$$

1.2 Logistic regression

Multiple logistic regression analysis is used for adjusting the association between determinant variables and having larvae of Dengue vectors in the container.

Logistic regression is a method of analysis that gives a particularly simple representation for the logarithm of the odds ratio association with risk factors, and when fitted to data involving dichotomous outcome and exposures, it automatically provides estimates of odds ratios and confidence intervals for specific combinations of risk factor. The model is defined as

$$\ln\left(\frac{P}{1-P}\right) = a + b_1x_1 + b_2x_2 \ldots\ldots + b_nx_n \qquad (2.19)$$

In this formula, P denotes the probability of occurrence of the outcome variable and X is the determinant variable, a is the constant coefficient, and b is the set of regression coefficients.

2. Graphical methods

The graphical methods are presented in the following steps.

2.1 Histograms with statistical summaries of raw data for all variables represents the distribution and summaries including the size, mean, standard deviation, minimum and maximum of a set of data.

2.2 Graphs of the odds ratios and 95% confidence intervals can be used to present the association between two nominal categorical variables.