Chapter 2

Methodology

In this chapter we describe the method used in the study. The methodology comprises the information on study design, population and sample, variables and conceptual framework, collection of data, method of data analysis and statistical method.

2.1 Study design

For convenience and internal validity, this study used a cross-sectional study design involving interviews of lower secondary school students from a selected study population.

2.2 Population and sample

The target population comprised all lower secondary school students (M1-M3) in Pattani province, southern Thailand.

The sample consisted of 244 lower secondary school students in Pattani province interviewed between 1 May and 31 October, 2006 from M1-M3.

Sample size calculations were based on the main outcome variable, which was bullying. The prevalence of physical bullying among students from grade four to nine in southern, Thailand was estimated to be 15.1%, from Tapanya (2006). Sample size was calculated by the following formula (McNeil, 1996):

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

PDF created with pdfFactory trial version www.pdffactory.com

where $Z_{a/2}$ is the critical value for the standardized normal distribution corresponding to a two-tail probability α, p is the prevalence of the outcome, and d is half of the width of the 100(1- α) % confidence interval. Choosing a prevalence of 15% based on previous studies and the ability to detect a difference in this prevalence of 5%, with 95% confidence interval, the sample size of the study is then calculated as

$$n = 1.96^2 \times \frac{0.15(1-0.15)}{0.05^2}$$

$$= 197$$

So a sample size of 197 lower secondary school students was needed. However in this study 244 subjects were obtained. Both samples had approximately equal numbers of boys and girls.

2.3 Variables and conceptual framework

The interest variables of the study comprised nine determinants and a binary outcome.

Determinants

The determinants in the study provided information on (a) characteristics of secondary school students (gender, age group, ethnicity, punishment (*defined as history of punishment by parents*) and school type), (b) family environments (father's occupation, mother's occupation, marital status of parents and parental violence (*defined as emotional and physical violence between parents*)).

Outcome variable

Physical bullying was taken as the outcome variable of interest. This outcome was measured by asking secondary school students whether or not they had experienced physical bullying (never bullied others or bullied others). The questionnaire is shown

PDF created with pdfFactory trial version www.pdffactory.com

in the appendix. Figure 2.1 shows a conceptual framework for the variables of interest.

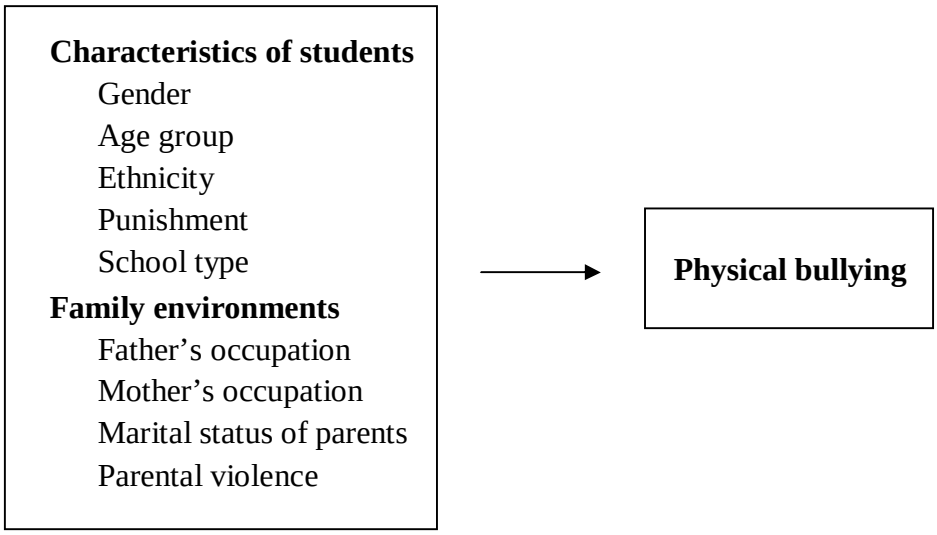


Figure 2.1: Conceptual framework showing variables in the study

2.4 Data collection

Instruments for collecting data

To collect data, a questionnaire was used in order to get information about physical bullying. The questionnaire was designed to determine the students' reported behaviour of bullying in schools. Since the topic was relatively new for Pattani province, there was no available research on which to base the construction of the research instruments. Consequently, the questions in the questionnaire were partly derived from and based on the studies by DeVoe and Kaffenberger (2005). This questionnaire used multiple choice questions and provided information on (a) characteristics of students, (b) family environments, (c) victims of bully and (d) bullying others.

PDF created with pdfFactory trial version www.pdffactory.com

Questions on bullying referred to events that occurred during 2006 in either the previous six months (1 May to 31 October, 2006) or the previous month (in November 2006) preceding the interview. Students answered by checking one or more responses. The bullying inventory measures levels of bullying and being bullied by using two key variables: In the last semester (1 May to 31 October, 2006), have you ever physically bullied other students at your school (slapped, hit, kicked, punched, or weapons)?, and In the last semester (1 May to 31 October, 2006), have you ever been physically bullied by other students at your school (slapped, hit, kicked, punched, or weapons)? Possible responses were “no” or “yes”. If “yes”, then use of a slap, hit, kick, punch, weapons or other was recorded, as well as the frequency.

Data collection

Primary data were collected from the selected lower secondary school students (M1-M3) in Pattani province, southern Thailand. Verbal consent to participate in the study was obtained from students after assurance that confidentially of interview was given. Before the interview, the purpose of the interview was explained, giving assurance of anonymity and stressing the importance of answering truthfully. All participants were interviewed in December 2006 by researchers at a neutral location outside the schools, such as in front of the school, department store or market. The interview took approximately 5-10 minutes to complete. Relevant data were collected, verified and recorded in a separate data record form and used to investigate factors associated with physical bullying.

PDF created with pdfFactory trial version www.pdffactory.com

2.5 Method of data analysis

Data management

The data were entered into an *EpiData* database and transferred to *Microsoft Excel* and finally analyzed using *Webstat* (a set of programs for graphical and statistical analysis of data stored in an SQL database, written in HTML and VBScript).

Data analysis

Statistics for descriptive analysis include percentages for measuring prevalence of bullying. The associations between the outcome and the determinant variables were examined in the preliminary analysis. Of the variables in the study, all were categorical and therefore chi-squared statistics and 95% confidence intervals were used to describe these associations. Logistic regression was used to model the association between the bullying and the determinants.

2.6 Statistical method

Descriptive statistics

The variables of interest are summarized by percentages. All determinants are categorical variables which are described by percentages.

Odds

The odds of an event or outcome is defined as the ratio of the probability of the event to the probability of the corresponding non-event. Here we are using the term "probability" to mean the proportion of outcomes in the population of interest. Thus if an event has probability p and the corresponding non-event has probability 1- p (so that there are only two possible outcomes) the odds of the event is p / (1-p).

PDF created with pdfFactory trial version www.pdffactory.com

The estimated odds of an outcome is obtained by replacing the population proportion p by the sample proportion p.

Graphs of estimated odds and 95% confidence intervals can be used to represent the proportions of physical bullying in the two groups. The graph of odds includes a 95% confidence interval. The confidence interval is graphed as a horizontal line containing a dot denoting the estimated odds. If p is the proportion of outcomes in a sample of size n, the estimated of the odds is (McNeil, 2005)

$$\text{Odds} = \frac{P}{1-P} \qquad (2.2)$$

An asymptotically valid (for large n) formula for the standard error of the log-odds, defined as $\ln(odds) = \ln(p) - \ln(1-p)$, is

$$SE(\ln(odds)) = \sqrt{\frac{1}{np} + \frac{1}{n-np}} \qquad (2.3)$$

A 95% confidence interval (95% CI) for the population odds is thus given by

$$odds \times \exp(-1.96 \times SE),\ odds \times \exp(1.96 \times SE) \qquad (2.4)$$

The estimated odds of an outcome is thus the ratio of the number of outcomes of interest to the number of non-outcomes.

The odds is always greater than the corresponding proportion or probability. But unlike a proportion, which is restricted to being between 0 and 1, the odds can be greater than 1.

PDF created with pdfFactory trial version www.pdffactory.com

Univariate analysis

The method for comparing the population proportions in two groups classified by a binary determinant extends in a straightforward way to situations involving more than two determinant categories. Pearson's chi-squared statistics for independence (i.e., no association) were used to assess the associations between the determinant variables and the outcome of this study. In this study the chi-squared statistic takes the form (McNeil, 2005)

$$c^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \qquad (2.5)$$

where O_{ij} is the observed count in category *i* of the determinant and category *j* of the outcome, and E_{ij} is the corresponding expected count, defined as before by dividing the product of the marginal totals by the overall total sample size, that is

$$E_{ij} = \frac{\sum_{j=1}^{c} O_{ij} \sum_{i=1}^{r} O_{ij}}{n} \qquad (2.6)$$

When the null hypothesis of independence is true, the right-hand side of equation (2.5) has a chi-squared distribution with (*r*-1) (*c*-1) degree of freedom (McNeil, 2005)

Logistic regression

Multiple logistic regression analysis is used for modeling the association between several determinant variables and physical bullying. Logistic regression is a method of analysis that gives a particularly simple representation for the logarithm of the odds ratio describing the association of a binary outcome with factors, and when fitted to data involving a dichotomous outcome and multiple determinants, it automatically provides estimates of odds ratios and confidence intervals for specific combinations

PDF created with pdfFactory trial version www.pdffactory.com

of the risk factor (McNeil, 1996). For a set of predictor variables x_1, x_2,…,x_p and a dichotomous outcome Y the logistic regression model takes the following form:

$$\ln\left(\frac{p}{1-p}\right) = a + \sum_{i=1}^{p} b_i x_i \qquad (2.7)$$

where p denotes the probability of occurrence of the specified outcome and x are the set of determinant variables, a is the constant coefficient, and b is the set of regression coefficients. This equation may be inverted to give an expression for the probability P as

$$\Pr ob[D] = \frac{1}{1+\exp(-a-bx)} \qquad (2.8)$$

The probability of the outcome $Y = 1$ can be expressed as

$$\mathrm{P[Y=1]} = \frac{\exp(a + \sum_{i=1}^{p} b_i x_i)}{1+\exp(a + \sum_{i=1}^{p} b_i x_i)} \qquad (2.9)$$

Using the logistic regression model for the data arising from a two-by-two table, we suppose $x_i = 1$ or 0, that is, the values of determinant X are taken to be 1 (exposure) and 0 (non exposure). Thus the logistic regression model can be written as

$$\ln\left\{\frac{P(Y=1/X=1)}{1-P(Y=1/X=1}\right\} = a + b \qquad (2.10)$$

$$\ln\left\{\frac{P(Y=1/X=0)}{1-P(Y=1/X=0}\right\} = a \qquad (2.11)$$

Equations (2.10) and (2.11) give the (natural) logarithms of the odds for the outcome given the exposure ($x = 1$) and non-exposure ($x = 0$), respectively.

PDF created with pdfFactory trial version www.pdffactory.com

After exponentiating each equation, the odds for the exposed and non-exposed groups can be written as exp(a+b) and exp(a), respectively. The odds ratio is therefore obtained from the simple formula

$$OR = \frac{\exp(a + b)}{\exp(a)} = \exp(b) \qquad (2.12)$$

PDF created with pdfFactory trial version www.pdffactory.com