Chapter 1

Introduction

Many medical and public health issues, using statistical models of varying complexity are used for analyzing these data. Logistic regression (Hosmer & Lemeshow, 2000; Kleinbaum & Klein, 2002) is a statistical method widely used to model the association between a binary outcome. These refer to events that either happen or don't happen, so they comprise factor variables with two levels. When the determinants are categorical factors can be structured as a multi-way contingency table of counts and the data for analysis comprise of adverse outcomes in the cells of this table. This method was applied to the first study, *Length of Stay of Patients Dying in Central Region Hospitals in Thailand*, which presented confidence intervals for adjusted proportions using logistic regression with weighted sum contrasts.

Since, linear regression assumes numerical outcomes. As for linear models, the determinants can be factors base on categorical variables, and the outcome variable is a linear function. If the normality assumption is not plausible, data transformations by take logarithms, the exponential relation takes the linear model can be fitted by Log-transformed linear regression. Next, the generalized linear models (GLMs) with Poisson and negative binomial assume outcomes are counts. Poisson and negative binomial distributions arise naturally as random counts with population-at-risk denominators. These statistical methods were used to the second study, *Muslim Victims of Terrorism Violence in Southern Thailand.* The incidence rate of Muslim victims as the outcomes are counts per person-year, unit population at risk in a

specific period of time, which can be used to compare the relative risks of the outcome for different factors such as sex, age groups, regions and years.

1.1 Rationale for study

Health expenditure increases substantially with LOS. The time spent in hospital by patient differs according to disease, treatment facility, cost of treatment and discharge status. Hospital stay terminated by death is an important outcome event, with death usually reflecting the severity and burden of disease.

LOS data has a highly positive skewed distribution (Xiao et al., 1999; Lee et al., 2003; Kulinskaya et al., 2005). The statistical analysis LOS data using Logistic regression model and log-linear regression model were used to handle skewed in the LOS distribution.

However, the empirical distribution of LOS is positively skewed so LOS models have used logistic regression to avoid symmetry assumptions (Ruttimann et al., 1996; Lim et al., 2009). Through a better understanding of the factors affecting longer LOS, it will be useful for health care services, treatment during hospitalization, and the allocation of resources on the basis of health care service at hospitals. By the way, to examine the variation in LOS of patients dying in hospital in the Thai central region, with respect to the patient's principal diagnosis and demographic status, and hospital size and geographic location were taken into account in the model.

Also, Terrorism violence is an important public health issue and in many respects is similar to a deadly disease that can reach epidemic proportions. The statistical models were used to describing the incidence rate of injuries to civilian resident victims of violence from terrorism in the target area defined as Pattani, Yala and Narathiwat

provinces and four eastern districts of Songkla province in Southern Thailand. For six years, from Jan 1, 2004 until December 31, 2009 there were 4,143 Muslim residents and 3,544 other (mainly Buddhist) residents of the target area have been recorded as victims by the Deep South Coordination Centre (DSCC). We focused on the Muslim population and fitted negative binomial and log-normal models to incidence rates classified by gender, age group, region and year, with the objective of comparing relative risk by these factors, after adjusting for other factors to remove confounding.

1.2 Literature review

This review covers articles that related with the 2 studies (LOS model and terrorism victims). The statistical issues include generalized linear models with logistic, Poisson and more general negative binomial distributions, a log-transformed data analysis. With hospital Length of stay (LOS), a common parameter used to indicate health resource utilization, health care cost and severity of disease (Wang et al., 2002; Lee et al. 2003). Based on a literature review Martin and Smith (1996) concluded that patient demographics and hospital characteristics were the two major factors that determine patient LOS. Among demographic characteristics, studies have reported that LOS varied according to age and disease group (Goldfarb et al., 1983; McMullan et al., 2004), whereas among hospital characteristics, LOS has been reported to vary by region, hospital size, and health care service (Xiao et al., 1997; Clarke, 2002).

LOS can be terminated by cure, transfer or death discharge. Many studies have been concerned only with LOS for patients with cured discharge. However, hospital stay terminated by death is also an important outcome event. Patient care in hospital is the most expensive way of providing palliative care (Huang et al., 2002). Longer stays are

more likely to indicate physicians' decisions or administrative inefficiencies than patients' need (Brownell & Roos, 1995). Reducing LOS is health policy in many countries (Clarke & Rosen, 2001). Setting up a proper palliative care or providing opportunity for patients to decide where they want to stay during the last stages of their life can reduce unnecessary hospital LOS. For inpatients spending their final days in hospital in developed countries such as Canada, England and Belgium, LOS has been studied by Huang et al (2002), Dixon et al. (2004) and Van Den Block et al. (2007), but similar studies for developing countries are limited.

Clark and Ryan (2002) developed a method for predicting length of stay for seriously ill or injured patients. They suggested that piecewise exponential models, which are adapted from the Cox's proportional hazard methodology for estimating competing risk, possibly would be useful in predicting length of hospital stay, especially if determinants of mortality are separated from determinants of live discharges. This finding is consistent with a study conducted by Sa et al (2007). They demonstrated that estimates of the standard survival models and competing risk models for length of hospital stay, with and without control for unobserved patient heterogeneity, can lead to a different conclusion.

Terrorism, a global public health burden occurred in many countries such as Iraq, Afghanistan, Pakistan, India and Sri Lanka, South America (Shane, 2007; South Asia Terrorism Portal, 2011; Baez et al., 2007) with emergence to southern Thailand. The most civilian residents have been victims of terrorism while the causes vary; the results are the same as death, disability, and suffering. Thailand's majority of population is ethnically Thai and religiously Buddhist. But three Thailand's southernmost provinces include Pattani, Yala, Narathiwat, and occasionally in part of

Songkla have taken place in terrorism violence, with proximately 80 percent of the population is Muslim. The difference religious and traditional culture is referred to connection with the conflict (Croissant & Trinn, 2009; Yusuf, 2007). The conflicts have been escalating and more violent confrontations took place more intensively from January 2004 up to the present time.

The studies that related to health and violence was conducted by Marohabout et al. (2009) that fitted a statistical model to events classified by location and month (in 2004 and 2005), using data files provided in police reports in the terrorism target area (defined as the three provinces Pattani, Yala and Narathiwat and the four easternmost districts of Songkla province). Lim et al. (2009) investigated the living conditions of victims' families from the unrest in Pattani, Thailand, and to evaluate the success of the healing victims' project. However, security in life and assets, mental health care, and financial supported need to urgent aid with healing strategies for the victims' families. Khongmark et al. (2011) fitted models and graphs injury incidence rates over regions for non-Muslims in the target area for years 2004-2009 inclusive, using data recorded in the database of the Deep South Coordination Centre (DSCC). Jitpiromsri (2010) provides statistical graphs and summaries of the 9,446 terrorism incidents resulting in approximately 4100 deaths and 6,500 non-fatal injuries, again using the Deep South Watch database for the 73 months from January 2004 to January 2010 inclusive.

Statistical issues

Log-transformed regression

Based on a literature review Kulinskaya et al. (2005) used Length of stay as a performance indicator of finding the most appropriate statistical analysis for data of the LOS type is addressed by comparing standard general linear model (GLM) methods and Truncated maximum likelihood (TML) using ln(LOS) as the dependent variable. However, before performing any analyses, the data will be trimmed to remove outliers and all same-day cases (LOS=0) were omitted.

Chertow et al. (2005) analyzed the marginal effects of acute kidney injury on hospital mortality, length of stay, and costs. Linear regression was used to evaluate the association of AKI with LOS and costs, log-transformed LOS and costs to accommodate data that were expectedly right skewed and used Mallow's Cp to access model fit. Residual diagnosis indicated few outliers, multiplicative interaction terms of AKI tested with other model covariates to evaluate for effect modification.

A Generalized Linear Mixed Model (GLMM) for the number of deaths was used by Lix et al (2006) conducted a study in Canada to test differences in mortality for small geographical regions over time using a generalized linear model with generalized estimating equations. The model provides one analytical tool for examining small area inequalities in health status.

Logistic regression

Logistic regression is the appropriate method for analyzing data from studies where the outcome variable of interest is a proportion. Lim and Tongkumchum (2009) focused on the study of length of stay for 40,498 mortality cases in hospitals in

southern Thailand from 2000 to 2003, with respect to age, gender, principal diagnosis, provinces and hospital size. Logistic and linear regression with log-transformed LOS was used to analyze the data. Both logistic regressions with LOS at least one week as the outcome and linear regression on appropriately log transformed LOS gave consistent results.

According to a model fitted, a general method of graphical display shows 95% confidence intervals for comparing several proportions after adjusting for categorical covariates from logistic regression model. This method is used with complicated adjustments for covariates (Rachatapantanakorn & Tongkumchum, 2009; Kongchoy & Sampantarak, 2010).

Poison and negative binomial

For count data, the Poisson and negative binomial generalized linear models are usually considered to be most statistically appropriate. Kongchoy et al. (2010) examined the trend, seasonal and geographic effects on tuberculosis (TB) incidence in southern Thailand from 1999 to 2004. Two models were compared; the TB incidence rates were modeled using negative binomial distribution for the number of cases and log-linear distribution for the incidence rate. However, the log-linear regression model could be used as a simple method for modeling TB incidence rates.

Marohabout et al. (2009) fitted negative binomial regression to incidence rates of terrorist event classified by location and time (moth in 2004 - 2005) occurring in the population in southernmost provinces of Thailand. Khongmark et al. (2011) used methods for modeling and graphically comparing incidence rates of adverse events that vary over both space and time and demographic characteristics (gender and age

group). Using two statistical models analyzed incidence rates, one assuming that the adverse event cells counts are fitted by a negative binomial distribution, the other assuming that the incidence rates follow a log-normal distribution. Each model provides adjusted incidence rates and confidence intervals for comparing subgroups with an overall mean.

1.3 Definition of terms

Length of stay (LOS) is a term commonly used to measure the duration of a single episode of hospitalization. Inpatient days are calculated by subtracting day of admission from day of discharge. However, persons entering and leaving a hospital on the same day have a length of stay of one (Wikipedia, 2011).

Victim is a person who suffers from a destructive or injurious action or death in unrest situation of Thailand's Deep South violence.

Terrorism refer to those violent acts which are intended to create fear (terror), are perpetrated for a religious, political or, ideological goal.

1.4 Objectives and plan of thesis

According to the outcome of interest such as numeric or counts, appropriate statistical models are used to analyze the data. These models attempted to identify the associations among demographic factors (gender, age and principal diagnosis), geographic factors (place), time and other characteristics with the outcomes. Also, log-linear regression, logistic regression, poison regression and negative binomial regression models were applied to fit these data.

The main objectives of studies were thus as follows.

1. To investigate the variation between factors of interest and outcome.
2. To develop statistical modeling for medical and public health data.

In this thesis contains four chapters. The introductory chapter discusses the rationale, the scope and the aim of the study, and also includes a review of some relevant literature. Chapter 2 includes the statistical methods for data analysis and the statistical models. Two manuscripts were shown in chapter 3, with the topic of "Length of stay of patients dying in central region hospitals in Thailand" and "Muslim Victims of Terrorism Violence in Southern Thailand". The last chapter gives overall summaries and conclusions. Suggestions for further research are also provided in this chapter.