Chapter 4

Conclusions and Discussion

This chapter summarizes the main results of the two studies included in Chapter 3 that some informative preliminary analysis of the data is not included in the publication due to journal space limitations and is thus included in this chapter. The first section of this chapter summarizes the overall findings from the two studies. The second section discusses the implications of the studies. In addition, limitations and suggestions are given in the last section.

4.1 Overall findings

The first study: Length of Stay of Patients Dying in Central Region Hospitals in Thailand.

A large sample sizes, 257,076 records of mortality cases were included to determine the effect of factors on LOS for the period of October 1999 to September 2007, reported from provincial hospitals in Central Thailand to the Ministry of Public Health. We considered principle diagnosis (ICD-10), age, gender, the hospital bed size, and the hospital region as predictors. Principle diagnosis was classified into nine disease groups (injuries, cerebrovascular disease (CVD), digestivesystem, infectious disease, respiratory disease, geneto-urinary disease (GUD), respiratory infection, malignant neoplasms, and other diseases) having median length of stay 2, 3, 4, 4, 5, 6, 6, 8 and 4 days, respectively. Since the corresponding distributions are heavily skewed we transformed LOS to ln(1+LOS), and found all predictors to be highly statistically significant in a multiple regression model. Also with the five factors of interest were found to have a statistically significant association with LOS based on chi-square test and multiple logistic regressions. As multiple logistic regressions model with weighted sum contrasts were used to compare proportions for each factor after adjusting for categorical covariates with an overall mean.

Distributions of LOS

Figure 4.1 shows a histogram of the distribution of ln(LOS+1). The lognormal distribution, being a continuous curve cannot accommodate the data where LOS discrete values 0 or 1, although if these shorter LOS were code in hours rather than days they might well be accommodate in the model curve.

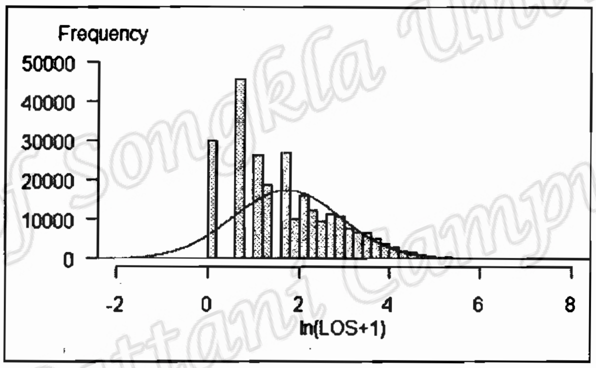

Figure 4.1: Distribution of ln(LOS+1) for hospital mortality in Central Thailand 1999 - 2007.

Regression models

Table 2 gives the estimated coefficients and their standard errors, t value and p-value for this linear model of the factors of interest, i.e. disease groups, gender, age groups, hospital bed size and hospital regions while give multiple R-squared: 0.07631, adjusted r-squared: 0.07624 by F-statistic: 1180 on 18 and 257057 DF, and p-value: <0.0001.

Figure 4.2 shows the adequacy of the model by the observed and residual plot against normal score plot.

Determinant	Coefficient	Std.error	t value	*p*-Value
Constant	1.8371	0.0078	236.49	< 0.0001
Disease group:				
Infectious diseases	Reference	-	-	-
Injuries	-0.3165	0.0089	-35.59	< 0.0001
CVA	-0.2897	0.0069	-42.27	< 0.0001
Digestive system	-0.0635	0.0098	-6.50	< 0.0001
Respiratory Disease	0.0703	0.0108	6.48	< 0.0001
GUD	0.1330	0.0124	10.76	< 0.0001
Respiratory Infection	0.2064	0.0098	21.00	< 0.0001
Malignant neoplasm	0.4027	0.0080	50.16	< 0.0001
All other disease	0.1338	0.0080	16.75	< 0.0001
Gender:				
Male	Reference	-	-	-
Female	0.0219	0.0046	4.78	< 0.0001
Age group:				
0-59	Reference	-	-	-
60-74	0.2002	0.0054	36.82	< 0.0001
75+	0.3460	0.0061	56.51	< 0.0001
Hospital bed size:				
0-60 beds	Reference	-	-	-
61-500 beds	-0.0540	0.0092	-5.87	< 0.0001
501+ beds	0.0971	0.0054	17.97	< 0.0001
Hospital region:				
Bangkok	Reference	-	-	-
South West	-0.2530	0.0082	-30.83	< 0.0001
North West	-0.3865	0.0081	-47.70	< 0.0001
North	-0.3143	0.0079	-40.03	< 0.0001
East	-0.3050	0.0079	-38.73	< 0.0001
South East	-0.4223	0.0071	-59.74	< 0.0001

Table 4.1 : Results for models fitted to ln(LOS+1).

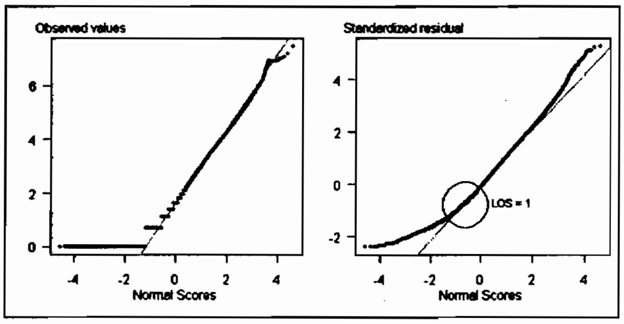

Figure: 4.2 Normal score plot against Standardized residual

The second study: Muslim Victims of Terrorism Violence in Southern Thailand

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. Over the 6 years from 2004-2009, there were 4,143 Muslim residents and 3,544 other (mainly Buddhist) residents in this area suffering injuries or death were officially designed as violence 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.

The frequency distributions of the characteristics of Muslim victims by gender in three age-groups also indicated the incidence rates per 100,000 populations which caused injury or death as shown in Table 4.2. Males had higher incidence rate than females (incidence rates 29–169 and 8–24 per 100,000, respectively). The heist risk had occurred in age-group 25 – 45 years.

Age group	Gender		Total
	Male	Female	
< 25	699 (29)	186 (8)	885 (19)
25 - 45	1,962 (169)	293 (24)	2,255 (95)
> 45	914 (116)	89 (11)	1,003 (63)
Total	3,575 (82)	568 (13)	4,143 (48)

Table 4.2: Characteristics of the Muslim victims and incidence rates per 100,000 populations classified by gender and age-group which caused injury or death

Table 4.3 shows the region-year risks of terrorist attacks to Muslim victims, there were the highest risks, especially in 2007, 2008 and 2009 (incidence rates 98, 53 and 48 per 100,000, respectively). The overall risk of becoming a victim for Muslim is 48 per 100,000.

RegionID/ Districts	Years (incidence rates:100,000 population)						
	2004	2005	2006	2007	2008	2009	Total
1: Chana/Thepha	2 (2)	1 (1)	9 (10)	9 (10)	3 (3)	2 (2)	26 (5)
2: SabaYoi/Na Thawi	26 (54)	4 (8)	2 (4)	57 (118)	4 (8)	6 (12)	99 (34)
3: Mueang Pattani	23 (34)	16 (24)	14 (21)	43 (64)	27 (40)	51 (76)	174 (43)
4: Kok Pho/Mae Lan	16 (39)	7 (17)	19 (47)	23 (56)	29 (71)	23 (56)	117 (48)
5: Nong Chik/Mayo/Kapho	3 5)	10 (16)	16 (26)	21 (34)	42 (69)	30 (49)	122 (33)
6: Yaring	5 (6)	16 (20)	8 (10)	35 (44)	32 (40)	47 (59)	143 (30)
7: Panare/Sai Buri/ Mai Kaen	2 (2)	11 (12)	24 (27)	54 (61)	48 (54)	64 (72)	203 (38)
8: ThungYang Dang	1 (1)	8 (11)	10 (14)	30 (43)	22 (32)	20 (29)	91 (22)
9: Yarang	3 (4)	14 (19)	24 (32)	38 (51)	29 (39)	39 (53)	147 (33)

10: Mueang Yala	35	61	57	145	65	28	391
	(44)	(77)	(72)	(183)	(82)	(35)	(82)
11: Betong/Than To	16	2	21	27	10	17	93
	(51)	(6)	(67)	(86)	(32)	(54)	(49)
12: Raman	10	30	31	86	98	22	277
	(18)	(55)	(57)	(158)	(180)	(40)	(85)
13: Yaha/Kabang/Krong	6	14	19	149	50	43	281
Pinang	(12)	(28)	(38)	(295)	(99)	(85)	(93)
14: Bannang Sata	32	23	51	176	84	59	425
	(46)	(33)	(73)	(252)	(120)	(84)	(101)
15: Mueang Narathiwat	6	12	9	43	28	8	106
	(8)	(17)	(12)	(59)	(39)	(11)	(24)
16: Tak Bai	132	4	4	13	11	7	171
	(288)	(9)	(9)	(28)	(24)	(15)	(62)
17: Bacho/Yi-ngo	2	9	13	35	16	63	138
	(2)	(110)	(16)	(42)	(19)	(76)	(28)
18: Rueso	2	18	60	107	31	38	256
	(4)	(34)	(113)	(201)	(58)	(71)	(80)
19: Rangae	6	42	29	133	39	32	281
	(9)	(60)	(42)	(191)	(56)	(46)	(67)
20: SiSakon/Chanae	3	5	11	59	22	26	126
	(6)	(10)	(22)	(118)	(44)	(52)	(42)
21: Sukirin/Waeng	14	6	8	18	41	8	95
	(27)	(12)	(15)	(35)	(79)	(15)	(30)
22: Su-ngaiPadi/Cho-airong	10	55	44	69	20	55	253
	(13)	(73)	(58)	(91)	(26)	(73)	(56)
23: Su-ngaiKolok	17	21	11	46	20	13	128
	(41)	(51)	(27)	(111)	(48)	(31)	(52)
Total	372	389	494	1416	771	701	4,143
	(26)	(27)	(34)	(98)	(53)	(48)	(48)

Table 4.3: Characteristics of the Muslim victims and incidence rates per 100,000 populations classified by region and year which caused injury or death

A bar chart shows the overall risks of becoming a victim varies substantially with region (Figure 4.2). The region with the highest risk of becoming a victim in Bannang Sata, Yaha / Kabang / Krong Pinang, Raman, Mueang Yala and Rueso (with incidence rate per 100,000 was 101, 93, 85, 82 and 80, respectively), while the lowest incidence rate in Chana and Thepa was 5 per 100,000. But the risk also varies by gender, religion, age-group and year.

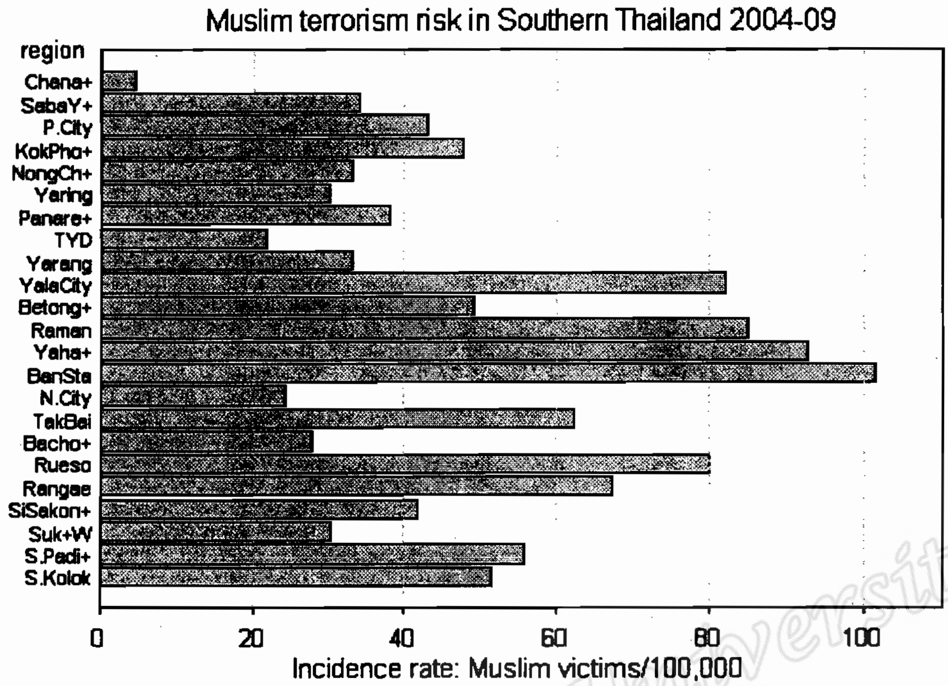

Figure 4.2: Risk of becoming a Muslim victim classified by region

A clustered bar chart can show such demographic variation (Figure 4.3). This graph gives a better indication of how the risk varied, with men aged 25 or more in the highest risk group, particularly in Yala province.

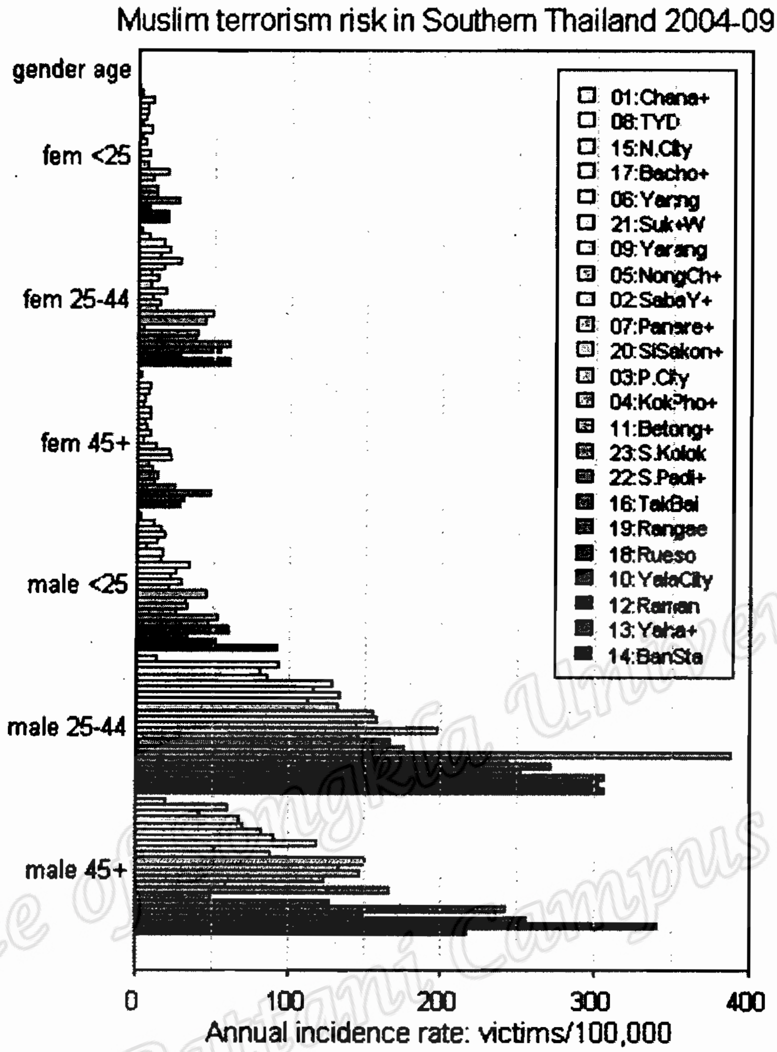

Figure 4.3: Risk of becoming a Muslim victim classified by gender and age group

A similar method can use to compare annual risks of injury from terrorist events for the residents in all 23 regions of the target area (Figure 4.4). These confidence intervals enable the regions to be classified into three risk groups:(a) above average, (b) average, and (c) below average. The information in the comparison confidence intervals is used to create stacked bar charts that classify the categories into three color-coded risk groups: (a) below average, (b) not evidently different from average, (c) above average (Figure 4.5).

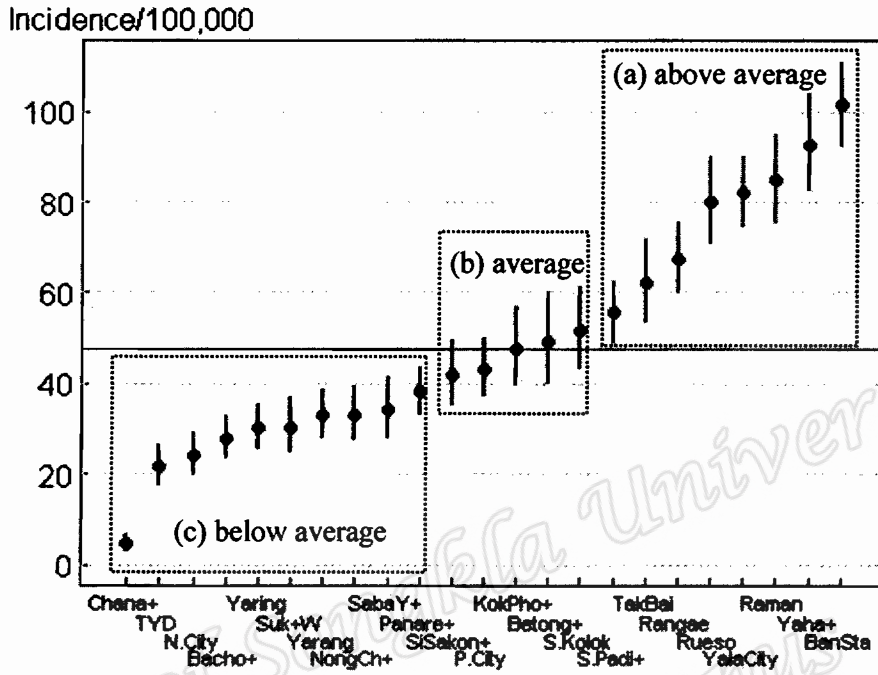

Figure 4.4: Three Risk groups of becoming a Muslim victim classified by region

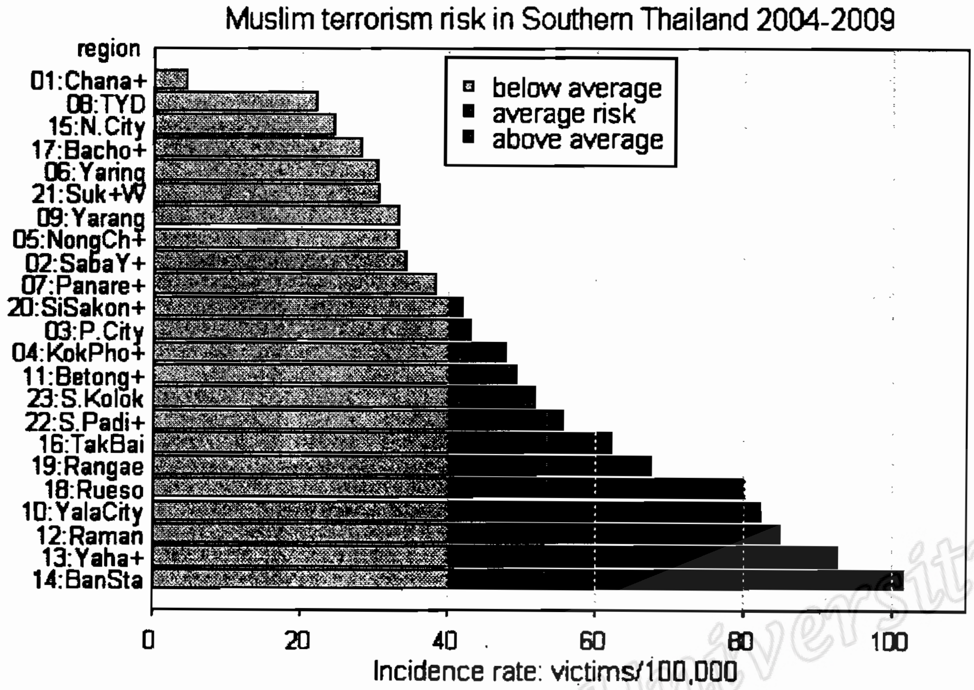

Figure 4.5: Stacked bar charts in three color-coded risk groups of becoming a Muslim victim classified by region

A method for modeling highly skewed count data is to transform the incidence rates by taking their logarithms, after replacing zero counts by 0.5 to ensure finiteness, to remove skewness in the residuals and just fit a linear regression model. This method is simpler and computationally faster, and can take advantage of the much more extensive body of theory supporting linear multiple regression models. The Poisson and negative binomial models also have problems with zero counts, failing to converge in situations where no events occur for a level of a factor (in contrast to linear models). This problem was circumvented by making a minimal change in the data by shifting a single non-zero count from one level to a neighboring level, thus keeping the total number of events constant.

4.2 Discussion of the implicated results

From the first study, the conventional linear model with logarithms transforming LOS, when used appropriately, give conformable results and validly identifies factors associated with LOS. However, LOS data contains outliers, extreme data and skewness (Xiao et al., 1999; Lee et al., 2003; Kulinskaya et al., 2005). Also, the lognormal distribution, being a continuous curve, cannot accurately accommodate data where LOS takes discrete values 0 or 1, if these shorter LOS were coded in hours rather days they might well be accommodated in the model curve (Lim & Tongkhumchum, 2009). Although the assumption of normality of the residuals was considerable satisfied and the residual plot from linear model was slightly satisfied but with low r-square of 7.63 %. This may because of the linear model was analysed by using individual data which have high variation between each individual.

We found that using a multiple logistic regression model to describe the factors associated with LOS of at least one week provided results consistent with a log-linear regression model. However, logistic regression has an advantage of no assumptions related to the distribution, a linear relationship, or equal variances that has seen in the first paper.

In the second study we used gneralized linear models (GLMs) were developed by McCullagh and Nelder (1989). They extend linear models by allowing non-normal outcome variables, through the influence of a single linear function of the determinants x_1, x_2, … (Venables and Ripley 1902). As for linear models, the determinants can be factors based on categorical variables. Poisson and negative binomial (that generalizes the Poisson) fit to the injury incidence rates from the terrorism violence to Muslim victims in the southern Thai region for 6 years (2004-2009).

Using combined two complex categorical factors: (1) age group & sex (6 levels); (2) region & year (90 levels) that were used to model fitted adjusting for region and year effects. The negative binomial and transformed linear models fit equally well in overall. There is little difference between the two models. There will be situations where one model fits better than the other, or where one model gives much shorter confidence intervals than the other.

An advantage of linear models with transformed outcomes is that they cover a wider range of variance-to-mean relations. Jansakul and Hinde (2004) developed a model with a different variance-to-mean formula than the negative binomial GLM, but it requires special software usually not available with standard packages. In this study we used R function for modeling and graph confidence interval from the model fitted.

4.3 Limitations and suggestions for further study

Although not all hospital deaths are reported by hospitals to the NHSO, and individual reporting practices differ from hospital to hospital and also over time. The data used in the first study might be criticised for being incomplete but it does represent a very high proportion of all deaths in hospitals in the region studied with a large sample size, which is required for modeling.

These findings could be useful for hospital management, particularly for prioritizing health care policies and improving health services, including when deciding the most appropriate allocation of health resources. It could be helpful to have more information about LOS with respect to patients' health conditions, and demographic and geographic factors. Further study is also needed to determine the relationship between regional LOS differences and health outcomes including health expenditures.

In the second study, the models gave different results, but each showed that while specific regions were at higher risk at different times and these patterns could not be easily predicted, risks in different demographic groups remained relatively constant. Both models fail to handle zero counts. Estimates of adjusted incidence rates in both models need to be sealed to ensure that the overall mean incidence rate is the same before and after adjustment for covariates. A problem with Poisson GLMs is that zero counts give rise to parameter estimates that converge to minus infinity. As the cases are combined factors gave zero count in the cells, which there were no injuries recorded. This "hole" in the data causes the Poisson GLM to fail to converge.

The limitation of the study is that injury incidence rates varied substantially by gender and age group, region and to a lesser extent by year. Also, some substantial interactions between the region-year factor and the other factors existed but it was possible to take account of them because it would have increased the number of parameters to such an extent that the model could not be easily fitted using available software. For the further study needed to determine the comparing relative risk in each region and year by the unadjusted ("crude") incidence rates with adjusted incidence rates, after adjusting for other factors to remove confounding.