Chapter 4

Statistical Modeling

This chapter describes statistical modeling used for estimating ill-defined mortality rates in Thailand during year 2000 to 2009. Negative binomial regression was chosen for modeling and identifying the strength of association between outcome and determinants as it provided the best fit. Sum contrasts were used to obtain confidence intervals for each level of each factor enabling comparison with the overall mean.

4.1 Statistical models

Ill-defined mortality rate was transformed by taking national logarithm as it had positively skew. Zero counts were replaced by 0.5 to avoid taking the logarithm of zero. After fitting the log-linear model the r-squared of the model was 90.9%. However, the residual plot shows few residual values depart from the diagonal line as shown in the top-left plot of Figure 4.1. Thus the normality assumption of residual was not satisfied. This means that the model was not fit well with the data. Log-liner model was no longer appropriate for this data.

Poisson model was then considered. We used number of deaths as its outcome with corresponding population per 100,000 as its offset. After fitting the model, residual deviance was 133950.4 on 1372 degrees of freedom as show in the top right plot of Figure 4.1. This model gave over-dispersion result. Then Poisson regression model was also no longer appropriate for this data.

Negative binomial regression was the model considered for fitting model further. The residual deviance of this model was 1450.1 on 1372 degrees of freedom with no over dispersion. Therefore negative binomial regression was considered as a better choice for fitting the data.

Figure 4.1 shows plots of residuals versus the normal quartile. The standardized residuals plot for the log linear regression model is on left panel, the deviance residual plot for Poisson regression is on right panel, and negative binomial regression is on bottom. As shown in this Figure, negative binomial model is more appropriate for modeling ill-defined death than log linear regression and Poisson regression model with less residual values depart from diagonal line resulted in more normally distributed residuals. However, two outliers appeared on the top-left of plots which are both genders in aged 0-9 years in Bangkok year 2000.

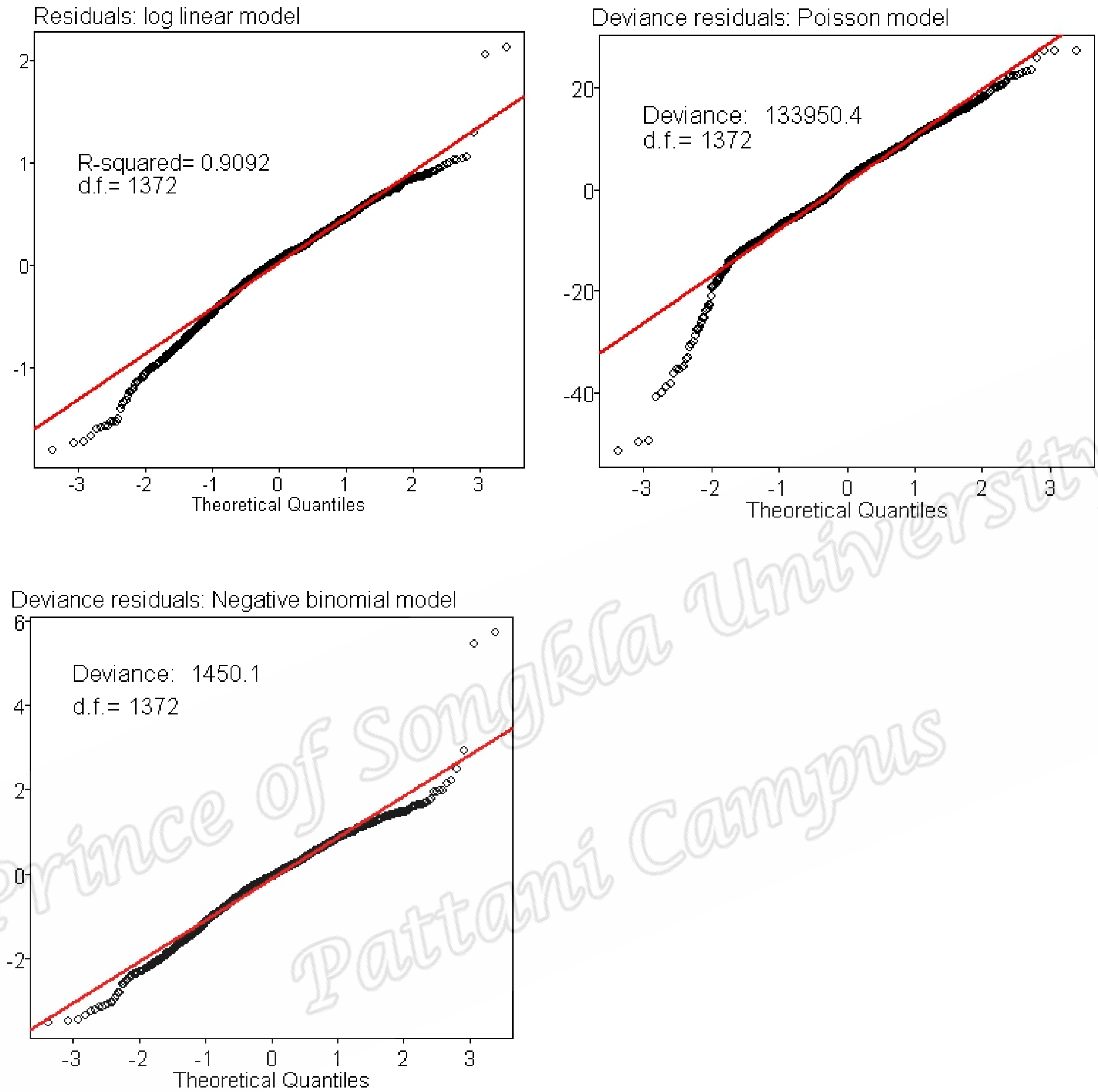


Figure 4.1: Residuals plot for log linear model (left panel), Poisson model (right panel), and negative binomial model (bottom panel)

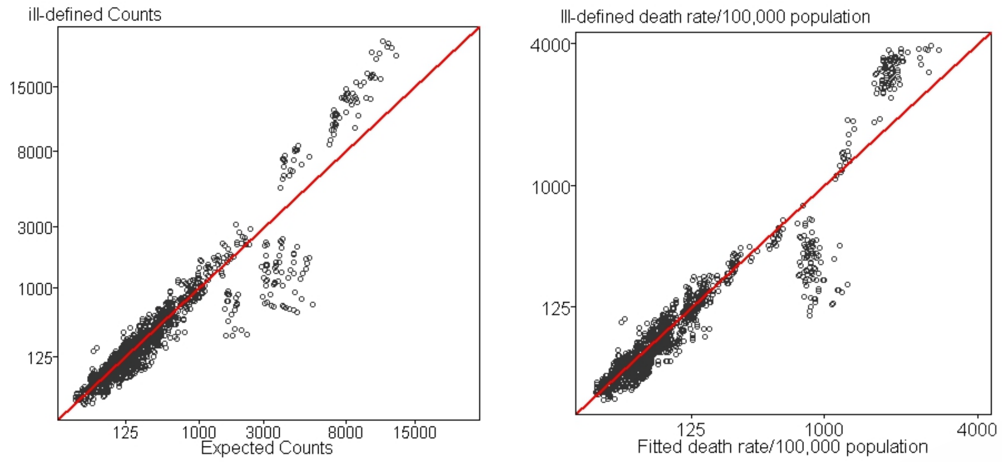


Figure 4.2: Plot of observed counts (left panel) and observed death rates (right panel) against fitted values from negative binomial model

Figure 4.2 shows plots of ill-defined observed counts (left panel) and ill-defined observed death rates (right panel) per 100,000 population against expected death rates obtained from negative binomial model. As shown in this Figure, negative binomial model fit with the data quite well.

Table 4.1: Coefficients, standard errors and p-values of ill-defined mortality rate in Thailand from negative binomial regression model.

Factor	Coefficient	SE	P-value
Constant	3.329	0.012	<0.001
Gender: Age group			
Male: 0-9	-0.852	0.045	<0.001
Male: 10-19	-1.593	0.046	<0.001
Male: 20-29	-0.439	0.044	<0.001
Male: 30-39	0.235	0.044	<0.001
Male: 40-49	0.579	0.044	<0.001
Male: 50-59	1.147	0.044	<0.001
Male: 60+	3.115	0.044	<0.001
Female: 0-9	-1.023	0.045	<0.001
Female: 10-19	-2.216	0.048	<0.001
Female: 20-29	-1.323	0.045	<0.001
Female: 30-39	-0.809	0.045	<0.001
Female: 40-49	-0.340	0.044	<0.001
Female: 50-59	0.465	0.044	<0.001
Female: 60+	3.054	0.044	<0.001

Table 4.1 (Cont.)

Factor	Coefficient	SE	P-value
Region			
region 1	-0.349	0.025	<0.001
region 2	0.072	0.025	0.003
region 3	0.052	0.025	0.033
region 4	0.157	0.025	<0.001
region 5	0.067	0.025	0.007
Year			
2000	0.287	0.021	<0.001
2001	0.026	0.021	<0.001
2002	0.033	0.021	<0.001
2003	-0.036	0.021	0.135
2004	-0.007	0.021	0.410
2005	-0.025	0.021	0.076
2006	-0.019	0.021	0.000
2007	-0.544	0.021	<0.001
2008	-0.098	0.021	<0.001
2009	0.416	0.012	<0.001
Place of death			
Inside-hospital	0.416	0.012	<0.001
Outside-hospital	-0.415	0.012	<0.001

Table 4.1 shows the coefficients of parameters and their standard errors from negative binomial model. The results showed that gender-age group, region, year and place of death were statistically significantly association with ill-defined death rates.

4.2 Confidence interval

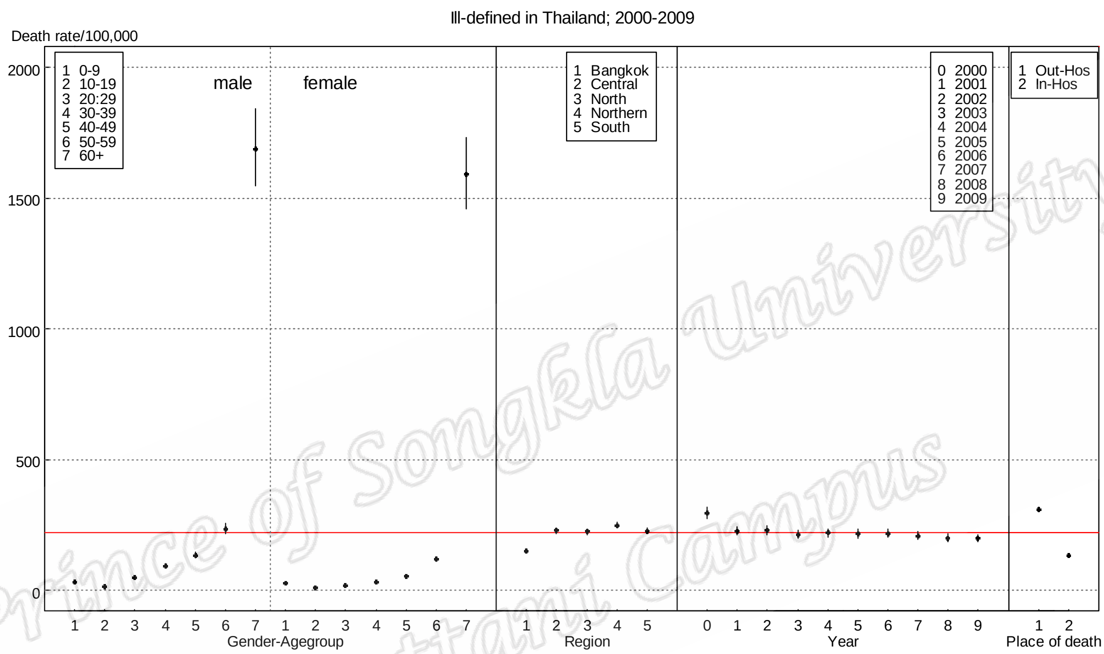
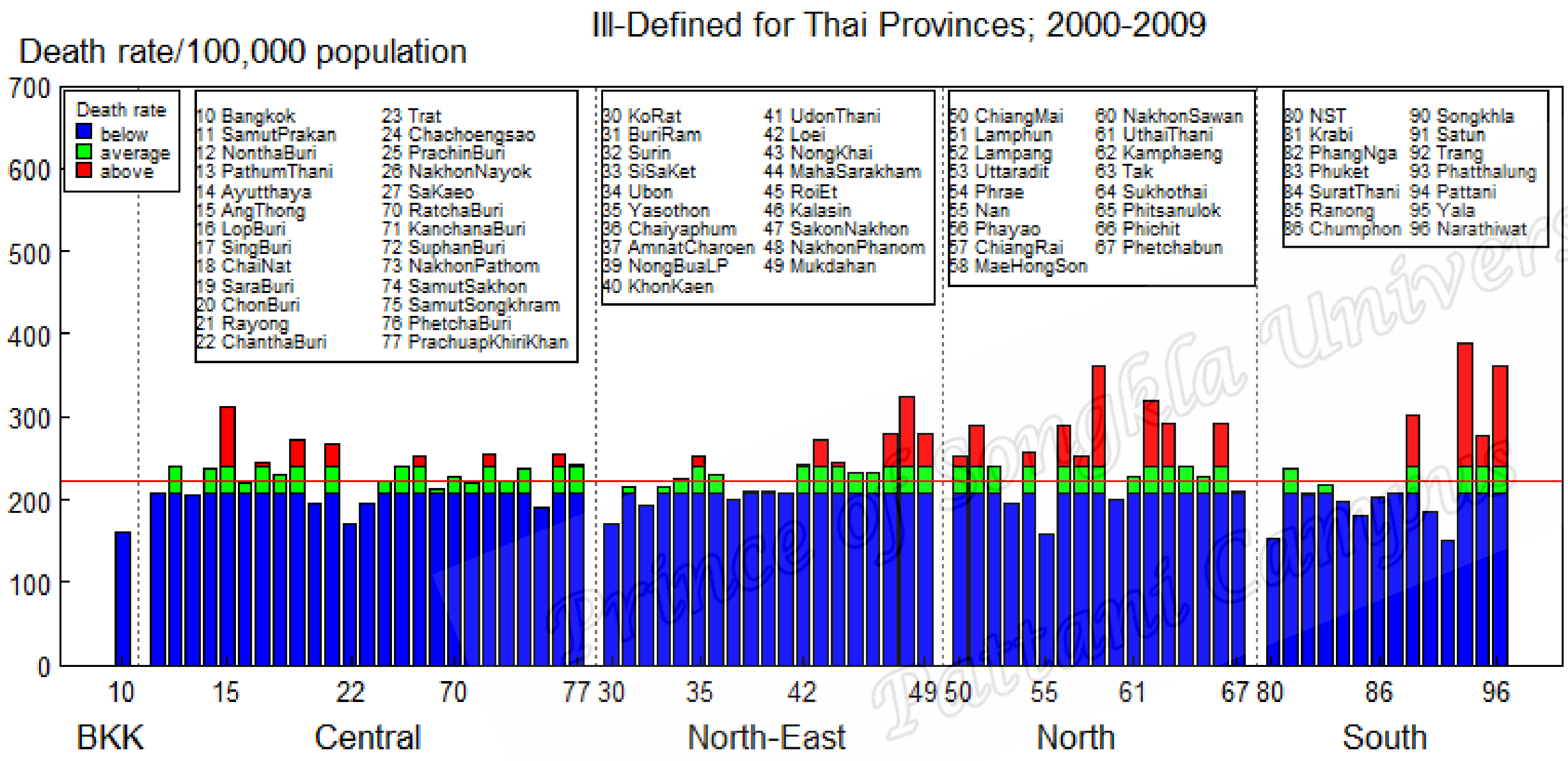


Figure 4.3: 95% Confidence interval plot for region, gender-age group and year of ill-defined death rates per 100,000 population during the period 2000-2009

Figure 4.3 shows the 95% confidence interval of death rates per 100,000 population for ill-defined from negative binomial model. Gender-age group, region, year and place of death were statistically significantly associated with ill-defined death rates. The red horizontal line on the graph represents the overall mean of mortality rate which was 222.7 per 100,000 population.

Age group in 0-9 to 50-59 years had lower ill-defined death rates than the average rate for both sexes whereas age group 60 and over had higher death rate than the average rate. An increasing trend was found in age-group 10-19 until aged 60 years and over for both sexes. Ill-defined death rate by in age group 10-19 years was lower than the average. Ill-defined mortality rate in Bangkok was lower than of other regions. Death rate in the North was higher than average death rate. Ill-defined death rates in year 2000 was higher than average whereas death rates in 2009 was lower than average. Ill-defined death rate outside hospital was slightly higher than average.

Figure 4.4 shows the bar chart for ill-defined death rates per 100,000 population by region based on confidence intervals from negative binomial regression model. The graph shows that ill-defined was found to be the leading cause of death in the Central (Ang Thong, Sing Buri, Sara Buri, Rayong, Nakhon Nayok, Suphan Buri and Phetcha Buri), the Northeast (Yasothon, Nong Khai, Maha Sarakham, Sakon Nakhon, Nakhon Phanom and Mukdahan), the North (Chiang Mai, Lamphun, Phrae, Phayao, Chiang Rai, MaeHongSon, KamphangPet, Tak and Phichit) and the South regions (Satun, Pattani, Yala and Narathiwat).



Figures 4.4: Bar chart of ill-defined death rates per 100,000 population

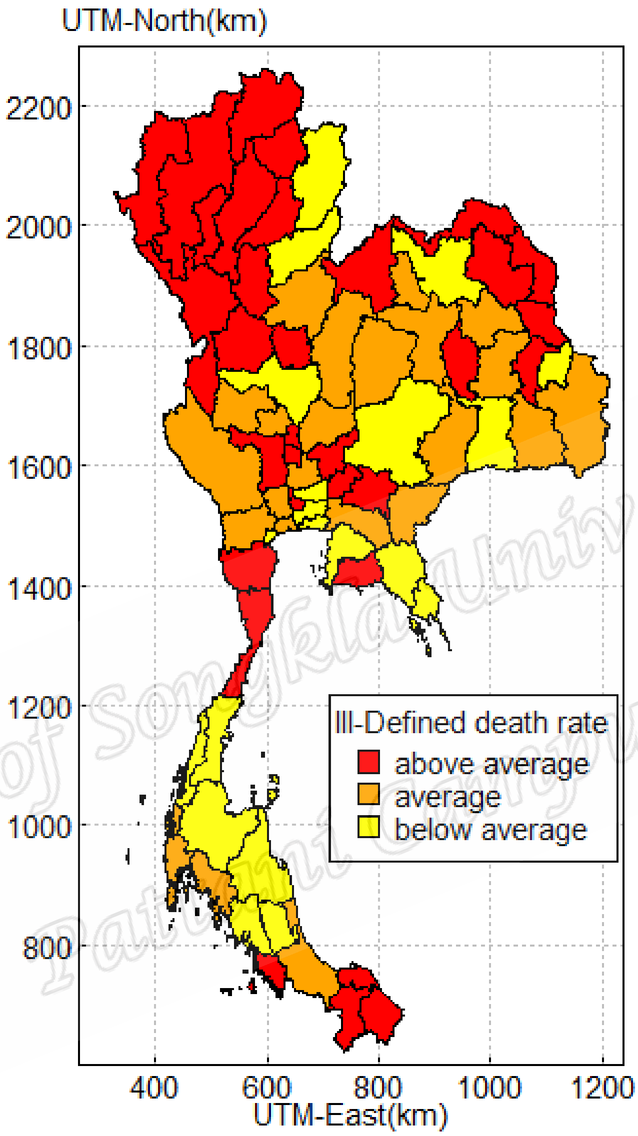


Figure 4.5 Schematic map of ill-defined death rate for each province

Figure 4.5 shows schematic map of ill-defined death rate per 100,000 population by province, based on confidence intervals. The map shows that 7 provinces in Central: Prachuap Khiri Khan, Phetchaburi, Prachin Buri, Samut Sakhon, Nonthaburi, Suphan Buri, Saraburi, Ang Thong, Sing Buri, Rayong and Nakhon Nayok, 11 provinces in the North: Chiang Mai,Chiang Rai, Mae Hong Son, Lampang, Lamphun, Phrae, Phayao, Kamphaeng Phet, Tak, Sukhothai, and Phichit, 7 provinces in the Northeast: Loei, Sakon Nakhon, Nakhon Phanom, Mukdahan, Maha Sarakham, Yasothon and Nong Khai and 4 provinces in the South: Satun, Pattani, Yala and Narathiwat had ill-defined death rates significantly higher than the overall mean.