Chapter 2

Methodology

2.1 Data sources and data management

2.1.1 Deaths

Gender-age-specific mortality for 926 districts over Thailand in years 1999-2001 were obtained from the vital registration database. This database is provided by the Bureau of Registration Administration, Ministry of the Interior and coded as cause-of-death using the tenth International Classification of Diseases (ICD-10) (World Health Organization, 1992) by the Bureau of Policy and Strategy, Ministry of Public Health.

2.1.2 Population

The population by gender, age group and district for Thailand was taken from the 2000 population and housing census conducted by the National Statistical Office (2002).

2.1.3 Data management

The two databases, death registration and population and housing census, used in this study have been collected from different format. The population from the 2000 population and housing census comprised the aggregated number by gender, age group, and district, while the registered deaths comprised individual records.

The age categories for both deaths and population are <1, 1-4, 5-9,…, 80-84 years. For simplicity, the open-ended age group 85 years and over was not included in our study.

The numbers of registered deaths in 1999-2001 were 362,607, 365,741, and 369,494 respectively. There were 0.54% with unknown age. These records were also excluded from our analysis.

The data analysis is based on place of residence rather than place of dying. Code of place of residence contains province, district, and sub-district, and we extracted only province and district code for our analysis. All death records with incomplete place of residence were ignored in our analysis.

Finally, the 1,086,142 (98.9%) completed records during 1999-2001 were used in this analysis. The aggregated deaths by gender, age group, and district were then calculated from this completed dataset and used as the number of all-cause deaths.

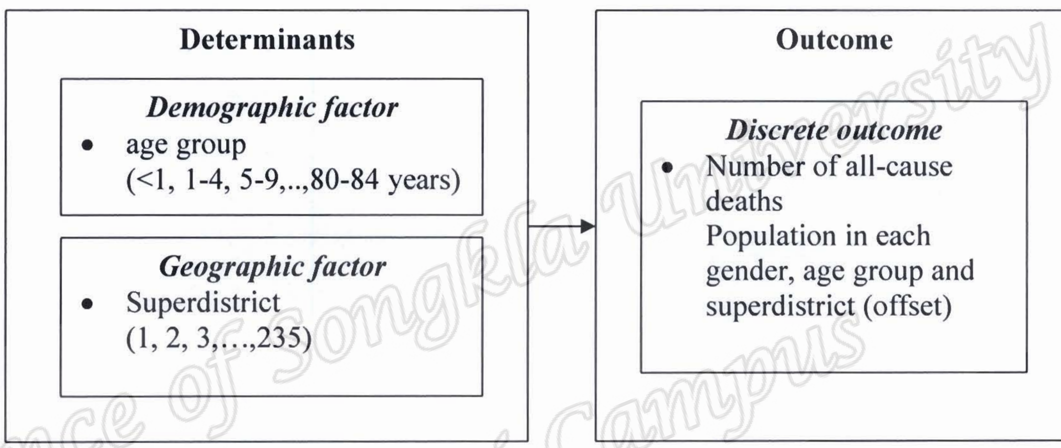
Proportional mortality from ill-defined cause was the focus of the second study. All records with cause-of-death coded in ICD10 as R00 – R99 were selected to summarize the number of ill-defined deaths by gender, age group, and district.

For simplicity, the population numbers by gender, age group, and district from the 2000 population and housing census were multiplied by 3 to use as the person-years at risk in our study.

2.2 Methods

Both descriptive statistics and statistical models were used to examine the mortality patterns in Thailand for 1999-2001. Appropriate graphics were developed for displaying the analysis results. Path diagrams for the two studies are shown in Figure 2.1. Since populations of districts in Thailand vary substantially, with the total resident under 85 population in 2000 ranging from a minimum of 2,088 (in King-Ko-Kut in Trat province) to a maximum of 451,447 (in Samut-Prakan City), the mortality data were analyzed in aggregated districts called "superdistricts", defined as regions comprising contiguous districts in the same province with the total population of at least 200,000 if possible. We thus obtained 235 superdistricts, with numbers varying from just one superdistrict in 14 provinces (Angthong, Singburi, Chainat, Nakhon-Nayok, Trat, Samut-Songkam, Amnat-Charoen, Mukdahan, Uthai-Thani, Phang-Nga, Phuket, Ranong, Krabi and Satun) to 24 in Bangkok province.

Study 1: Geographical variation of mortality in Thailand



Study 2: Geographical variation of ill-defined mortality in Thailand

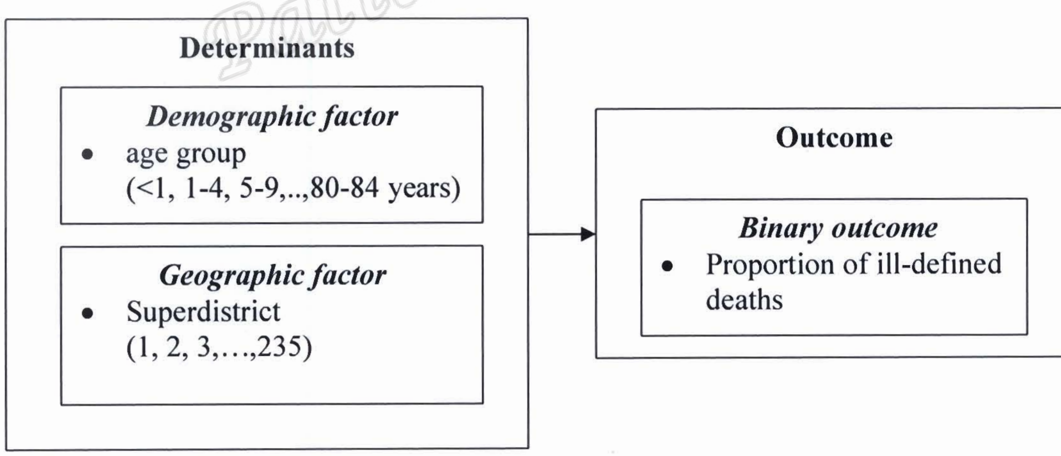


Figure 2.1: Path diagrams for variables used in our two studies

2.2.1 Crude death rates, age-specific death rates and cause-specific mortality

The *crude death rate* is the total number of deaths to residents in a specified geographic area (country, province, district, etc.) divided by the total population for the same geographic area (for a specified time period, usually a calendar year) and multiplied by 100,000.

$$\frac{Total\ resident\ deaths}{Total\ population} \times 100{,}000 \qquad (1)$$

The *age-specific death rate* is the total number of deaths to residents of a specified age or age group in a specified geographic area (country, province, district, etc.) divided by the population of the same age or age group in the same geographic area (for a specified time period, usually a calendar year) and multiplied by 100,000.

$$\frac{Total\ deaths\ in\ specific\ age\ group}{Total\ population\ in\ the\ same\ specific\ age\ group} \times 100{,}000 \qquad (2)$$

The *cause-specific death rate* is the number of deaths from a specific cause divided by the total population.

$$\frac{Total\ deaths\ in\ specific\ cause}{Total\ population} \times 100{,}000 \qquad (3)$$

The *proportional mortality* is used to compare the deaths from specific cause to total deaths (all causes).

$$\frac{Number\ of\ deaths\ due\ to\ cause\ X}{Total\ deaths\ (all\ causes)} \qquad (4)$$

2.2.2 Poisson regression

The number of deaths D_{ij} per 100,000 population E_{ij} in district i and age group j, may be modeled using the Poisson generalized linear model (Venables and Ripley, 2002). For an additive model, this distribution has mean

$$\lambda_{ij} = E_{ij} \exp(\alpha_i + \beta_j) \ , \quad (5)$$

where α_i are district specific parameters, and β_j are age-group parameters, one of which is redundant.

The contrast matrix used in the model was "sum contrasts" so that each parameter estimate has a standard error enabling comparison with the mean (Venables and Ripley, 2002; Tongkumchum and McNeil, 2009).

2.2.3 Logistic regression

Logistic regression is used for model fitting in which the outcome variable is binary. In this study, the adjusted proportions of ill-defined and unknown cause mortality by age group and superdistrict were estimated using a logistic regression model. The probability p_{ij} that a reported death for age group i and superdistrict j was ill-defined is thus modeled as

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \mu + \alpha_i + \beta_j \quad (6)$$

where μ is a constant and α_i and β_j are parameters associated with individual age groups (i = <1, 1-4, 5-9, .. ,80-84 years) and superdistrict (j = 1, 2, 3, .. ,235), respectively, that sum to 0. These coefficients were estimated from the data and the adequacy of the model was assessed using statistical methods described in Venables and Ripley (2002).

2.2.4 Graphical methods

Residuals plots

The plot of standardized residuals (y-axis) against normal quantiles (x-axis) is used to check the normality assumption. This plot should be a straight line if the method is corrected.

Plots of 95% Confidence Intervals

A 95% confidence intervals plot can be used to display the pattern of mortality rates for each factor of interest. These confidence intervals are produced from the standard errors of differences between the mortality rate and its overall mean, drawn as a vertical line.

Thematic maps

Thematic maps are essentially data maps – maps that show not only the location and shape of a feature, but also one or more values associated with the feature.

In our study, a thematic map was produced based on the model results. Gender-specific death rates were used for comparison of mortality between geographical areas. The mortality rates from all such areas in the sample were then classified into three groups, according to whether the confidence interval was (a) totally above the mean, (b) crossing the mean, or (c) totally below the mean. A thematic map was used to display this information using corresponding colours, (a) darkest shade, (b) intermediate shade, and (c) lightest shade. Statistically valid conclusions can be made using this map, that is, the mortality in each darkest-shaded superdistrict is greater than the average mortality, and the mortality in each lightest-shaded superdistrict is less than the average mortality.

Mortality Graphical Display GUI

The statistical model and graphical methods for analysing all-cause and cause-specific mortality rates as shown in the previous sections were developed using R commands. To facilitate to generating of analysis results of Thai mortality data with customized options, all R commands were then encapsulated for processing by the Mortality Graphical Display GUI (Figure 2.2).

The Graphical User Interface (GUI) enables the user to create side-by-side graphs comparing patterns of mortality for a specified sample (selected using an SQL-type database retrieval based on gender, region, place of death (inside or outside hospital) and diagnosis group) with respect to geographical variation and age pattern, and also to see diagnostic plots of residuals that show special features of the data (outliers) or a poor fit of the model. Two model choices are available: (a) Poisson, and (b) weighted transformed linear. The comparison choices are:

(1) gender (males or females);

(2) place of death (inside or outside hospital);

(3) cause of death (e.g., preventable cause or other);

(4) model (Poisson or weighted transformed linear).

Figure 2.2: Mortality Graphical User Interface