Chapter 1

Introduction

1.1 Overview of the thesis

The study presented in this dissertation focuses on using statistical methods for forecasting demographic time series counts of infectious disease (ID) mortality and for examining variation in length of hospital stay (LOS) with respect to principal diagnosis, demographic, and geographic and hospital size factors for patients dying in hospital. All subjects involved in this study were persons who died in 14 provinces in Southern Thailand over the period January 1, 2001 to December 31, 2004. Data were obtained from two official sources: vital registration and in-patient hospital databases.

This dissertation contains four chapters, including this introductory chapter.

Chapter 2 provides a description of methodology, geographic area of the study, data sources, and statistical methods used.

Chapter 3 illustrates and reports on statistical methods for modelling incidence of infectious diseases mortality and length of hospital stay, among patients dying in Southern Thailand.

Chapter 4 involves a general discussion of the results and statistical analysis methods used, together with limitations, strengths and applications evident from this study.

Also in this chapter, the overall conclusion includes implications of the study and recommended directions for further studies.

This introductory chapter proceeds with the background of the data, including the characteristics of the mortality data from each source. Subsequently, the background of infectious disease and length of hospital stay for those who died in hospital are revealed. There is a review of existing literature including statistical methods used and summary of the findings. This is followed by the rationale, research questions and objectives.

1.2 Background of the data

We consider two national sources of data used in this thesis: civil registration and hospital mortality databases.

1.2.1 Vital registration mortality data

Mortality data in Thailand has been computerized for over 10 years since the 1995 initiation of a programme for vital registration. The main source of mortality data is from the civil registration database kept by the Bureau of Registration, Department of Local Administration, Ministry of Interior, which was initiated in 1917 (Rukumnuaykit, 2006). The civil registration law in 1991 requires every death to be registered within 24 hours after occurrence. The head of the household or a close relative is responsible for reporting the death at a civil registry in the province where the death occurred. If the death occurred outside the house, the person who finds the body is responsible for reporting the death to the local registrar, usually the head of the village. Doctors or health personnel will issue the medical death certificate if the

death occurs in hospital (Ministry of Public Health 2004). Vital statistics and cause of death from civil registration are essential for public provisions such as making health policy and planning, since good quality of mortality data can result in better policy decisions and legislation (Setel et al, 2007). The qualities of the vital registration database in Thailand have been discussed by several studies (Tangcharoensathien et al, 2006; Rukumnuaykit, 2006; Prasartkul and Vapattanawong, 2006). The qualities of civil registration data are discussed as the following.

1.2.1.1 Completeness of death registration

The quality of death registration is a main problem in many countries especially in developing ones. At the end of 2003, there were 115 countries reporting deaths from their vital registration systems to World Health Organization (WHO). However, only 23 countries, mostly developed, have high quality death registration data (more than 90% complete, with ill-defined causes accounting for less than 10% of total deaths). Seventy five countries have no information available on causes of death for any year after 1990, with more than 90% from Africa (Mathers et al, 2005).

Thailand has also faced problems of unregistered deaths. Data from the Survey of Population Change (SPC) during 1974-1976 revealed that 41% of deaths were not registered officially (Rukumnuaykit, 2006). Incomplete registration of deaths reduced from 23-25% during 1985-1986 to 5% during 1995-1996 (Hill et al, 2006). After the official registration system was initiated in 1996, incompleteness of death registration has shown further gradual reduction (Mathers et al, 2005) A less ideal completeness

of death registration in Kanchanaburi province in Thailand during the five year period from 1999 to 2003 was found in a study comparing the deaths found in the SPC with those recorded in the registration system. It was estimated that 12.5 % of deaths were not registered (Prasartkul and Vapattanawong, 2006). The completeness of death registration in municipal areas was higher than in non-municipality areas. According to the SPC in 1985-1986 and 1995-1996, the highest death registration was in the North and the lowest was in the South (Rukumnuaykit, 2006). The most incomplete death registration was found to be for infants. Deficiencies in the death registration system may be due to incomplete coverage, late registration, missing data and errors in reporting or classifying the cause of death (Sibai, 2004).

1.2.1.2 The quality of cause of deaths

The underlying cause of death in most developed countries is certified by a medical practitioner, whereas in developing countries a substantial number of deaths are registered without involvement by a medical practitioner. Moreover, laws and public sanctions have influenced the underreporting of sensitive causes of death such as suicide and HIV/AIDS (Mathers at al, 2005).

The prevailing perception in society is that the most reliable mortality data are those with cause of death certified in a hospital by medical professionals. However, such information is available for less than 30% of the estimated 50.5 million deaths occurring in the world each year (Murray and Lopez, 1997). In addition, hospital statistics tend to under-estimate the prevalence of mortality in under-developed countries, such as Ethiopia (Reniers et al, 2005). In Thailand, the civil registration

database of the Ministry of Interior was officially recognized by the Ministry of Public Health (Ministry of Public Health, 2005). The cause of death is coded using the International Classification of Disease in its 10th revision (ICD10) by Bureau of Policy and Strategy, Ministry of Public Health. In Thailand, inaccurate recorded cause of death is still a major problem. Approximately 70% of deaths occur outside hospital (Chuprapawan et al, 2003) and a substantial proportion of these deaths are ill-defined with "senility (senescence)" (ICD10 code R54) recorded as the cause. Mathers et al (2005) reported that 49% of deaths reported in Thailand were of ill-defined cause and classified Thailand cause of death information as being of low quality. Figure 1 shows how Thailand compare to other countries in the quality of cause of death information.

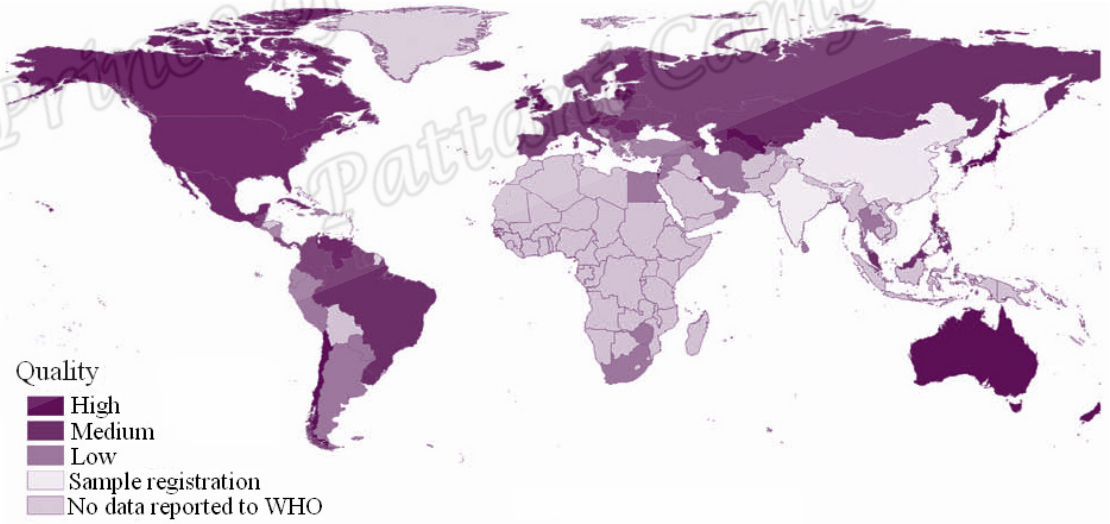

Figure 1.1 Quality of cause of death information from national civil registration systems 2003 (Mathers et al, 2005; WHO, 2007).

Tangcharoensathien et al (2006) conducted a study to evaluate the collection and flow of mortality and cause of death data in Thailand and found that 65% of deaths in 2004 occurred outside hospital. They suggested that errors in recoding deaths and errors in transferring information from death certificates to death registers were two reasons for incompleteness of death registration and inaccuracy of cause of death. In addition, the quality of cause of death records depends on the person who records the cause. If the cause is recorded by the head of village, the validity of the cause of death is questionable. Although the most reliable reporting of cause of death comes from physicians, there are still problems due to physician inexperience and the different practices for identifying cause of death among them. Porapakkham (1986) and Rukumnuaykit (2006) suggested that in statistical analysis a focus on death causes with the large numbers of deaths can be used to make the analyses more consistent and robust to systematic error. Furthermore, the quality of data should be investigated before analysis proceeds.

1.2.2 Inpatient mortality data

Inpatient data from various hospitals in Thailand, both government and private, are sent to the National Health Security Office (HNSO) for reimbursement of health expenditure. The causes of death for patients who die in hospital are recorded by physicians. Therefore mortality data is more valid and accurate than vital registration data. These data are only available from persons who are covered by the health insurance system. The Thai government established a universal coverage scheme called the “30 Baht” system in October 2001 (completed in April 2002) for the benefit of any Thai citizen who does not belong to the Social Security Scheme (SSS)

insurance scheme. An estimated 5% of the Thailand population in 2003 was not covered by any health insurance scheme, compared to 30% of the total population not covered before 2001 (International Labour Office, 2004). Data for inpatient mortality from NHSO should provide reliable results and can be used to validate or check the results from vital registration data.

1.3 Infectious diseases

Infectious diseases account for a quarter to a third of all global deaths. Human behaviour such as lifestyle choices, land-use patterns, increased trade and travel, and inappropriate use of antibiotic drugs such as from mutations in pathogens, are the causes of the spread of infectious diseases (National Intelligence Council, 2000). The leading infectious disease causes of death are from HIV/AIDS, tuberculosis, malaria, pneumonia, and diarrheal diseases, particularly in developing countries (World Health Organization, 2004). By 2020, HIV/AIDS is expected to cause more deaths than any other single infectious disease worldwide and may account for one-half or more of infectious disease deaths in the developing world alone (National Intelligence Council, 2000). The trend of infectious disease mortality has declined for both developed and developing countries due to rising living standards, use of antibiotics, vaccination and good medical technology (Armstrong et al, 1999; Arkwright, 2005; Pinner et al, 1996; Dore et al, 1998; Serraino et al, 2004; Bi et al, 2003; Gage, 1994; Wolleswinkel-van den Bosch et al, 2001). In developing countries, infectious diseases have re-emerged (Chow et al, 2003). The projection, carried out by WHO, of global deaths between 2002 and 2030 from selected diseases revealed large declines for the entire principal communicable diseases, except HIV/AIDS which has shown a

dramatically increasing trend. Global deaths from HIV/AIDS were projected to rise from 2.8 million in 2002 to 6.5 million in 2030 under a baseline scenario that assumes antiretroviral drug coverage reaches 80% by 2012 (Mathers and Loncar, 2006; WHO, 2007). Thus HIV/AIDS should be considered separately due to the difference trends of mortality between HIV/AIDS and other infectious diseases as shown in Figure 1.2.

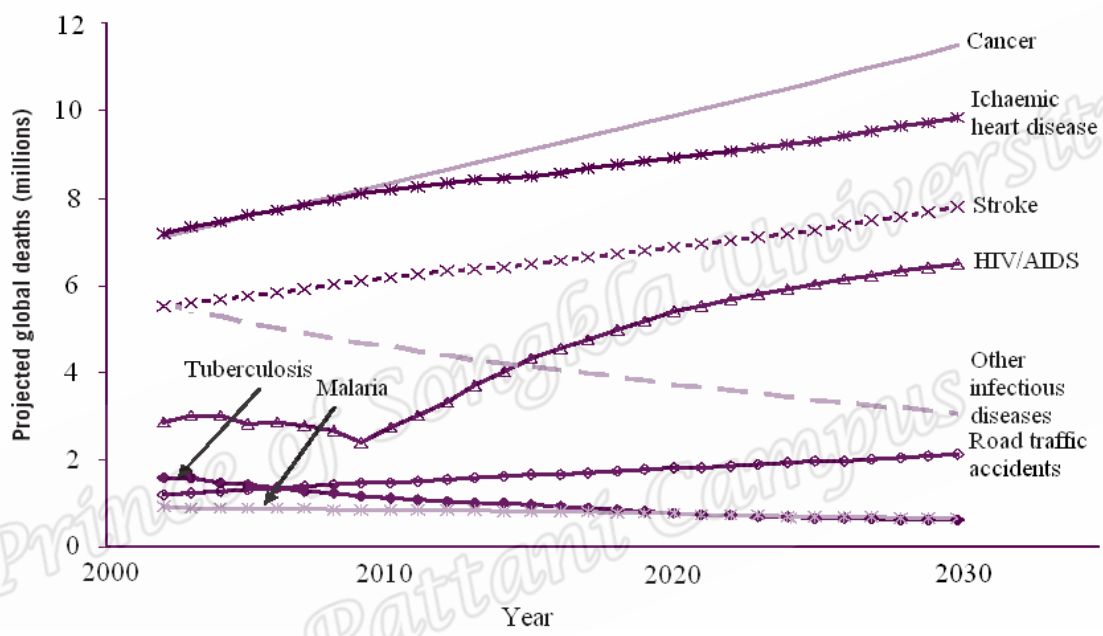

Figure 1.2 Projected global deaths for selected causes of death, 2002–2030 (Mathers and Loncar, 2006; WHO, 2007)

Infectious disease is the third leading cause of deaths in Thailand (Rumakom et al, 2001: Rukumnuaykit, 2006) in which HIV infection was the greatest component, especially in the 25-44 year age group (Bureau of Policy and Strategy, 2002).

1.4 Length of hospital stay (LOS)

Length of hospital stay (LOS) is a common parameter used to indicate health resource utilization, health care cost and severity of disease (Li, 1999; Wang et al, 2002; Lee et

al, 2003). LOS is a strong determinant of medical care cost and health expenditure. Therefore, reducing unnecessary LOS is one of the policies in many hospitals (Cannoodt and Knickman, 1984; Clarke and Rosen, 2001). Clarke and Rosen (2001) suggested that a low LOS in a hospital indicates greater general efficiency. The average length of hospital stay varies according to severity of diseases, patients' characteristics, geographical location, hospital characteristics, the effectiveness of treatment and physician experiences (National Center for Health Statistics, 1973). Many studies have considered LOS according to diagnosis-related groups (DRG) (for example: Horn et al, 1991). However each DRG is grouped according to the similarity of health care cost and LOS. Since LOS is taken into account in grouping DRG, LOS should not be used in analysis again as a separate determinant. The duration of staying in hospital depends on type and severity of disease and has a very wide range. A study conducted by Setrakian et al (1999) reported that average LOS in Canada was 14.4 days. There has been only limited study of LOS in Thailand and further investigation is needed.

1.5 Literature review

Several publications relevant to this study were reviewed including statistical methods used and their findings. Mortality from infectious diseases, length of hospital stay of patients dying in hospital, and the statistical methods used for these studies have been reviewed.

1.5.1 Statistical methods for mortality trends

Several statistical methods have been used for analyzing the spatial-temporal change of mortality trends. The Poisson model is commonly used for modelling the number of deaths in a specific population within a certain time. Poisson models in Bayesian and empirical Bayes analyses have been used by various authors (Besag and Molli'e, 1989; Maiti 1998, Dellaportas, 2000). Bayesian hierarchical modelling is suitable for combining neighbouring regions. Another approach for analyzing mortality trends was proposed by Arato et al (2006), based on a conditional binomial distribution for the number of age-dependent deaths, with zero mean Gaussian Markov Random Field models for incorporating spatial correlations between neighbouring regions and an intrinsic Gaussian model for including correlations between age-dependent mortality rates. This approach applies to general mortality in cohorts with a small population. A Generalized Linear Mixed Model (GLMM) for the number of deaths was used by Chan and Tong (2006) for estimating the death rate of an emerging infectious disease with censored aggregate data.

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. Condon (2006) set out a framework for modeling health outcomes over area, age, and time dimensions that takes account of spatial correlation, interactions between dimensions, and cohort as well as age effects.

Most mortality forecasting studies concentrate on life expectancy for the purpose of social insurance and pension payment. For example, Booth et al (2006) and Wang

(2007) applied the Lee-Carter model for mortality forecasting. Babel et al (2006) applied the Kannisto model for forecasting German life expectancy. A study in Kenya using data from the 1989 population census examined the regional differentials in under-five mortality in Kenya using negative binomial regression (Rutaremwa, 2000). Autoregressive terms were not considered in that study.

Choi et al (2003) suggested that an adequate analysis should depend on the scale at which the epidemiological field is considered rather than being limited by the scale at which the data are available. In our study, a large number of deaths are required for adequate analysis of mortality at the regional level. Negative binomial generalized linear models are used for handling over-dispersion of the number of deaths. This method is general and powerful and is available in the widely used R system, freely downloadable from the Internet.

1.5.2 Infectious disease mortality trends

There are several studies of the mortality trends from infectious disease. A study in England showed a decrease trend of past mortality from infectious diseases (Arkwright 2005). This result agrees with a study of infectious disease mortality trends in United States conducted by Armstrong et al (1999). The study revealed that death from infectious diseases declined during the first 8 decades of the 20th century. The last 2 decades saw an increased trend in the 1980s and early 1990s in persons aged 25 years and older and was mainly due to the emergence of the acquired immunodeficiency syndrome (AIDS) in 25 to 64 year olds. A study in Italy conducted between 1969 and 1999 described infectious disease mortality excluding HIV/AIDS, with particular emphasis on sex, age, and geographic differences and revealed a very

strong downward trend, with a 6-fold decline. This trend leveled off in 1995-1999, mainly due to increasing deaths due to septicaemia, heart infections and hepatitis (Angeletti, 2004). Dore et al (1998) conducted a study on the trend of infectious disease mortality in Australia. The results showed that mortality from infectious disease in Australia declined overall between 1979 and 1994. However, there were significant increases in mortality from septicaemia and HIV/AIDS which is expected to become a leading cause of infectious disease mortality, particularly among adult males.

Curns et al (2005) conducted a study of infectious disease hospitalizations among older adults in the United States from 1990 through 2002. The result of this study revealed that hospital death rate due to infectious disease decreased during that period. The overall hospital death rates during the last 3 years of their study were similar across genders. They revealed that septicaemia accounted for 12% of hospitalization and accounted for approximately one third of infectious disease related hospital deaths from 2000 through 2002. A study in Israel using death certificate data for 1979-1992 found that there was a 30% increase in the mortality rate from septicaemia (Shohat et al, 1999).

1.5.3 Statistical methods for length of hospital stay

Generally, LOS has a highly positive skewed distribution (Liu et al, 2001; Lee et al, 2003) and some patients stay in hospital for a very long time. Therefore it is important to develop appropriate methods for the statistical analysis of LOS data. Using linear regression to predict LOS is likely to seriously violate the assumptions of the model (Li, 1999). Various methods have been suggested to handle outliers. For caesarean

delivery LOS Lee et al (2003) used median regression to analyze LOS without transforming and trimming the data. However, this method is used for analyzing LOS only for the one group of diagnosis with slight skew of LOS. Marazzi et al (1998) examined the adequacy of models based on lognormal, Weibull and gamma distributions and found that the lognormal model was most appropriate but cases with LOS less than one day were omitted from their analysis because of computational problems when the LOS is zero. Harrocks (1999) had developed parametric models for LOS that accommodate for the LOS skewness and allow for multiple discharged statuses using data extracted from medical charts. This method is equivalent to competing risk analysis. Clark and Ryan (2002) developed a piecewise exponential model for predicting length of stay, with discharge alive or dead, for seriously ill or injured patients. Applying this method with a larger sample size will provide more precise results. They suggested that interactions and data transformations may also provide more insight into the factor accountable for longer hospital stay.

1.5.4 Variation of length of hospital stay (LOS)

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 (Health Technology Case Study 24, 1983; Xiao, 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 (see, for example, Cabre et al, 2004). 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.

1.6 Rationale

This section describes the reasons for conducting this study, identifying issues, knowledge gaps, and expected usefulness for future research and health policy decisions.

Technologies of computerization have improved substantially. Computers can now store huge mortality datasets, however utilization of these available data is low, especially in developing countries. Accuracy and completeness of mortality data are important for sound health planning and policy making. More accurate and comprehensive reporting on the death certificate will increase the analysis utility and crucial indicators of health, including demographic characteristics and causes of

death, can be tracked. Information from mortality data is invaluable for setting budget priorities, ranking the burden of diseases and establishing suitable research topics (King and Lu, 2007).

Even though mortality data in Thailand has problems with quality, they can provide useful information. Several studies have used mortality data from the vital registration system to investigate the burden of disease and mortality trends, such as one by Rumakom et al (2002), which estimated levels and patterns of tuberculosis and pneumonia mortality and found an increasing trend in these diseases. Rukumnuakit (2006) investigated overall mortality trends in Thailand and found significant improvement in survival probability.

Although causes of death are more accurately recorded in hospitals, there is still substantial room for improvement. For example, for persons living in southern Thailand over the six years between1999-2004, of the 6,885 deaths from infectious diseases, excluding HIV/AIDS certified in hospitals, 61.2% had principal diagnosis recorded as “septicaemia, unspecified” (ICD10: A41.9), which says little about the actual cause.

HIV mortality was the greatest infectious disease component among those aged 25-44 years (Bureau of Policy and Strategy, 2002). Therefore, looking at HIV/AIDS separately from other infectious diseases should be useful for investigating pattern of disease trends.

Not only do mortality trends provide useful information for public health planning but it is also important to understand the factors that affect the distribution of the length of stay for patients who die in hospital. Such knowledge will assist policy planning and

enable health funds to better allocate resources in the community. Results from in-patient data can be used to confirm and support the finding from vital registration data. In addition, mortality from several diseases in hospital can be an indicator of the overall health situation in the population and a major determinant of the cost of medical care. Information on hospital bed utilization for different diseases in Thailand is also inadequate. Accurate knowledge could be used to rearrange the facilities of health services according to the severity and burden of each disease.

1.7 Research questions

This thesis is focussed on forecasting incidence of infectious diseases mortality and investigating length of stay for patients dying in hospital. The research questions are the following.

1. What is the trend and pattern of infectious disease mortality in Southern Thailand?
2. What statistical methods appropriate for forecasting infectious mortality rate in Southern Thailand?
3. What are the demographic and spatial-temporal factors associated with infectious disease mortality in Southern Thailand?
4. What diseases cause longer stays among patients dying in hospital?
5. What is the pattern of the variation in length of stay (LOS) with respect to principal diagnosis and demographic, geographic and hospital size factors for patients who died in hospital in Southern Thailand?
6. What are the factors associated with length of hospital stay?

7. What statistical methods appropriate for investigating the variation of length of hospital stay?

1.8 Objectives

1. To investigate and forecast regional and temporal patterns of mortality reported as infectious disease (including HIV/AIDS) in the 14 provinces of southern Thailand over the period 1999-2004.
2. To develop statistical model for forecasting of infectious mortality rate in Southern Thailand.
3. To identify the association between regional and temporal patterns of mortality reported as infectious disease and mortality rate in Southern Thailand by using the appropriate statistical methodology for obtaining the true underlying spatial pattern.
4. To investigate inpatient length of stay (LOS) for patients dying from all diseases.
5. To develop statistical model for investigating the variation of length of hospital stay.
6. To identify the factors associated with length of hospital stay for inpatients who died from all diseases.