Chapter 1

Introduction

1.1 Overview of the thesis

"*Good care prevents accidents*"

(Albanian Proverb)

This Ph.D. thesis focuses on using statistical methods to create an appropriate model for estimating the actual number of transport accident deaths in the national vital registration (VR) data by gender-age, province, VR cause-location group over the years 1996 to 2009. Afterwards, this thesis focuses on using statistical methods for exploring trend directions and patterns of transport accident mortality rates in Thailand during the years 2004 to 2009. This thesis proceeds in four chapters as follows:

Chapter 1 is an introduction to the thesis with rationale, research questions, objectives and literature review.

Chapter 2 describes the methodology including study design, data sources, data management, statistical analysis and conceptual framework.

Chapter 3 describes and illustrates results of statistical methods for estimating transport accident deaths and external causes of deaths based on verbal autopsy (VA) study and for exploring patterns of transport accident mortality rates in Thailand.

Chapter 4 is an overall conclusion and discussion of the findings and statistical analysis methods used, together with limitations, recommendations for policies to address transport accident death problems in Thailand and suggestions for further study.

1.2 Rationale

Transport accident deaths have long been recognized as a major public health problem affecting much of the world. This is a problem which is preventable and controllable (Anjuman *et al*., 2007, World Health Organization, 2009a; World Health Organization, 2011a). However, transport accident deaths are projected to increase from the ninth leading cause of death globally in 2004 to be the fifth leading cause of death in 2030 (World Health Organization, 2009b). Moreover, these accidents have a direct impact on human physical and mental health, quality of life and property (World Health Organization, 2006). In addition, transport accidents cause a burden to health care systems globally, especially in low-income and middle-income countries (Anjuman *et al*., 2007, World Health Organization, 2009a).

Thailand is an upper middle income country in the South-East Asian region (Lopez *et al*., 2006; World Health Organization, 2008). A survey of 11 countries in the South-East Asia region in 2007 and 2011 reported that Thailand had the highest estimated road traffic accident mortality with the rate of 25.4 per 100,000 population in 2007 (World Health Organization, 2009b) and the highest estimated road traffic accident mortality with the rate of 38.1 per 100,000 population in 2011 (World Health Organization, 2013). Although, as previously mentioned, various concerned agencies have placed an emphasis on addressing these problems since 2003 (Tanaboriboon and Satiennam, 2005), the rate of road traffic accident mortality has also risen.

It is well known that Transport accident deaths vary with demographic and regional, and remains a challenging issue to all concerned agencies to address this problem. However, death data from VR system in Thailand are limited for their public utility (Porapakkham *et al*., 2010; Bundhamcharoen *et al*., 2011). Good quality data for transport accident deaths remain essential to aid health planners with appropriate planning, problem solving and decision making. Death certificates from VR system in Thailand are potentially a good source for death data from injuries (Aekplakorn, 2003). However, these data are of limited use as approximately 40 to 49% of deaths have the cause of death reported as “ill-defined”, unknown or unspecified causes (Mathers *et al*., 2005; Tangcharoensathien *et al*., 2006; Rukumnuaykit, 2006; Polprasert *et al*., 2010). This is important because accurate measurement of mortality risk is the basis for effective disease control and prevention.

To address the misclassification of the causes of deaths, the VA study was conducted in 2005 by the SPICE project (The Setting Priorities using Information on Cost-Effectiveness analysis) to verify causes of deaths reported by VR in Thailand (Pattaraarchachai *et al*., 2010; Polprasert *et al*., 2010; Porapakkham *et al*., 2010; Rao *et al*., 2010).

This study applied statistical methods to the available VR data to produce more reliable data on cause of death. The actual number of transport accident deaths was used for investigating trend patterns of transport accident mortality rates in Thailand.

Research questions

This thesis focuses on epidemiological patterns of transport accident mortality rate in each province in Thailand. These search questions are the following;

1. How well does the method predict transport accident deaths, are the results the same as the number of transport accident deaths in the Thai 2005 VA study?
2. What is the quality of transport accident death reporting in the VR system in Thailand?
3. How well does the method predict external causes of death (i.e. other accident and suicide), and are the results the same as the number of external causes of death in the Thai 2005 VA study?
4. What is the quality of external causes of death reporting in the VR system in Thailand?
5. How many patterns of transport accident death are there in Thailand?

1.3 Objectives

1. To create an appropriate model for transport accident death based on the Thai 2005 VA study in 2005.
2. To estimate the actual number of transport accident deaths from VR data in Thailand years 1996 to 2009.
3. To create an appropriate model for estimating external causes of death (i.e. other accident and suicide) based on the Thai 2005 VA study.
4. To estimate the corresponding number of external causes of death (i.e. transport accident, other accident and suicide) from VR data in Thailand 1996 to 2009.
5. To explore patterns of transport accident mortality rate in Thailand between 2004 and 2009.

1.4 Literature Review

This literature review covered four major topics relevant to this study including mortality data, verbal autopsy, transport accident mortality and statistical methods used for this study.

1.4.1 Mortality data from the national Vital Registration (VR) system

Two major government agencies are involved in civil registration in Thailand, the Ministry of the Interior (MOI) and the Ministry of Public Health (MOPH) (Kijsanayotin *et al*., 2013). The VR system, part of civil registration, is the most important source of mortality data in Thailand, and has been enforced throughout Thailand since 1917 (Rukumnuaykit, 2006; Vapattanawong and Prasartkul, 2011). Fetal deaths were included for the first time in the VR system in 1936 (Arnold and Kfiner, 1980; Kijsanayotin *et al*., 2013). The civil registration act of 1991 has required deaths must be registered within 24 hours at the place of occurrence. For deaths occurring in a health institution, health personnel (physician, nurse or medical coder) will issue a medical death certificate, while for deaths occurring outside a health institution, the head of village will issue a notification report based on information from the head of the household or family members. This notification is then passed on to local registrars. For deaths occurring outside the house, the person who found the body or the head of the village reports the death to local registrars (Prasartkun and Vapattanawong, 2006; Rukumnuaykit, 2006; Kijsanayotin *et al*., 2013) as shown in Table 1.1.

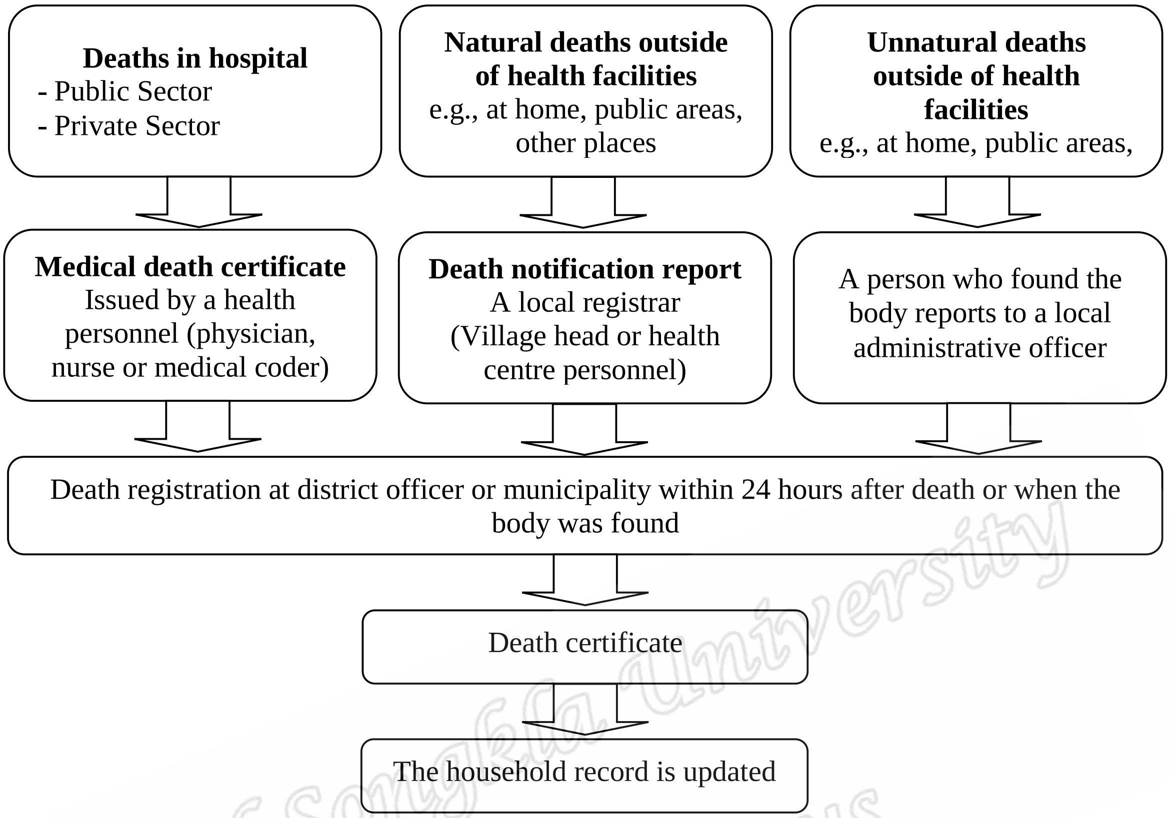
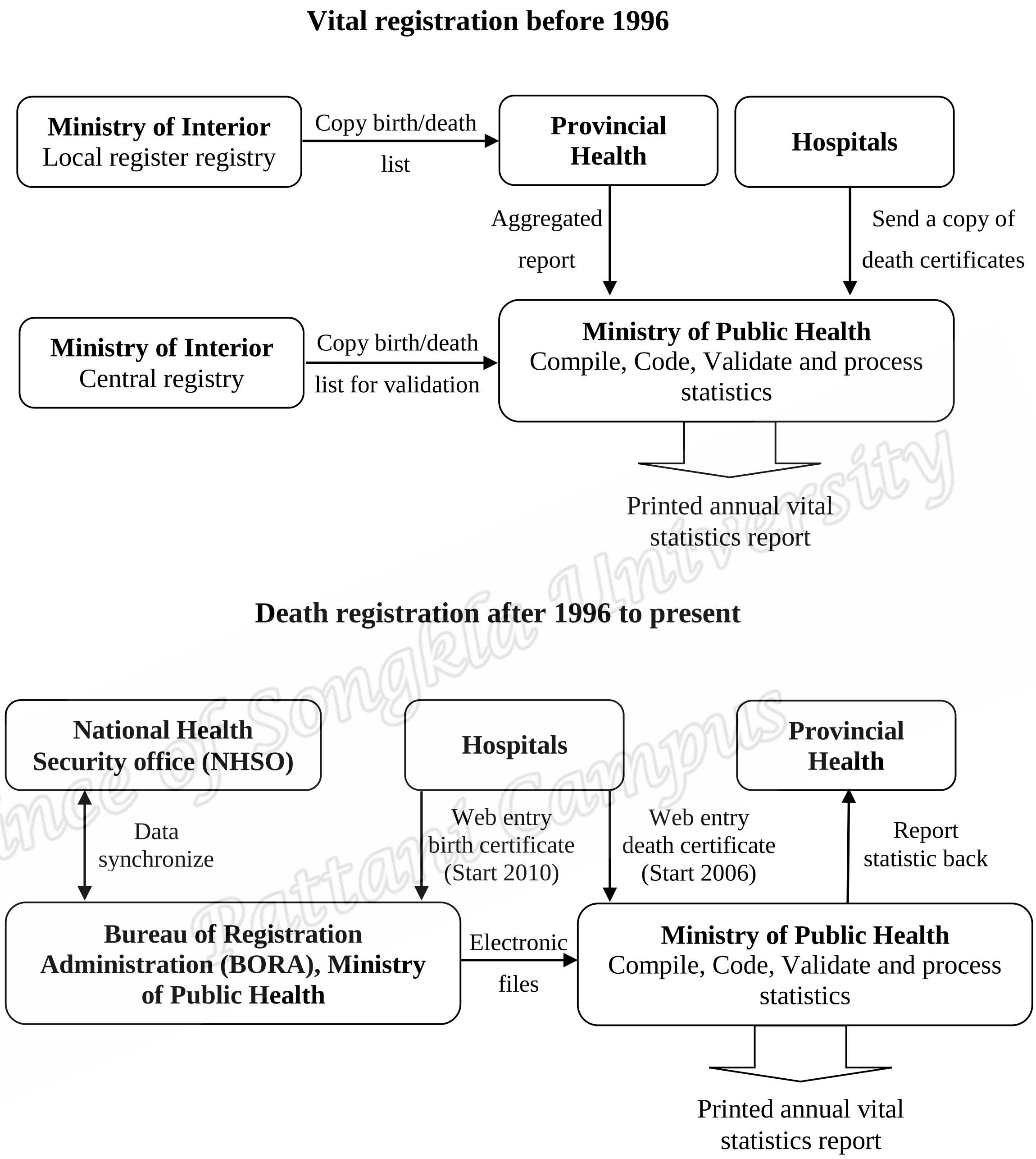


Figure 1.1 Process of death registration in Thailand

However, mortality data before 1996 was severely impacted by under-registration and poor causes of deaths data. In 1996, the VR system was improved and implemented when the MOI began providing electronic files of death data and the MOPH signed an agreement regarding utilization of the VR system (Rukumnuaykit, 2006; Kijsanayotin *et al*., 2013) as shown in Figure 1.2.



Modified from: Thai Health Information Standards Development Center, Health Systems Research Institute

Figure 1.2 Vital registrations before 1996 and after 1996

Although the system of death registration is improving in Thailand two major problems of data quality remain. There are the incompleteness of death registration and inaccuracy of cause of death recording with a high proportion of ill-defined or non-specified cause (Tangcharoensathien *et al*., 2006). These problems severely limits the public utility of VR data.

Incompleteness of death registration

The incompleteness of death registrations involved two issues: the incomplete death registration and the incomplete data regarding the deceased persons, such as age, gender, place of residence and cause of death (Prasartkul and Vapattanawong, 2006). Under-reporting of death registrations has influence on the estimation of actual mortality.

The Survey of Population Change (SPC) has been conducted every 10 years since 1964 to estimate the completeness of the VR system. The first three SPCs (i.e. 1964-1966, 1974-1976, 1984-1986) used dual records procedure (i.e. identifying deaths recorded in both sources of data or recorded in one source). The next two SPCs (1995 to 1996 and 2005 to 2006) used direct questions in the survey to ascertain whether vital events were registered or not. The surveys found that the completeness of death registration was 60.0% in the Thai population during the 1960s and 1970s, rose to 76.5% in 1985 to 1986, and 95% in 1995 to 1996 and 2005 to 2006) (Prasartkul and Vapattanawong, 2006; Vapattanawong and Prasartkul, 2011). The study conducted by Vapattanawong and Prasarkul (2011) used cross-matching data from SPC and VR for estimating under-registration of death during 2005 to 2006. They found that under-registration of deaths was 9.0% for males and 8.4% for females. This was 1.7 times in males and 1.9 times in females, higher than was found in the SPC. The Kanchanaburi project in 1999 to 2003 was conducted to compare a survey of deaths with those recorded in the VR system. The result found that 12.5% were not registered. Most of those were in children less than five year olds (20.8%) (Prasartkul and Vapattanawong, 2006).

Other errors in the death registration system in Thailand from 1996 to 2009 include, from 0.01 to 6.82 of unknown province and from 0.05 to 4.78 of unknown age. The highest proportion of deaths in an unknown province occurred in 2004, while the highest of unknown age occurred in 1996 (Chutinantakul *et al*., 2014a).

Quality of cause of deaths

Reliable cause of death data from the VR system is essential to understanding the relative importance of various diseases and injuries and is thus of high utility for health policy makers. In 2003 a study of death registration data from 115 countries found only 23 countries had more than 90% complete data and less than 10% with cause of death as ill-defined cause or unknown cause. Thailand was classified as one of the 28 countries with low quality of data, with registration data consisting 49.0% with ill-defined or unspecified causes of death (Mathers *et al.*, 2005).

There are two causes of inaccuracy of cause of death recording in death registration, from deaths occurring in hospital and outside hospital. In 2004, 65% of death occurred outside hospital and 35% of death occurred in hospital (Tangcharoensathien *et al*., 2006). For deaths occurring in hospital and deaths of unnatural causes, the cause of death is specified by a physician or health personnel who may have had little contact with the deceased person. While for deaths of natural causes occurring at home outside a municipal area, the cause of death is specified by a village headman who likely lacks expertise in determining cause of death (Prasartkul and Vapattanawong, 2006; Rukumnuaykit, 2006; Tangcharoensathien *et al*., 2006).

Therefore, Thai cause of death data are not of value for disease prevention and control in health information systems.

The limited utilities of VR data derive from the large proportion of vague causes of deaths. However, several previous studies have attempted to analyze geographical variations in mortality in Thailand despite this limitation. These include studies of all-cause mortality by Faramnuayphol *et al*. (2008); Odton *et al*. (2010) and a study of transport accident mortality by Sriwattanapongse *et al*. (2013). Some of these studies manipulated the data for analysis, For example, Angkurawaranon *et al*. (2013) assessed the association between urbanization and three non-communicable diseases (cardiovascular, cerebrovascular and malignant neoplasms) which were among the top ten leading known causes of death using 2009 VR data. This study had less problems with misclassification because there was no evidence of misclassification of these causes of death in hospital. Prameprart *et al*. (2013) modeled unintentional drowning mortality rates in Thailand using VR data for the years 2000 to 2009. They regrouped missing age (0.01%) into aged 80 years and over and omitted unknown province before doing further analysis.

1.4.2 Verbal autopsy (VA)

Determining causes of death has been significant to public health for problem solving. VA is a technique for determining cause of death in cases which are not medically certified in the vital registration system. In developed countries, VA is a routine process of medical diagnosis of each death unlike in developing countries. The VA method has been widely used for correcting causes of death in several countries, including China, India, Kenya, Tanzania, Thailand, Zambia (Gajalakshmi and Peto, 2004; Yang *et al*., 2005; Setel *et al*., 2006; Rao *et al*., 2010; Mudenda *et al*., 2011; Ndila *et al*., 2014;).

Two major categories are used for validation and validity of VA including physician-certified verbal autopsy (PCVA) and computerized coding of verbal autopsy (CCVA) (Chandramohan, 2011). They found VA method can be used for correcting causes of death from VR data. Several studies have linked VA and VR to improved causes of death assignation (Setel *et al*., 2005; Setel *et al*., 2006; Porapakkham *et al*., 2010; Mudenda *et al*., 2011; França *et al*., 2011).

In Thailand in 1999, a VA study was initially used for verifying causes of death reported by VR in 15 provinces outside Bangkok which provided data for analysis of the burden of disease (BOD) (Choprapawon *et al*., 2005). However, the reliability and validity of causes of death reported in VR have remained a priority and also a major problem. Thus, in 2005 an improved VA study for verifying causes of deaths from VR in Thailand was begun by the SPICE project (Setting Priorities using Information on Cost-Effectiveness analysis). The SPICE project verified cause of death from the national VR data in Thailand in 2005 using a multistage stratified cluster sampling technique. The data were based on a sample of 3,316 hospital deaths and 6,328 home deaths from 28 districts in nine provinces (i.e. Bangkok, Chiang Rai, Phayao, Loei, Ubon Ratchathani, Nakhon Nayok, Suphanburi, Chumphon and Songkhla). For nine provinces, the samples were selected from 28 of the 926 districts in Thailand. Districts were selected by two-stage stratification where Bangkok and pairs of provinces from the four regions (i.e. North, Northeast, Central and South) were divided into two strata at the 50th percentile and similar districts were selected in each province for representation of urban or rural communities. Finally, death certificates to be assessed were randomly selected from the 28 districts using the probability-proportional-to-size (PPS) method as shown in Figure 1.3. More detail of the Thai 2005 VA study is available from previous publications (Byass, 2010; Pattaraarchachai *et al*., 2010; Polprasert *et al*., 2010; Rao *et al*., 2010).

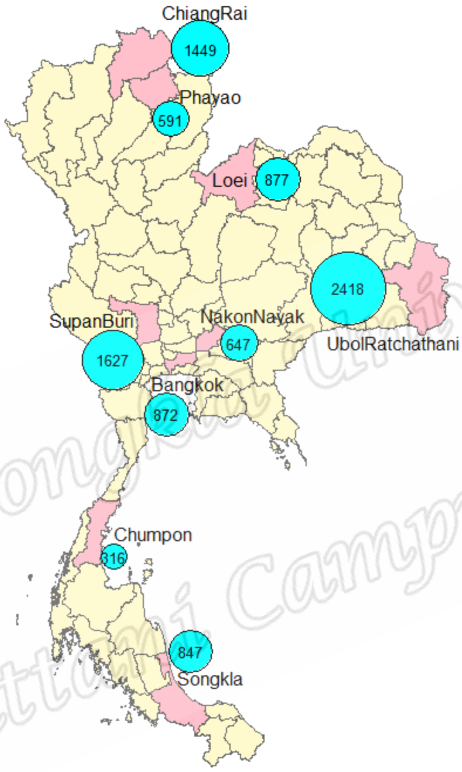


Figure 1.3 Map of the number of samples in the nine provinces of the Thai 2005 VA study

The 2005 VA study found that almost 30% of deaths in hospital were assigned causes of death as ill-defined or septicemia (Pattaraarchachai *et al*., 2010). More than 53% of deaths outside hospital were assigned causes of death as ill-defined (Polprasert *et al*., 2010). The study reduced the percentage of ill-defined causes of deaths from 33% to 4% in males and from 45% to 6% in females (Rao *et al*., 2010). Other findings were that transport accident was the second leading cause of death in males (8.1%) (Rao *et* *al*., 2010*)*. After the reassignment of cause of death, transport accident deaths increased from 2.8% to 6.4% for in hospital deaths (Pattaraarchachai *et al*., 2010) and increased from 2.7% to 5.2% for outside hospital deaths (Polprasert *et al*., 2010). These studies have limitations in the method of data collection and the VA method (e.g. the responses to the VA interview, the diagnostic guidelines by physician reviewers) to derive final mortality.

However, these showed that VA study can verified the actual transport accident deaths from VR data.

1.4.3 Transport accident mortality

Transport accidents are classified using the International Statistical Classification of Disease and Related Health Problem 10th Revision (ICD-10) codes V01-V99 relating to land transport accidents (V01-V89), water transport accidents (V90-94), air and space transport accidents (V95-97), and unspecified transport accidents (V98-99) (World Health Organization, 2004). Of those, road traffic accidents cause the majority of deaths.

Transport accidents are a major global health problems, notably in low-income and middle-income countries (Nantulya and Reich, 2003; World Health Organization, 2007; Naci *et al*., 2009; World Health Organization, 2009a; World Health Organization, 2013). It is well known that this problem varies with demographic and regional factors, and have a direct impact on human physical and mental health, quality of life and property (World Health Organization, 2006). For example, an estimate of global road fatalities in 1999 found that between 750,000 and 880,000 people died from road crashes, almost half of all estimated deaths occurred in the

Asia-Pacific region. In addition, forecasts of the number of deaths between 2010 and 2020 suggest that over the next ten to twenty years the number of people dying annually from road crashes may rise from 1 million to 1.3 million (Jacobs *et al*., 2000). The Global Burden of Disease (GBD) in 2004 predicted that road traffic accidents rise from the 9th leading cause of death in 2004 to the 5th in 2030 as shown in Figure 1.4 (World Health Organization, 2010).

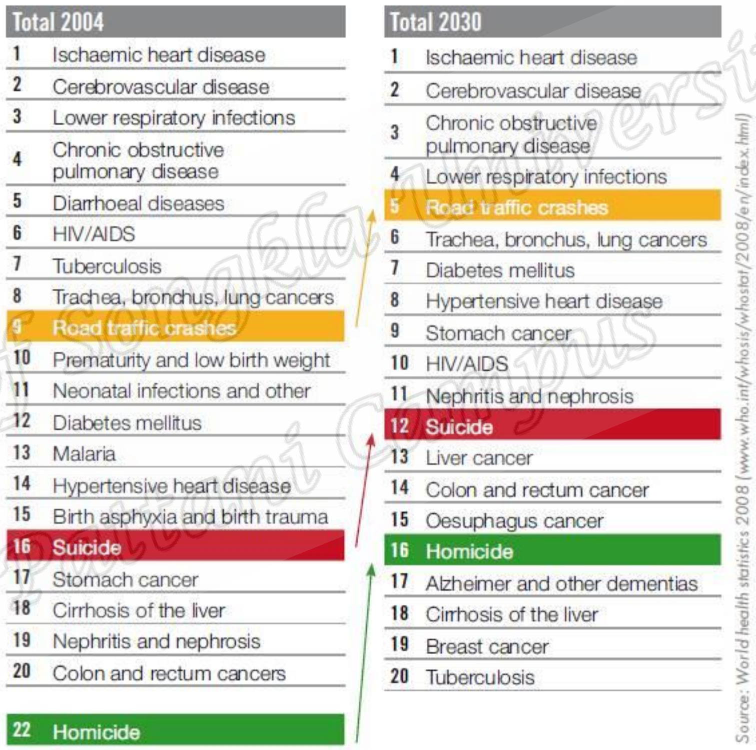


Figure 1.4 Compared leading causes of death between 2004 and 2030

The number of road traffic accidents increased in several countries. Kim *et al*. (2012) using National Statistical Office data from 1991 to 2006 in Korea found that the transport accident mortality rate increased from 47.7 per 100,000 persons in males and 15.0 per 100,000 persons in females in 1991, to 56.8 per 100,000 persons in males and 20.6 per 100,000 persons in females in 1996. After that, they decreased from 36.9 per 100,000 persons in males and 14.2 per 100,000 persons in females in 2000, to 23.1 per 100,000 persons in males and 8.7 per 100,000 persons in females in 2006. From 1994 to 1998, road traffic accidents were a leading cause of death and 61.2% of road traffic fatalities and 52.3% of road traffic injuries occurred on roads in rural areas in Ghana (Afukaar *et al*., 2003). The WHO survey of 178 countries in 2008 reported that low- and middle-income countries have road traffic mortality rates of 21.5 and 19.5 per 100,000 population respectively, while high-income countries have road traffic mortality rates of 10.3 per 100,000 population (World Health Organization, 2009a). Ninety-one percent of road accident mortalities occurred in low-income and middle-income countries, especially in the South-East Asian region, from a rapidly increasing number of road accidents (World Health Organization, 2009a).

Road traffic accidents affect people of all ages, both males and females. Peden *et al*. (2004) presented an overview of risk factors and impact of road traffic accidents based on a collaboration of the WHO and the World Bank and found that in 2002, more than 50% of global traffic accident mortality affected those aged between 15 and 44 years and road traffic mortality rates in males are three times higher than in females. Whereas, a report in South Africa in 2001 to 2006 found the highest road traffic accident mortality rate in those aged 35 to 49 years and males had a higher road traffic accident mortality rate than females by 2.5 times (Lehohla, 2009).

Thailand is classified as an upper middle country in the South-East Asian region. A survey covering 10 of the 11 countries of the WHO South-East Asia region in 2007 reported that Thailand had the highest mortality rate at 25.4 per 100,000 population (World Health Organization, 2009b), and over 80% of those killed on roads in

Thailand in 2007 were vulnerable road users (pedestrians, motorcyclists and cyclists) (World Health Organization, 2009a). In 2008, the age-standardized death rate from road traffic accidents in Thailand was 42.9 per 100,000 population (World Health Organization, 2011b).

The seriousness of the transport accident mortality threat is shown in several databases. Data from the Injury Surveillance System of 28 provincial hospitals in the years 1998 to 2007 showed that most of injuries were from transport accidents, and rates in males were three times higher than females. The peak of transport accidents was in those aged between 15 and 24 years (Siripanich *et al*., 2009). Based on data from the Royal Thai Police in 1983 to 2002 the pattern of road traffic accident had varied substantially in this period. Road traffic mortality rates increased from 6.1 per 100,000 population in 1983 and peaked at 28.2 per 100,000 population in 1995 before steadily decreasing to 20.9 per 100,000 population in 2002 (Taneerananon *et al*., 2003; Tanaboriboon and Satiennam, 2005). In addition, a study on death certificate reports in the national VR database years 2000 to 2009 found that males had an increasing rate of transportation accident mortality higher than females, and the highest mortality rate was in males aged 20-29 years. However, this study forecasted based on the VR databases which were not improved for the actual causes of death (Sriwattanapongse *et al*., 2013).

Spatial analyses of transport accident mortalities have been useful for identifying the risk areas and used to compare the effectiveness of prevention measures for reducing the death rate. The annual report of the Thai Roads Foundation and Thailand Accident Research Center in 2011 found that Bangkok was a lower risk area for death from transport accidents than other provinces. A similar result was also reported by Choiejit and Teungfung, (2005) and Yiengprugsawan *et al*. (2014).

1.4.4 Statistical method for estimating transport accident deaths

There are many statistical methods for analyzing mortality data depending on the purpose of the studies and the data distribution such as the capture-recapture method. Many studies used the capture-recapture method in resource limited settings (Hest *et al*., 2011), for estimating incomplete data (Khazaei *et al*., 2012), and for matching the number of both sources (Aptel *et al*., 1999; Razum, *et al*., 2000; Tercero and Andersson, 2004; Van *et al*., 2006; Porapakkham *et al*., 2010). In addition to other previous studies (Wittes *et al*., 1974; Robles *et al*., 1988), the capture-recapture method is applied for estimating the size of the undercount in a population or estimating the number of duplicate records on a database (Chao *et al*., 2001).

The logistic regression model is widely used in several studies of biomedicine, epidemiology and public health. It is always used together with a receiver operating characteristic (ROC) as a diagnostic test to measure the accuracy of the model such as in studies conducted in prostate cancer probability (Wang *et al*., 2011), pregnancies (Condous *et al*., 2004) and serious injuries (Kononen *et al*., 2011). It has the advantage in that normal distribution is not necessary for the dependent variable. However, there is discussion about how well a model fits the data between the gold standard and diagnostic test (Hosmer and Lemeshow, 2000). Thus, the goodness of fit is usually measured by ROC for assessing the fit of the logistic regression model.

However, the logistic regression model has a limitation when used with small sample size (under about 200) and with regard to the number of events per factor when was adding or deleting might change the selected factors (King and Zeng, 2001).

1.4.5 Statistical method for investigating the patterns of transport accident deaths

Poisson regression is generally used for modelling mortality rates which is not many zero of observed data. Several studies used Poisson regression on small study areas. For example, Lix *et al*. (2004) used Poisson regression to investigate the relation of population health status to demographic, geographical, and temporal explanatory variables in Manitoba, Canada between 1985 and 1999. This study suggested that Poisson regression model enables analysis to test differences of population health status between small areas and time periods. This is consistent with Yang *et al*. (2005) who used a Poisson regression model to estimate trends and compare age- and sex-specific mortality rates due to injuries in Guangxi province of South Western China from 1997 to 2001, using death certificates. This model calculated a probability model for an observed number of cases having a Poisson distribution within the study period. The result showed a marginally significant trend of increasing injury mortality within the period.

In other studies, Poisson regression models were applied to analyze specific cause mortalities such as Odton *et al*. (2010). They used a Poisson regression model for analyzing variations in all-cause mortality in Thailand in 18 age-groups and 235 superdistricts during 1999 to 2001 based on death certificates. The results found that the model could provide a good fit to these data. Whereas, the study of

Sriwattanapongse *et al*. (2013) could not use a Poisson regression model for forecasting transportation accident mortality rates in Thailand based on death certificate reports from 2000 to 2009 due to the many zeros observed on small cells (55.8%).

1.4.6 Summary of literature review

Overall, this literature review can be summarized as follows:

1. Transport accident deaths are a serious and prolonged threat which have substantially increased mortality rates beyond the target rates of safety in 2020 in Thailand.

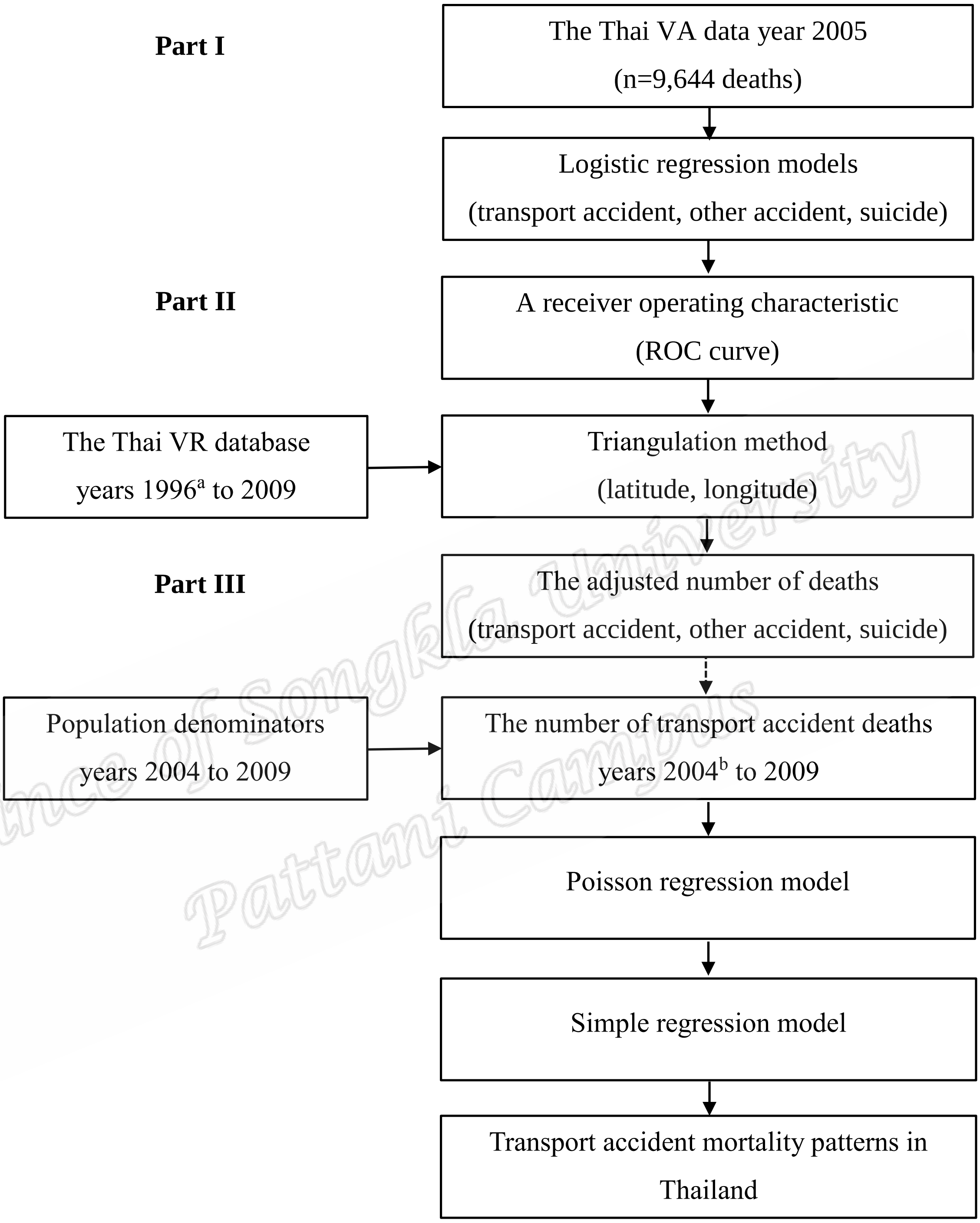
2. Thailand has both unreliable cause of death from VR data and more reliable cause of death from the Thai 2005 VA data.

3. Logistic regression model is appropriate for estimating the number of transport accident deaths based on the VA data and the model can be verified using the ROC method.

4. Poisson regression model is appropriate for count data which are not many zeros on small cell.

This thesis consists of three parts: part I, creating an appropriate model for estimating transport accident death and external causes of deaths based on VA study, part II, estimating the actual number of transport accident death and external causes of deaths from VR data and part III, exploring patterns of transport accident mortality rates in Thailand.

To illustrate the workflow of this study, a path diagram, shown in Figure 1.5, is used to display each step of the overall approach of this study.



[a] The starting point of the electronic death data files

[b] The analysis is restricted to 2004-2009, period of least error for VR data

Figure 1.5 Workflow of this study

1.5 Path diagram and variables

The first and second parts of this thesis involve, creating an appropriate model based on VA study and estimating the actual number of transport accident death and external causes of deaths from VR data. The target population was all reported Thai deaths during the years 1996 to 2009. Samples were the Thai 2005 VA study, which assessed true causes of death for 9,644 selected residents from nine provinces. Logistic regression was used for analysis with three determinants; province, gender-age group and VR cause-location group. The path diagram is used to summarize the variables considered in part I and part II as shown in Figures 1.6 – 1.8.

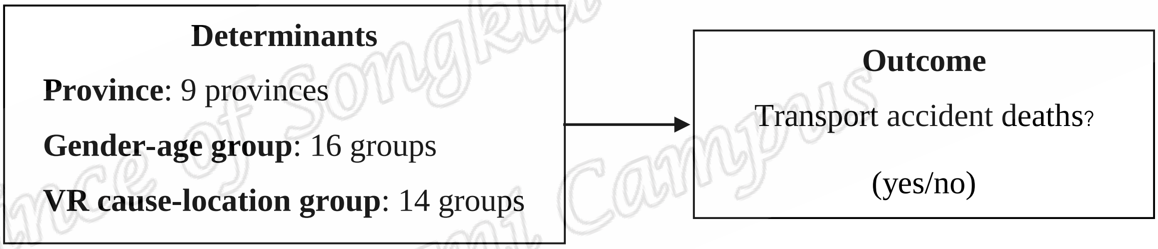


Figure 1.6 Path diagram for transport accident deaths

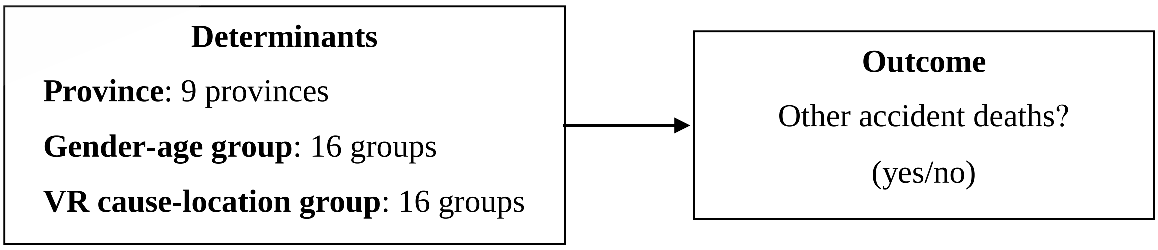


Figure 1.7 Path diagram for other accident deaths

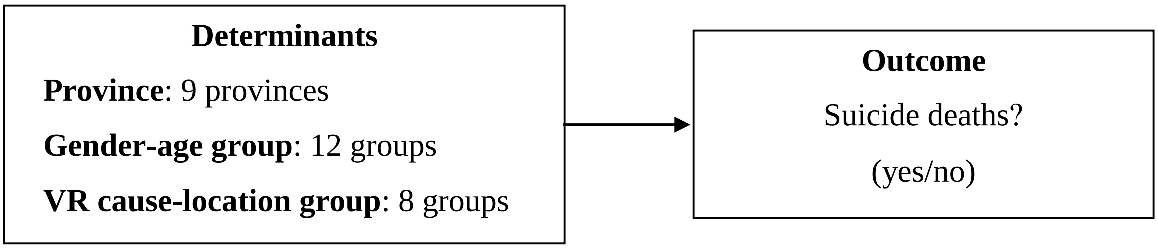


Figure 1.8 Path diagrams for suicide deaths

For the third part of the thesis, exploring patterns of transport accident mortality rates in Thailand. The target population was the whole Thai population during the years 2004 to 2009. Samples were the adjusted number of transport accident deaths during the years 2004 to 2009, which were estimated in the second part of the thesis. A Poisson regression model was used for exploring patterns of transport accident mortality rates with two determinants; gender-age group and province-year group. The path diagram is used to summarize the variables considered in part III as shown in Figures 1.9.

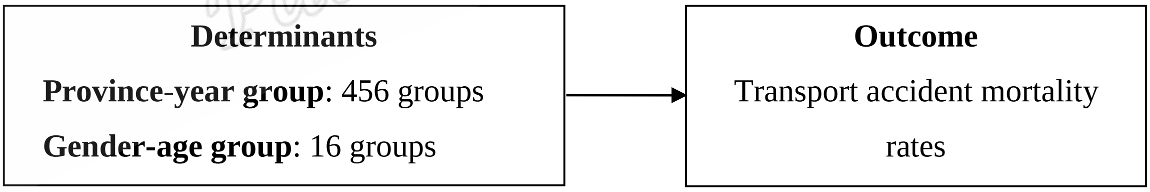


Figure 1.9 Path diagram for transport accident mortality rates