Statistical Methods for Comparing Mortality between Populations with Application to Mortality Data in Thailand in 2005 and Japan in 2006

Arinda Ma-a-lee

A Thesis Submitted in Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Research Methodology

Prince of Songkla University

2016

Thesis Title Statistical Methods for Comparing Mortality between Populations with Application to Mortality Data in Thailand in 2005 and Japan in 2006

Author Miss Arinda Ma-a-lee

Major Program Research Methodology

Major Advisor

…………………………......................

(Assist. Prof. Dr. Phattrawan Tongkumchum)

Co-advisor:

....

(Asst. Prof. Dr. Chamnein Choonpradup)

…………………………………

(Emeritus Prof.Dr.Don McNeil)

Examining Committee:

……………………………....Chairperson

(Assist. Prof. Dr. Apiradee Lim)

…………………………..........Committee

(Assist. Prof. Dr. Phattrawan Tongkumchum)

....Committee

(Asst. Prof. Dr. Chamnein Choonpradup)

....Committee

(Emeritus Prof. Dr. Don McNeil)

....Committee

(Assoc. Prof. Dr. Yothin Sawangdee)

The Graduate School, Prince of Songkla University, has approved this thesis as fulfillment of the requirements for the Doctor of Philosophy Degree in Research Methodology.

....……………………………….....

(Assoc. Prof. Dr. Teerapol Srichana)

Dean of Graduate School

This is to certify that the work here submitted is the result of the candidate’s own investigations. Due acknowledgements has been made of any assistance received.

..…………..………….........Signature

(Asst. Prof. Dr. Phattrawan Tongkumchum)

Major Advisor

...…………..……………........Signature

(Miss Arinda Ma-a-lee)

Candidate

I hereby certify that this work has not been accepted in substance for any degree, and is not being currently submitted in candidature for any degree.

……………..…….……………Signature

(Miss Arinda Ma-a-lee)

Candidate

ชื่อวิทยานิพนธ์ วิธีการทางสถิติสาหรับเปรียบเทียบการตายระหว่างประชากร ประยุกต์ใช้ข้อมูลการตายในประเทศไทย ปี พ.ศ. 2548 และ ญี่ปุ่น พ.ศ. 2549

ผู้เขียน นางสาวอารินดา มะอาลี

สาขาวิชา วิธีวิทยาการวิจัย

ปีการศึกษา 2558

บทคัดย่อ

วิทยานิพนธ์นี้ ได้อธิบายวิธีการทางสถิติที่เหมาะสมสำหรับเปรียบเทียบการตายระหว่างประชากร โดยแบ่งการศึกษาเป็นสามส่วนดังนี้

ส่วนที่หนึ่ง มีวัตถุประสงค์เพื่อเปรียบเทียบการตายจาก 21 สาเหตุหลัก ของประเทศไทย ปี พ.ศ. 2548 และการตายของประเทศญี่ปุ่น พ.ศ. 2549 ส่วนที่สอง เป็นการศึกษาเปรียบเทียบอัตราการตายจากโรคอัมพฤกษ์ อัมพาต ในจังหวัดสุพรรณบุรีกับกรุงเทพมหานคร ปี พ.ศ. 2548 และส่วนที่สาม เป็นการสรุปวิธีการปรับแก้ข้อมูลรายงานสาเหตุการตายที่คลาดเคลื่อนจากระบบทะเบียนราษฎร์ โดยใช้วิธีการทางสถิติประมาณจำนวนการตายของ 21 สาเหตุในประเทศไทย

สาเหตุการตายที่บันทึกไว้ในฐานข้อมูลจากทะเบียนราษฎร์มีความคลาดเคลื่อน เนื่องจากบันทึกสาเหตุเป็น “ไม่ชัดแจ้ง” ประมาณ 40% การศึกษานี้ จึงนำข้อมูลจำนวนการตายตามสาเหตุจากทะเบียนราษฎร์ มาประมาณใหม่โดยอาศัยสาเหตุการตายจากการสอบสวนสาเหตุการตายด้วยการสัมภาษณ์ (Verbal autopsy) ใน ปี พ.ศ. 2548 โดยใช้การวิเคราะห์การถดถอยโลจิสติก (logistic regression) และเนื่องจากประชากรแต่ละกลุ่มมีโครงสร้างต่างกัน เมื่อจำแนกตามเพศและอายุ จึงใช้วิธีการปรับฐานอายุ (age-standardization) เพื่อให้สามารถเปรียบเทียบกันได้

ส่วนที่หนึ่ง เป็นการใช้กราฟโดนัท เพื่อแสดงการตายส่วนเกิน (excess deaths) ในแต่ละสาเหตุของไทยและญี่ปุ่น จำแนกตามเพศ ผลการศึกษาพบว่า โดยส่วนใหญ่ การตายส่วนเกินของประเทศไทยที่มากกว่าญี่ปุ่น ได้แก่ การตายจากสาเหตุ อัมพฤกษ์ อัมพาต อุบัติเหตุการจราจรในเพศชาย ระบบภายในของเพศหญิง โรคติดเชื้อและมะเร็งตับ ส่วนญี่ปุ่น โดยส่วนใหญ่ การตายส่วนเกินที่ญี่ปุ่นมากกว่าประเทศไทย ได้แก่ การตายจากสาเหตุ การฆ่าตัวตายและมะเร็งในระบบทางเดินอาหาร ดังนั้น หน่วยงานของรัฐที่ดูแลเกี่ยวกับระบบสุขภาพ ควรตระหนักและให้ความสนใจ ในการกำหนดเป็นนโยบายสุขภาพเร่งด่วน เพื่อลดการตายส่วนเกินในแต่ละสาเหตุ ที่ประเทศไทยมากกว่าญี่ปุ่น

ส่วนที่สอง เป็นการตรวจสอบความมีนัยสำคัญทางสถิติ โดยการใช้การถดถอยพัวซอง (Poisson regression) เพื่อสร้างตัวแบบเปรียบเทียบอัตราการตายจากโรคอัมพฤกษ์ อัมพาต

หลังจากปรับอิทธิพลของปัจจัยทางประชากรแล้ว ของกรุงเทพมหานครเปรียบเทียบกับสุพรรณบุรี ผลการศึกษาพบว่า อัตราการตายของโรคอัมพฤกษ์ อัมพาต ในจังหวัดสุพรรณบุรี สูงกว่ากรุงเทพมหานครอย่างมีนัยสำคัญทางสถิติ และอัตราการตายในเพศชายสูงกว่าในเพศหญิง ในทุกกลุ่มอายุที่น้อยกว่า 80 ปี ดังนั้น การเพิ่มความพยายามในการดูแลสุขภาพและการป้องกันโรคอัมพฤกษ์ อัมพาต ควรกำหนดเป้าหมายพื้นที่ชานเมือง และวิธีการทางสถิติที่ใช้ในการศึกษานี้ สามารถนำมาใช้กับพื้นที่อื่น ๆ สำหรับการเปรียบเทียบอัตราการตายสาเหตุเฉพาะโรค ระหว่างประชากร

ส่วนที่ สาม เป็นการสรุปวิธีการปรับแก้ข้อมูลรายงานสาเหตุการตายที่คลาดเคลื่อน โดยใช้วิธีการทางสถิติประมาณจำนวนการตายของ 21 สาเหตุ เนื่องจากตัวแปรตามที่สนใจ เป็นแบบกลุ่มที่มี 21 ระดับ ตัวแบบที่เหมาะสมสำหรับการวิเคราะห์ คือ การถดถอยพหุนาม (multinomial regression) ในการศึกษานี้ แสดงให้เห็นว่า สามารถใช้การวิเคราะห์ถดถอยโลจิสติก สร้างตัวแบบของสาเหตุการตายแต่ละสาเหตุจำนวน 21 ตัวแบบ จากข้อมูลการสอบสวนสาเหตุการตายด้วยการสัมภาษณ์ในประเทศไทย ปี พ.ศ. 2548 หลังจากนั้น ปรับขนาดจำนวนการตายโดยรวมในแต่ละสาเหตุที่ประมาณได้ ให้สอดคล้องกับจำนวนการตายโดยรวมจากทะเบียนราษฎร์

การใช้การวิเคราะห์ถดถอยโลจิสติก สร้างตัวแบบกับข้อมูล ในแต่ละสาเหตุ สามารถแสดงช่วงความเชื่อมั่นของร้อยละของการเสียชีวิต จากแต่ละสาเหตุในแต่ละระดับของปัจจัยเสี่ยง หลังจากปรับอิทธิพลของปัจจัยอื่น ๆแล้ว และนำไปเปรียบเทียบกับแผนภูมิแท่งของร้อยละของการเสียชีวิตเพื่อประเมินอิทธิพลของปัจจัยกวน กราฟพื้นที่ (Area plot) ใช้แสดงผล การตายจำแนกตามเพศ กลุ่มอายุและปี วิธีการในส่วนที่สามนี้ ได้ประยุกต์ใช้กับข้อมูลการตายจาก โรคอัมพฤกต์ อัมพาต

Thesis Title	Statistical Methods for Comparing Mortality between Populations with Application to Mortality Data in Thailand in 2005 and Japan in 2006
Author	Miss Arinda Ma-a-lee
Major Program	Research Methodology
Academic Year	2015

ABSTRACT

In this thesis, appropriate statistical methods were described to compare mortality between populations. The thesis comprises three parts.

The first part aimed to compare 21 cause-specific deaths in Thailand in 2005 and Japan in 2006. The second part aimed to compare stroke death rates in Suphan Buri with Bangkok in 2005. Deaths from the Thai death registration (DR) database were adjusted for misclassified cause of death using logistic regression based on the verbal autopsy (VA) data. Age-standardization was used to adjust for population differences in gender and age group. The third part aimed to summarize methods for adjusting misclassification causes of deaths in the DR data.

In the first part, a useful doughnut plots was used to display 21 cause-specific excess deaths by gender. In Thailand, there were cause-specific excess deaths for transport accident for males, endocrine for females, infectious diseases, stroke, and liver cancer. In contrast, there were cause-specific excess deaths for suicide and other digestive cancer for Japan. The more cause-specific excess deaths in Thailand impose a significant public health burden that demands attention and action by health policy makers.

In the second part, the high excess death of stroke from part 1 was further investigated. A Poisson regression model was used to compare stroke death rates after adjusting for demographic factors. The model provides a good fit to the data. The stroke death rate in Suphan Buri was statistically significant higher than Bangkok. The rates were higher for males than for females in all age groups under 80 years. Increasing efforts in health care and stroke prevention should target suburban areas. The statistical methods used in this study can be applied to other regions for comparing a cause-specific mortality between populations.

In the third part, causes of deaths in Thailand are misreported because about 40% of deaths have been recorded as ill-defined. This study aims to offer statistical methods to correct misreported multinomial outcome. The methods involve analysis of verbal autopsy (VA) data. Since the outcome is a nominal variable with 21 levels, the appropriate model for systematic analysis of death by ICD-10 code is multinomial regression. However, it is simpler and more informative to separately fit logistic regression models to the 21 outcome cause groups, and then rescale the results to ensure that the total number of estimated deaths for each group match those reported in the corresponding populations. This method also gives confidence intervals for percentages of deaths in cause groups for levels of each risk factor adjusted for other risk factors. These confidence intervals are compared with bar charts of sample percentages to assess evidence of confounding bias. Area plots are used to show results by gender, age group and year. The methods were illustrated using stroke deaths.

Acknowledgments

I would like to express my gratitude and appreciation to my supervisors, Asst. Prof. Dr. Phattrawan Tongkumchum, Emeritus Prof. Dr. Don McNeil and Asst. Prof. Dr. Chamnein Choonpradub for their invaluable assistance and helpful guidance on this thesis.

My special thanks to Prof. Val Gebski from NHMRC Clinical Trial Centre, university of Sydney, Prof. Malcolm Hudson and Dr. Hilary Green from Macquarie University for their very kindness help and very useful suggestions to my thesis.

I would like to acknowledge the program of Strategic Scholarships Fellowships Frontier Research Networks (Specific for Southern Region) for the PhD. Program Thai Doctoral degree from the Office of the Higher Education Commission of Thailand, and the Centre of Excellence in Mathematics, the Commission of Higher Education, Thailand, for funding this study and awarding me a scholarship and also thank you to the Graduate School, Prince of Songkla University for partially funding this study.

I am grateful to Ministry of Public Health and Dr. Kanitta Bundhamcharoen for providing the data. I would like to thank all of students in Research Methodology program for the encouragement. Finally, I would like to express my thankfulness to my family and all friends for cheering me up throughout this study.

Arinda Ma-a-lee

Contents

\

List of Tables

List of Figures

CHAPTER 1

Introduction

The differences in life expectancy between countries are well known. What are the implications of people in developing countries dying earlier than those in developed countries? And what are the causes of deaths that account for this difference?

This study compared 21 cause-specific deaths in Thailand in 2005 and Japan in 2006 which was adjusted for misclassified cause of death using 2005 verbal autopsy (VA) data. Appropriate statistical methods were used to compare stroke death rates in Suphan Buri with Bangkok in 2005. We also summarize methods for correcting misclassified cause of deaths in a multinomial outcome and the methods were illustrated using stroke deaths data.

This chapter presents overall background, Thai mortality and 2005 VA data and Japan data. Literatures were summarized on statistical methods for comparing mortality between populations and for correcting misclassified cause of deaths. A roadmap of the present study is also provided.

1.1 Background

Estimating overall mortality and why people have died are essential for monitoring health and planning appropriate health services. Mortality statistics reflect how various diseases and injuries are affecting the living. Having cause-specific mortality help health authorities determine whether they are focusing on the right kind of public health actions.

Developed countries have systems in place for assessing causes of death in the population and their mortality statistics are available. Most developing countries do not have such systems. Although some countries including Thailand have such systems, causes of deaths are often misclassified (Mathers *et al.*, 2005). Therefore, numbers of deaths from specific causes have to be estimated from incomplete data. VA is widely used to verify cause of deaths (Setel *et al.*, 2005; Ngo *et al.*, 2010; Rao *et al.*, 2010) in order to get more accurate causes.

The gap of life expectancy between developed and developing countries are well known. Japan is one of the developed countries. In 2011, Japan had life expectancy at birth of 79 years for males and 86 years for females whereas in Thailand life expectancy at birth was 71 years for males and 77 years for females (WHO, 2013). The difference in life expectancy between these two countries corresponds to males and females in Thailand dying 8 and 9 years earlier, than those in Japan. It is important to understand why this is so. Moreover, comparison of cause-specific deaths to those developed countries is crucial for improving health and reducing preventable deaths in developing world.

Stroke is a preventable and the leading cause of death among males (9.4%) and females (11.3%) in Thailand in 2005 (Rao *et al.*, 2010). There is also a variation of stroke deaths between five regions of Thailand with highest prevalence in the central region (2.4%), followed by the southern (2.3%), northern (1.5%), and north-eastern (1.1%) (Suwanwela, 2014). The central region of Thailand comprises of 11 provinces. Of these, Suphan Buri and Bangkok were sampled in the 2005 VA survey (Rao *et al.*, 2010).

Standardized mortality ratios (SMR) are widely used to compare all-cause and cause-specific deaths including stroke between and within population (Fukuda *et al.,* 2004; Torrey and Haub, 2004; Famnuayphol *et al.,* 2008). However, these studies did not compare deaths based on the statistical model. In this study, we used appropriate statistical methods for comparing cause-specific deaths between populations.

Comparison of cause-specific deaths between countries requires good quality data. This study offers methods for correcting multinomial outcome data (21 cause groups) with application to stroke mortality in Thailand. The methods involve analysis of 2005 VA sample. Since the outcome is nominal variable with 21 levels, the appropriate model for systematic analysis of death by ICD-10 code is multinomial regression. However, it is simpler and more informative to separately fit logistic regression models to the 21 outcome cause groups, and then rescale the results to ensure that the total numbers of estimated deaths for each group match those reported in the corresponding populations.

1.2 Literature review

1.2.1 Thai mortality registration data

Mortality data is one of the most important health indicators for measuring country's health and development. Mortality data are used for identifying and monitoring health program, and also used to measure and compare mortality rates between and within populations.

In Thailand, the most important source of mortality data is vital registration (VR) system, which was initiated in 1916. Civil registration law mandates that every vital event (births and deaths) be registered at the offices of the district or municipality registrars, which are under the Interior Ministry. According to this law, deaths must be registered within 24 hours. The registration process for deaths is not straightforward because two steps are involved: (i) notifying and authorized person about the death, and (ii) validating and registering the death (Prasartkul and Vapattanawong, 2006). After both registration steps are completed, the name of the deceased is deleted from the household roster and the details concerning the deaths are entered into the registration system. Before the end of each year the Interior Ministry counts the births, deaths and total population and makes the figures publicly available (Vapattanawong and Prasartkul, 2011).

However, mortality data are often poor in developing countries because of inadequate death registration (DR) (Huong *et al.,* 2006; Rukumnuaykit, 2006). Recently mortality statistics in Thailand were classified as low quality with over 30% of deaths unregistered and about 40% of deaths remained attributed to ill-defined cause (Mathers *et al.,* 2005; Fottrell and Byass, 2010). The high ill-defined rate is usually found in rural areas where physicians are not available to identify the cause of death. Therefore, VA is an important method to verify cause of death (Fottrell and Byass, 2010; Setel *et al.*, 2006).

1.2.2 Verbal Autopsy study

Verbal autopsy (VA) studies are widely used throughout the developing world to estimate cause-specific death. The VA method gave more precise cause of death than those reported in the DR (King and Lu, 2008). The VA data set can be used to inform the implementation of VA to more reliably verify cause of death in national health information systems (Rao *et al.,* 2010).

The standard VA study and medical records review procedures were used to verify registered causes. They have been successfully applied in China and Iran (Rao *et al.,* 2005; Ngo *et al.,* 2010). In Thailand, the latest VA has been conducted in 2005 (Choprapawon *et al.,* 2005; Pattaraarchachai *et al.,* 2010; Polprasert *et al.,* 2010; Porapakkham *et al.,* 2010; Rao *et al.,* 2010). Mortality based on the 2005 VA study has been published. The five leading causes of death among males are stroke, transport accident, HIV/AIDS, ischemic heart diseases, and chronic obstructive lung diseases. Among females, leading causes of death are stroke, diabetes, ischemic heart diseases, HIV/AIDS, and renal diseases (Rao *et al.,* 2010).

1.2.3 Japan mortality data

Japan mortality data in 2006 were available from the World Health Organization, which is includes details of deaths registered by Japan civil registration (VR) (WHO, 2011). Cause of death in Japan data was classified as high quality with less than 4% of deaths are ill-defined (Carmichael, 2011) and cause of death for years 1950-2000 has 97% coverage whereas coverage for Thai data is 89% (Mathers *et al.*, 2005).

1.2.4 Principal methods for mortality comparison

The total number of deaths is a useful measure for quantifying the burden of disease in a population. However, the absolute number of deaths is less useful for comparison between populations because of population size. Larger population group tends to generate more deaths than smaller group (Robson *et al.*, 2007). Consequently, to compare deaths among group, the number of deaths must be related to the “population at risk” of dying to calculate death rates (Curtin and Klein, 1995).

The *crude death rate (CDR)* is a summary measure, defined as the total number of deaths divided by the total midyear population and often expressed per 100,000 populations. Although it does relate the number of deaths to the population, comparison of crude death rate would be misleading since this rate does not take into account the age distribution of the population. As such, it is not appropriate measure for comparing differences between populations. Because the risk of dying is much higher in older than younger ages, populations with a higher proportion of older age group tend to generate higher crude death rate than younger populations (Anderson and Rosenberg, 1998).

The *age-specific death rate (ASDR)* can be used as an alternative to CDR. The most reliable method of comparing death rate between populations is to compare individual ASDR for each age group. However, this method mostly requires a large number of comparisons (Curtin and Klein, 1995). Therefore, age-standardization is a method that controls for the population age distribution, and make meaningful comparison between populations.

Age-standardization

Age-standardization is a method used to depict the health of populations and essentially to establish rates for comparison purposes (Li *et al.*, 2008; Julious *et al.*, 2001). It aims to control for variation in age distribution. It is classified into direct and indirect approaches that controls for variation in age distributions. *Direct age-standardization* is a method that ASDRs from study population are applied to a standard population distribution. The resultant summary index is call age-standardized death rate. *Indirect age-standardization* is frequently used in several comparison studies between populations (Chah *et al.,* 2014; Fukuda *et al.,* 2003; Torrey and Haub, 2004; Famnuayphol *et al.,* 2008). This method uses population with lower mortality rate as a standard population. A standard set of ASDRs are applied to the study population providing an "expected" number of deaths in a study population. The index most frequently used is *the standardized mortality ratio (SMR)*, which is the ratio of the observed to the expected number of deaths.

To use the age-standardization requires a "standard population" which is a set of arbitrary population weights. There is closely no conceptual justification for choosing one standard over another. The standard population should be chosen depending on the study population being examined. Generally, the standard population is chosen from the population with lower age-standardized death rate. Tsai and Wen (1986) pointed that the choice of selection of standard population is crucial because it may lead to different conclusions.

Shah *et al.* (2014) investigated the mortality differences according to marital status (married, never married (single), divorced and widowed) among men and women in Kuwait. It was found that married persons have lower mortality than the others status, so this study used married persons as standard population. Another study compared mortality in US with Canadian in 1998. The Canadian population was used as a standard population to adjust for population differences in gender and age group (Torrey and Haub, 2004). The reason for choosing standard is based on that the life expectancy of Canadian is longer than US population.

Excess deaths: a mortality comparison

A comparison analysis always provides the result of excess deaths. Savesh *et al.* (2013) studied about comparison of mortality trend and pattern between India and Japan from 1950 to 2010. This study used data from various published studies of India and Japan. India population has a younger age distribution whereas Japan has older age distribution. However, this analysis does not take into account the different in population age distribution. Therefore, it can give biased results (McNamee, 2004).

Studies on excess mortality using indirect age-standardization have been reported (see, for example, Torrey and Haub, 2004; Satcher *et al.*, 2005; Fukuda *et al.*, 2003; Faramnuayphol *et al.*, 2008).

Torrey and Haub (2004) compared US and Canadian mortality in 1998. Indirect age-standardization was used to adjust for population differences in age distribution, by applying Canadian gender age-specific death rates for all-cause and cause-specific deaths to the US population. The difference between the observed deaths and the expected deaths represents the excess deaths in 1998 US population.

Satcher *et al.* (2005) examined trend in Black-White standardized mortality ratios for each gender-age group from 1960-2000. The difference between the number of observed deaths and the expected deaths were calculated to determine the excess deaths. They found that Black-white gap measured by SMR changed very little between 1960 and 2000. The estimated excess deaths were higher in Black population than White.

Fukuda *et al.* (2003) quantified the geographical variation of mortality in Japan in 1973-1998. This study estimated excess deaths attributable to the variation among municipalities across Japan. The municipalities were divided into quintiles according to the SMR. The lowest quintiles as the standard, the number of excess deaths and its ratio to observed deaths were estimated.

Faramnuayphol *et al.* (2008) analyzed the geographical variation of mortality in Thailand. SMR was calculated by dividing the observed number of deaths by the expected number of deaths for the overall mortality and cause-specific mortality at district level. Expected number of deaths for each area was calculated from national age-specific mortality rate and age structure of population in that area. They found that overall mortality was concentrated in the middle part of upper north. Clustering of cause-specific in a single region was found for liver cancer (in the upper northeast) and chronic obstructive pulmonary disease (in upper north).

The principal measures of excess mortality as considered above are the difference between the observed and the expected deaths or the ratio between the observed and the expected deaths (SMR) (Peter and Haberman, 1993; Howel, 1995). However, these analysis methods were not based on the statistical significant test. Many studies suggested that when using SMR, its precision should be calculated (Breslow and Day, 197s; Court and Cheng, 1995; Howel, 1995). Since this method does not directly provide 95% confidence interval and p-values therefore an appropriate statistical model is needed.

1.2.5 Statistical methods

Statistical methods are useful for analyzing mortality data. Linear regression is used to investigate the associations between continuous outcome and covariates, assuming independent error terms. However, Mortality data is often involved counts as outcome variables rather than continuous outcomes. The Poisson regression model is generally applied to situation in which the outcome variable is count (e.g., number of stroke deaths in a population) (McCullagh and Nelder, 1989; Venable and Ripley, 2002). The modeling approaches, based on the theory of generalized linear models, allow us to build models containing several explanatory variables. The statistical significance of these variables can be tested and the effect of the interactions between the variables can be assessed rigorously. Moreover, the results based on the model provide conclusion of the study.

However, mortality count data are always over dispersed, with the variance greater than the mean. To accommodate such over-dispersion, several studies were turned to over-dispersed Poisson and negative binomial regression models. As a natural extension of the Poisson distribution, the negative binomial distribution is more flexible and allows for over-dispersion (Davis and Wu, 2009).

Poisson and negative binomial regression has been applied in several studies. Li *et al.* (2008) used Poison model for comparing trends in cancer rates across regions. Another study conducted in Spain to investigate the burden of deprivation-associated excess deaths by cause groups in each gender and age-groups, using Poisson model (Benach *et al.* (2003). Steward *et al.* (1999) compared stroke mortality between ethnic groups. The incident rates adjusted for age and sex for stroke with 95% CI were calculated using Poisson regression model.

Kaewsompak *et al.* (2005) used Poisson and Negative binomial model to investigate relation between the incidence rates in terms of geographical patterns. Lim and Choonpradub (2007) studied the temporal patterns of death reported from infectious diseases in 14 southern provinces of Thailand based on Poisson regression and negative binomial regression.

1.2.6 Methods for correcting misclassified cause of deaths

There are several studies for estimating causes of death based on the 2000 VA data (Rao *et al.*, 2010; Pattaraarchachai *et al.*, 2010; Porapakkham *et al.*, 2010). However, these studies were not based on statistical models. A benefit of using statistical models is that many factors can be investigated and the effect of one factor can be adjusted for other factors. Moreover, a statistical model can detect confounding factors that biases the results.

1.2.7 Summary of literature review

The literature review can be summarized as follows.

Thai DR data is low quality because of misclassified cause of deaths. VA is commonly used to verify cause of death in many studies in developing countries where DR data are of poor quality. Therefore, it is necessary to adjust for the misclassification to get more accurate cause using VA data.

Age-standardization is an essential method that adjusts for differences in population age distribution, and to make meaningful comparison.

Several comparison studies always provide the results of “excess deaths” based on the preliminary analysis of the differences between the observed and expected deaths, and the SMR.

Statistical model is needed to detect the statistical significant excess. Generally mortality counts data has Poisson distribution. There are some situations that over dispersion occur. In this case, negative binomial can be applied (McCullagh and Nelder, 1989).

Since causes of deaths were categorized into 21 groups the causes of death outcome is nominal variable with 21 levels, the appropriate model for systematic analysis of death by ICD-10 code is multinomial regression. However, it is simpler and more informative to separately fit logistic regression models to 21 outcome cause groups, and then rescale the results to ensure that the total numbers of estimated deaths for each group match those reported in the corresponding populations.

1.3 Rational of study

Comparison of mortality is important for monitoring population health status. It can be used for planning appropriate health services for improving population health. In Thailand, most of comparison studies have analyzed mortality data to explore all-cause and cause specific excess deaths based on the descriptive and preliminary analysis using SMR. This method does not directly provide 95% confidence interval and p-value. Therefore, appropriate statistical methods need to be used to study mortality comparison and corresponding graphical methods need to be developed for visually comparison mortality data. The end of the study summarized methods for correcting misreported multinomial outcome data with application to stroke deaths.

1.4 Objectives

The objectives of this study are:

1. To compare cause-specific deaths in Thailand in 2005 with Japan in 2006.
2. To describe statistical methods for comparing stroke death rates in Suphan Buri with Bangkok in 2005.
3. To summaries methods on adjustment of misclassification causes of death.

1.5 Roadmap of the present study

This thesis consists of four chapters to describe details of the knowledge that found in the study. Chapter 1 contains introduction, background, literature review, rational and objectives of the study. Chapter 2 describes a detail of methodology including data source and management and analysis methods used in the study. Chapter 3 contains the preliminary data analysis and includes these studies comprising a comparison of cause-specific deaths in Thailand in 2005 and Japan in 2006, statistical methods for comparing stroke death rates in Suphan Buri with Bangkok in 2005, and methods for adjustment of misclassification on cause of death. Finally, the discussion and conclusion are addressed in chapter 4.

CHAPTER 2

Methodology

This chapter describes the data sources and management, and summarizes the statistical and graphical methods used for analyzing the data.

2.1 Data sources and data management

Deaths and Population

Thai Data

The 2005 verbal autopsy (VA) and Thai death registry (DR) data were used to estimate number of deaths. Both databases were obtained from Bureau of Health Policy and Strategy, Ministry of Public Health. The DR data is misclassified cause of death with more than 30% of ill-defined cause (Mathers *et al.*, 2005). The VA data contain information of deceased person on province of residence, gender, age group, location (death occurred in or outside hospital), DR reported ICD-10 code and VA-assessed ICD-10 code.

The DR-reported ICD-10 code is from register cause of death where the VA-assessed ICD-10 code is based on verification cause of death by medical expert team under the SPICE project. Therefore, causes of death in the VA data are more reliable than those in the DR.

The misclassification cause of death in the DR data was adjusted using the VA data. This adjusting method is described in detailed elsewhere (Chutinantakul *et al.*,2014;

Klinjun *et al.*, 2015). Figure 2.1 shows a flow diagram for estimating number of deaths for the whole of Thailand in 2005.

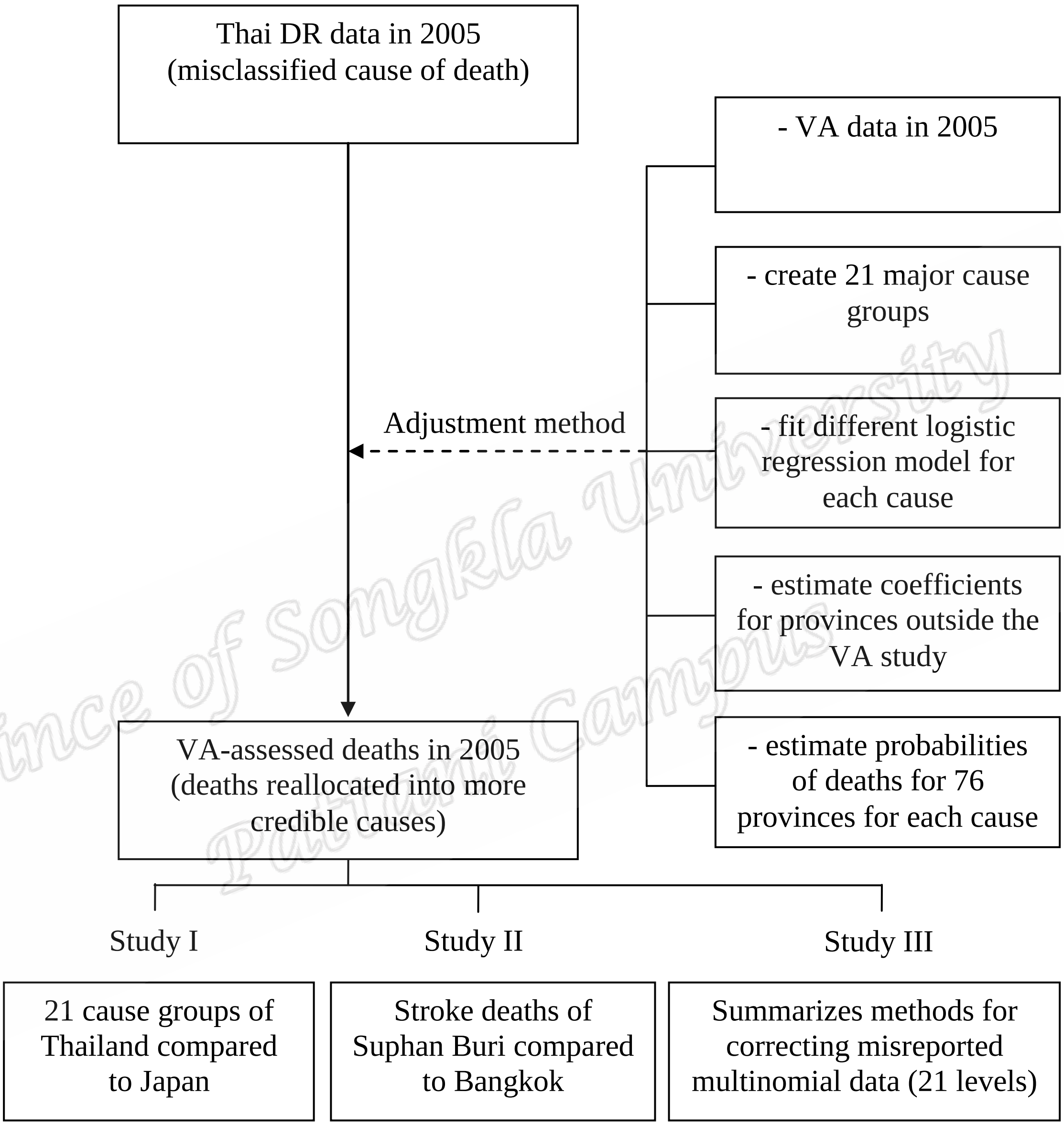


Figure 2.1 Diagram for estimating deaths in 2005 for each 21 cause groups

Data management began with systematic analysis of the VA data used chapter-block classification of ICD-10 codes (WHO, 2004), which is consisting of blocks classified mainly by human organs, creating 21 major cause groups with VA counts is shown in Table 2.1.

Table 2.1 Definition of 21 major cause groups with VA counts

Cause groups	Count	Cause groups	Count
1: Transport Accident (V)	536	11: Genitourinary (N)	412
2: Other Injury (W, X0-59)	327	12: Ischemic (I20-25)	617
3: Suicide (X60-84)	158	13: Stroke (I60-69)	1076
4: All other	358	14: Other CVD (I^-)	540
5: TB (A15-19)	195	15: Liver Cancer (C22)	500
6: HIV (B20-24)	512	16: Lung Cancer$^+$(C30-39)	320
7: Other Infectious (A, B)$^-$	219	17: Other Digestive (C15-26$^-$)	290
8: Endocrine (E)	647	18: Other Cancer (C^-, D0-48)	697
9: Mental, Nervous (F, G)	223	19: Respiratory (J)	801
10: Digestive (K)	489	20: Septicaemia (A40-41)	77
		21: Ill-defined (R)	501

$^+$ Respiratory/thoracic, $^-$ exclude above

Logistic regression model was fitted to death for each cause with geographic (province), demographic (gender, age-group) and medical (DR-reported cause group and location) factors as determinants. A Receiver Operating Characteristic (ROC) was used to assess the model (Fan *et al.*, 2006; Sakar and Midi, 2010). A spatial triangulation method was used to estimate coefficients for the provinces outside the VA study. The estimated probabilities from the model were applied to the DR data giving the unscaled estimated deaths for the whole of Thailand. Before further analysis, we scaled the unscaled totals estimated deaths to match the totals reported deaths. Then the estimated deaths for the whole of Thailand were used to compare deaths with Japan as present in study I, and then Suphan Buri and Bangkok were chosen for study II. The adjustment methods were described and illustrated using stroke death in study III.

The projected populations were used. The projection method is based on the assumptions about mortality and female fertility (Vapattanawong and Prasartkul, 2011; Health Research System Institute, 2003). We developed a mathematical model for projection, which is formulated in terms of the size of the population, the mortality and fertility rate per unit population. To avoid the estimation of the age-specific fertility; we assume that births decrease by 2% annually. From the model we start with gender-specific population census age distributions in 2000, and we used reported death rates by age and gender from 2000 to 2009. Thus projected populations for the whole of Thailand from 2000 to 2009 were obtained. Then, the 2005 projected population for 76 provinces was used for study I, and for study II Suphan Buri and Bangkok were chosen for comparison.

Japan Data

The numbers of deaths for Japan in 2006 by gender, five year age group and ICD-10 codes, with corresponding population at risk, are available from the World Health Organization (WHO, 2011).

2.2 Variables

The path diagrams for the variables used in the three studies are shown in Figure 2.2. The study I and study II, the VA-assessed deaths outcome is counts. Determinants are demographic (gender-age group), medical (cause group) and geographical factors.

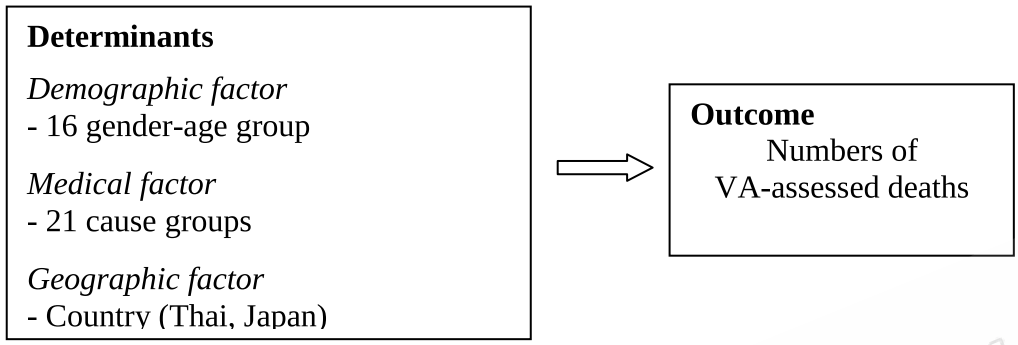
Study III, the misclassification was assessed by cross tabulation between VA-assessed and DR- reported ICD-10 codes. The DR-reported cause was then used as a main determinant for the VA-assessed specific cause. Other determinants are location, gender, age, and province.

For analysis of application to stroke death data, the binary outcome is VA-assessed ICD-10 code as deaths from stroke or other. The DR-reported ICD-10 codes and location of death were combined and categorized into 18 groups: 9 DR-reported ICD-10 code groups each for the two locations (in or outside the hospitals).

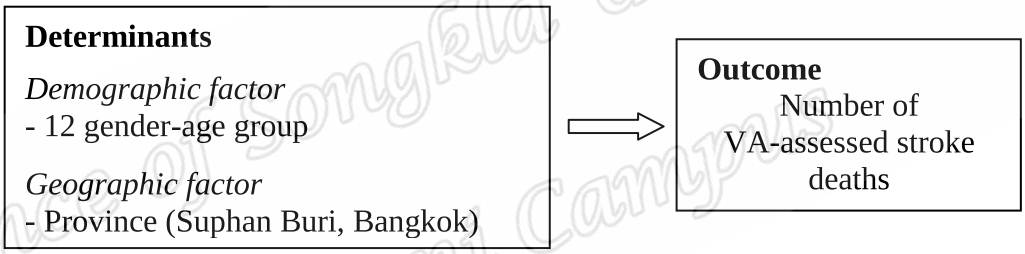
Gender and age were combined and categorized into 6 groups by sex: ages 5-39, 40-49, 50-59, 60-69, 70-79 and 80+years for each sex. Nine provinces (Bangkok, Nakhon Nayok, Suphan Buri, Ubon Ratchathani, Loei,

Phayao, Chiang Rai, Chumphon, and Songkhla) were samples in the VA survey.

Study I: A Comparison of 21 cause-specific deaths in Thailand in 2005 and Japan in 2006



Study II: Statistical methods for comparing stroke death rates in Suphan Buri with Bangkok in 2005



Study III: A Statistical method for correcting misreported multinomial outcome data with application to stroke mortality in Thailand

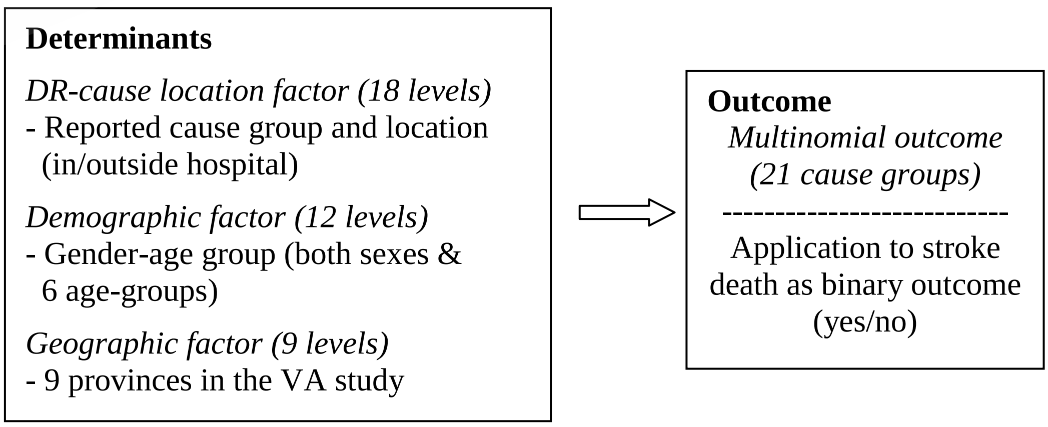


Figure 2.2 Diagram for variables used in three studies

2.3 Analysis methods

2.3.1 Crude death rates and age-specific death rates

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

$$CDR = \frac{Total\ resident\ deaths}{Total\ population} \times 100{,}000 \quad (2.1)$$

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

$$ASDR = \frac{Total\ deaths\ in\ specified\ age\ group}{Total\ population\ in\ the\ same\ specified\ age\ group} \times 100{,}000 \quad (2.2)$$

2.3.2 Age-standardization and SMR

There are a number of methods that can take account of age structure differences when comparing deaths between two different populations. This study focuses on indirect age-standardization, which is a standard set of age-specific death rates are applied to the study population. This method provides the expected number of deaths in a population. The Standardized Mortality Ratio (SMR) was then calculated by dividing the total number of observed deaths by the total number of expected deaths.

The SMR equals 1.0, indicating neither an excess nor a deficit of risk. If the SMR is greater than 1.0, there is said to be "excess deaths" in the study population.

$$SMR = \frac{number\ of\ Observed\ deaths}{number\ of\ Expected\ deaths} \qquad (2.3)$$

2.3.3 Poisson regression analysis

The Poisson regression model is appropriate for modelling count data. The formulation for the Poisson model is based on two set of predictors, if λ_{ij} denotes the mean death rate in province *i* and the gender-age group *j*, *P* is the population at risk per 100,000 for each data cell, an additive model with this distribution has mean λ_{ij}, where

$$\ln(\lambda_{ij} / P_{ij}) = \mu + \alpha_i + \beta_j \qquad (2.4)$$

In this model, α_i and β_j are province and gender-age group, respectively, and μ is a constant. The Poisson model has only the single parameter λ, and the variance is also λ.

2.3.4 Logistic regression model

Logistic regression model is appropriate when the outcome takes one of only two possible values representing presence or absence of an attributes of interest. The model formulated the logit of the probability that a person died from the specific cause of death as an additive linear function of determinants (Hosmer and Lemshow 2002, Venable and Ripley 2002). The simple model when the determinant is a categorical variable is expressed as

$$\ln(p_i /(1 - p_i)) = \mu + \alpha_i \qquad (2.5)$$

Where p_i the probabilities of death due to the specific cause of death, μ is a constant, and α_i is the parameter of DR- cause location group *i*. the simple model was compared with the full model (2.6), which includes an additive linear function of determinant factors. The full model with three categorical determinants was formulated as

$$\log it(p_{ijk}) = \log(p_{ijk}/(1-p_{ijk})) = \mu + \alpha_i + \beta_j + \gamma_k \quad (2.6)$$

Where p_{ijk} is the probabilities of death due to the specific cause of the *i*, *j* and *k* groups of predictor factors, μ is a constant, α_i, β_j and γ_k refer to province, gender-age group, and DR-cause location group, respectively. The selected cause group can be obtained as follows:

$$p_{ijk} = 1/(1+\exp(-(\mu + \alpha_i + \beta_j + \gamma_k))) \quad (2.7)$$

2.4 Graphical methods

For preliminary analysis of population age distribution, we use bar plot. The bar plots were used to explore age distribution for each population to be compared (Thai/Japan or Suphan Buri/Bangkok).

Doughnut plots

We used doughnut plots to display excess deaths by cause and gender. This plot is a useful method for visually comparison data from two populations. It comprises inner and outer coloured circles. The areas of yellow inner circles represent numbers of deaths by cause groups in the country with the lower age-standardized death rate. The outer rings, with areas denoting the difference between inner and outer circles, represent the magnitude of excess deaths in the other country (using red for Thailand

and blue for Japan) for each cause.

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 plots are produced from the standard errors of differences between the mortality rates and its mean.

A Receiver Operating Characteristic (ROC) curve assesses how well a model predicts a binary outcome. Denoting the predicted outcome as 1 (deaths due to selected cause) if $p_{ijk} \geq$ c or 0 if $p_{ijk} <$ c, the ROC curve plots sensitivity against the false positive rate, as *c* varies. The ROC curve passes through the upper left corner, providing area under the curve (AUC) close to 1.

All data analysis were under taken using R program version 2.15.2 including graphical displays.

CHAPTER 3

Results

This chapter contains the preliminary analysis for comparing mortality between populations in three studies.

The first study concerned with a useful graphical method for visually comparison data which is used to display cause-specific excess deaths by gender and was submitted to the *Pertanika Journal of Social Sciences & Humanities.*

The second study is provided on the use of a statistical model for comparing stroke death rates for Suphan Buri with Bangkok in 2005 and was submitted to *the Journal of Research in Health Sciences.*

The third study summarizes methods for correcting misreported multinomial outcome data with application to stroke mortality in Thailand and was submitted to *the Walailak Journal of Science and Technology*.

3.1 Preliminary results

The first study

Table 3.1 shows the distribution of numbers of deaths by cause and gender between Thailand and Japan, which is used for analysis. For comparison, these deaths cannot be compared directly due to population size and age distribution. The larger population will generate more deaths than smaller population.

Thailand's crude death rates in 2005 (652 per 100,000) was less than those in Japan in 2006 (892 per 100,000). Comparison of these rates would be misleading because these rates do not take into account the population age distribution.

Table 3.1 Distribution of numbers of deaths by cause and gender for aged five years or more in Thailand in 2005 (total population 58,972,812) and Japan in 2006 (total population 121,102,000)

Cause groups	Numbers of deaths			
	Thai		Japan	
	male	female	male	female
1: Transport Accident	18475	4335	6206	2762
2: Other Injury	9360	4545	16914	12032
3: Suicide	4954	1232	21419	8502
4: All Other	8726	6317	6808	8013
5: TB	5609	2733	1517	752
6: HIV	14805	8006	55	5
7: Other Infectious	4448	3535	6362	6378
8: Endocrine	8780	15463	10030	9523
9: Mental and Nervous	5751	3510	8978	10320
10: Digestive	12594	7420	23072	19264
11: Genitourinary	7352	8324	12090	15748
12: IHD	13478	11720	41295	34132
13: Stroke	24475	22371	61340	66911
14: Other CVD	9483	10619	53065	67872
15: liver Cancer	14137	6513	22576	11084
16: lung Cancer	8581	3492	47448	17716
17: Other Digestive Cancer	6879	5269	86536	58713
18: Other Cancer	11806	16167	46547	48655
19: Respiratory	20340	10734	89915	72707
20: Septicemia	1172	1699	4204	4563
21: Ill Defined	8275	11206	12323	24964
All-cause	219478	165211	578700	500616

Figure 3.1 shows that both male and female population age distributions in Thailand are different from those in Japan. Women live longer than men, especially in Japan, where 31.3% of females aged five years or more were aged 60 years or more

compared with 25.8% of males, and 7.3% of females aged five years or more were aged 80 years or more compared to only 3.7% of males.

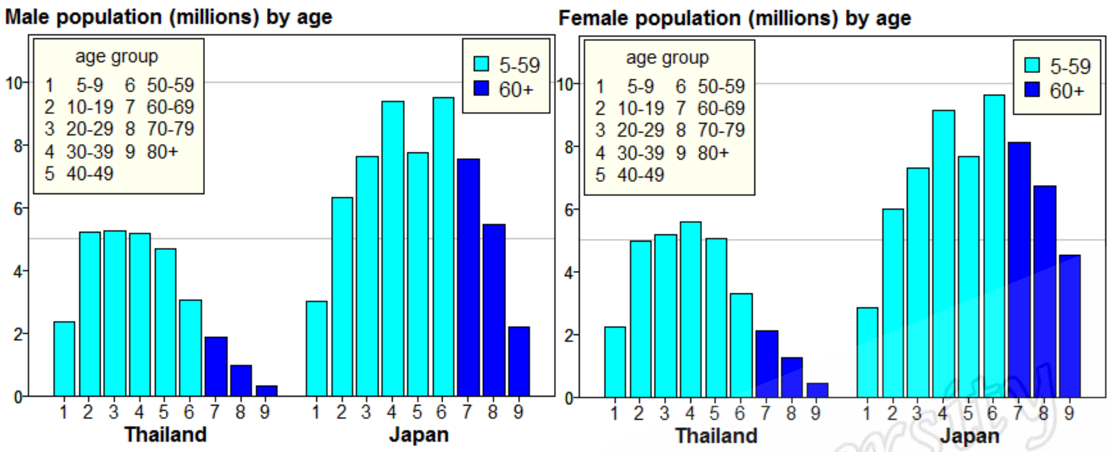


Figure 3.1 Population age distributions (millions) by gender between Thailand (2005) and Japan (2006)

Comparison of 21 cause groups between Thailand and Japan, age-specific death rates and age-standardized deaths for each cause are computed. For simplify, we choose stroke deaths to demonstrate the comparison methods and then apply to the other causes.

Figure 3.2 shows gender age-specific stroke death rates for Thailand and Japan. In both males and females, Thailand has higher age-specific death rate at every age group except age 5-29 years. The excess mortality for Thailand in 2005 over Japan in 2006 is highlighted in magenta shading in the bar strips. For clarity, a cube-root *y* axis scale is used, so that graphs do not accurately depict the extra burden of death in Thailand compared to Japan.

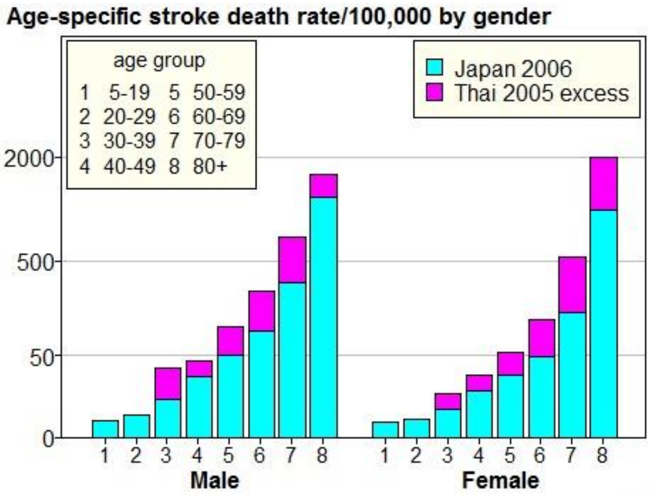


Figure 3.2 Age-specific stroke death rates by gender between Thailand (2005) and Japan (2006) showing excess deaths

Age-standardized mortality comparison

Applying Japan's age-specific stroke death rates to the Thai population, the total number of age-standardized stroke deaths for Japan would be only 21,395 (11,989 for males and 9,406 for females). This is 47% of the number of stroke deaths in Thailand (46,846) in 2005 as shown in Figure 3.3. For gender excess deaths, Thailand has 2.38 times as many female age-standardized stroke deaths in Japan compared with 2.04 times as many male age-standardized stroke deaths.

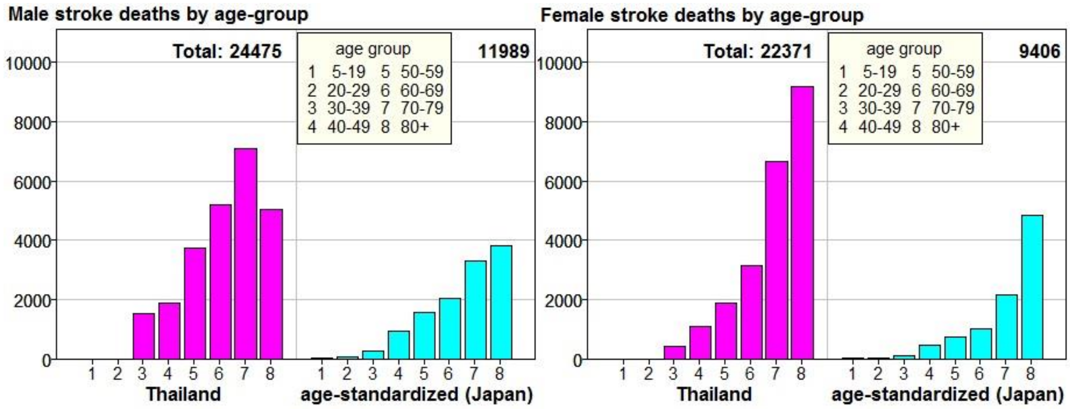


Figure 3.3 Numbers of stroke deaths in Thailand and age-standardized deaths for Japan by gender and age group

Comparison of cause group

After applying Japan's age-specific death rates for each remaining cause groups, deaths by cause were presented using bar plots as shown in Figure 3.4.Deaths caused by injuries, infectious diseases, internal causes (especially female endocrine), stroke, and liver cancer are all much higher in Thailand. In contrast, deaths from suicide and other digestive cancer are higher in Japan.

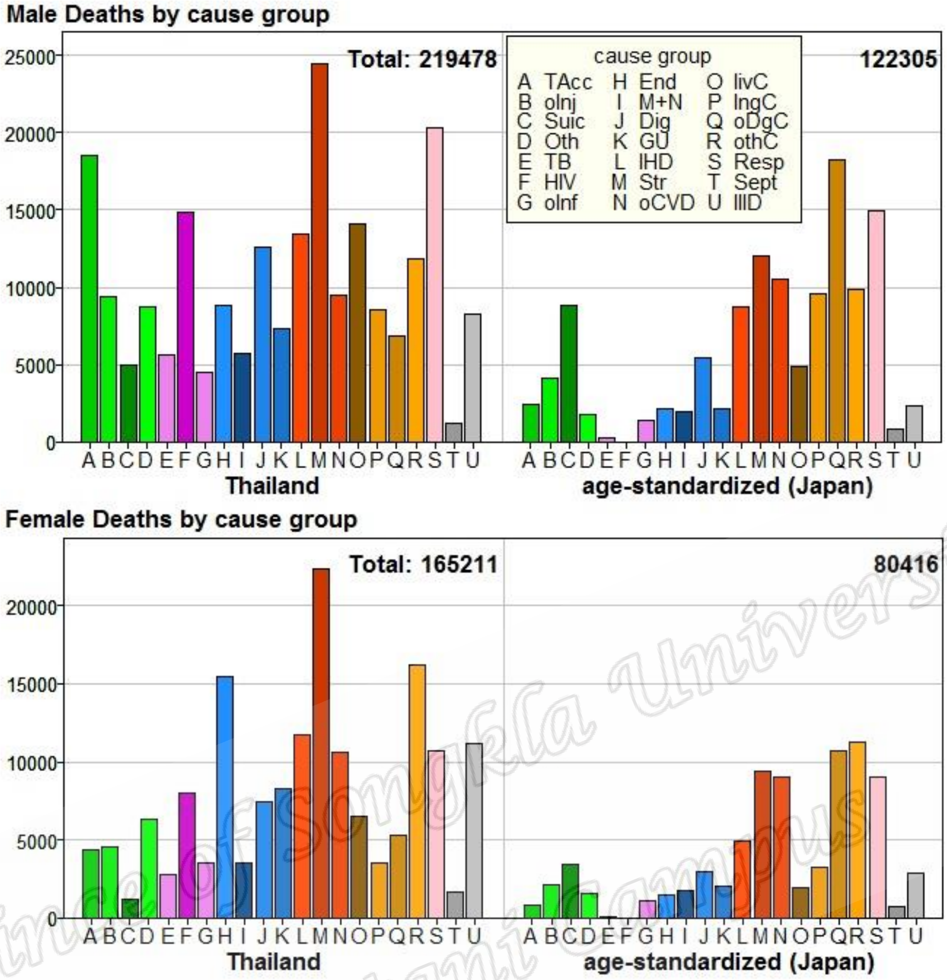


Figure 3.4 Deaths by cause groups in Thailand (2005) and Japan (2006) by gender

The second study

Table 3.2 shows percentage of population age distribution, observed number of stroke deaths (O), age-specific stroke death rates (ASDR) per 100,000 population for Suphan Buri and Bangkok by gender, and age-standardized stroke deaths (E), respectively.

Suphan Buri and Bangkok have different population age distribution. Both males and females, Bangkok has a higher percentages of people aged 30-49 years whereas Suphan Buri has a higher percentages of people aged 60 years or more than Bangkok, especially for females.

For age-specific stroke death rates, Suphan Buri has higher death rates at every age group. However, these excess deaths do not accurately depict the extra burden of deaths in Suphan Buri compared to Bangkok.

Applying Bangkok's age-specific stroke death rates to the Suphan Buri population, the total number of age-standardized stroke deaths (E) for Bangkok would be only 929 (483 for males and 446 for females). This is 83% of the stroke deaths in Suphan Buri (1,119) in 2005.

Table 3.2 Percentage of population age distribution, observed numbers of stroke deaths (O), age-specific stroke death rates (ASDR) per 100,000 population for Suphan Buri (SP) and Bangkok (BKK) by gender, and age-standardized stroke deaths (E), respectively

Gender	Population (%)		O		ASDR/100,000		E
age group	SP	BKK	SP	BKK	SP	BKK	
m.30-39	28.0	37.1	15	86	24	13	8
m.40-49	28.3	29.4	44	256	68	50	32
m.50-59	19.8	18.2	77	454	171	143	64
m.60-69	13.6	9.4	110	478	355	291	90
m.70-79	7.6	4.5	194	781	1118	1000	174
m.80+	2.6	1.4	119	468	2022	1952	115
f.30-39	27.1	36.3	6	30	9	4	3
f.40-49	25.2	28.9	24	136	37	23	15
f.50-59	20.5	17.8	44	199	84	54	29
f.60-69	14.4	9.7	43	207	117	105	38
f.70-79	9.1	5.3	193	716	836	660	152
f.80+	3.7	2.0	250	913	2674	2235	209

The third study

Figure 3.5 shows that the unscaled VA-estimated and DR-reported deaths in 2005 for each 21 cause group. Using separate logistic models to 2005 VA data gives the totals VA-estimated deaths (379,165 cases) slightly less than those reported deaths (384,689 cases). The degrees of misclassification of causes of death were observed for example, septicaemia is over-reported but stroke and others are under-reported.

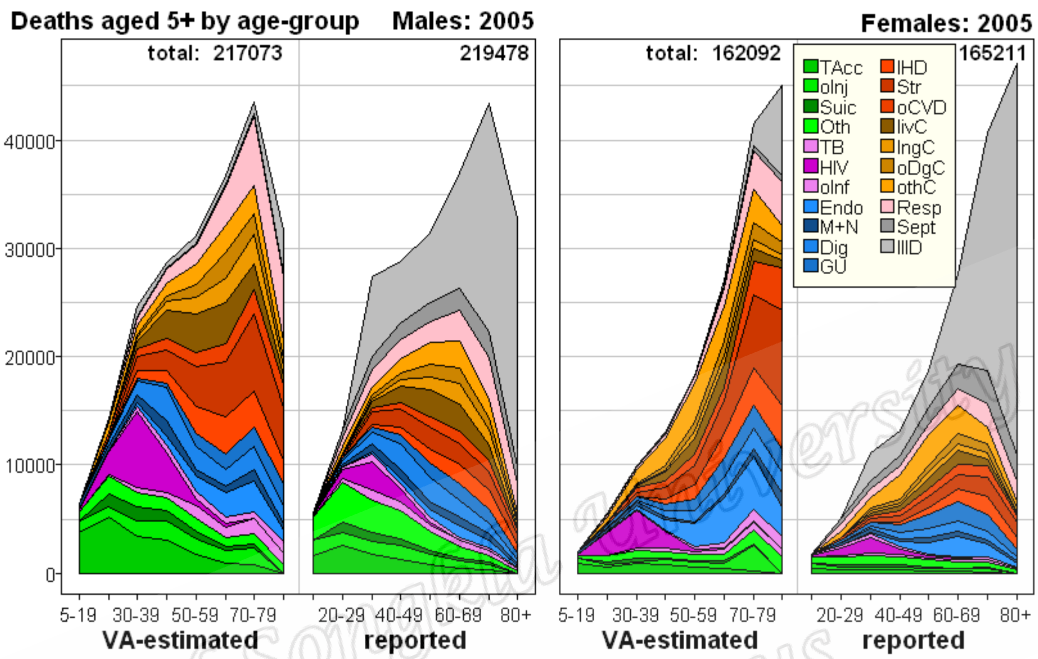


Figure 3.5 Deaths by cause groups and gender in 2005

3.2 Manuscripts

3.2.1 A Comparison of cause-specific deaths in Thailand in 2005 and Japan in 2006

Manuscript was submitted to the *Pertanika Journal of Social Sciences & Humanities*

Arinda Ma-a-lee[1], Phattrawan Tongkumchum[1], and Attachai Ueranantasan[1]

[1]Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Thailand

Abstract

This study compared mortality aged five years or more between Thailand in 2005 and Japan in 2006. Deaths from the Thai death registration (DR) database were adjusted for misclassification using logistic regression based on the verbal autopsy (VA) data. Thailand and Japan have different population age distribution so age standardization rates was used to make the comparison possible by applying gender-age specific death rates of Japan to the Thai population. The total number of age-standardized deaths for Japan was 203,133 that is, 53% of the number of deaths in Thailand in 2005 (384,689). Gender excess deaths were observed with Thailand having 2.05 times as many female deaths as Japan, compared with 1.79 times as many male deaths. In Thailand, there were cause-specific excess deaths for transport accident for males, infectious diseases, endocrine for females, stroke and liver cancer. In contrast, there were cause-specific excess deaths for causes related to suicide and other digestive cancer for Japan. The more cause-specific excess deaths in Thailand impose a significant public health burden that demands attention and action by health policy

makers.

Keywords: age-standardized deaths, causes of death, Japan**,** Thailand

INTRODUCTION

According to the World Health Organization statistics (WHO, 2004a) Thailand ranked 77th among 192 countries, with age-standardized death rate double that of Japan, which is the lowest age-standardized death rate among all countries in the world. During 2008, the age-standardized death rate for Thailand (935 deaths per 100,000 population) exceeded that of Japan (349 deaths per 100,000 population) (OECD/WHO, 2012). In 2011, Japan life expectancy at birth of 79 years for males and 86 for females whereas in Thailand life expectancy at birth of 71 years for males and 77 for females (WHO, 2013). These gaps are unaccepted and can be improved.

The difference in life expectancy between two countries corresponds to males and females in Thailand dying 8 and 9 years earlier, respectively, than those in Japan. It is important to understand why this is so. Moreover, comparison of cause-specific deaths between countries is crucial for improving human health and reducing preventable deaths in developing countries.

Comparison of mortality between two populations with different age distributions requires a method known as age-standardization (Naing, 2000; Torrey & Haub, 2004; Robson *et al.*, 2007). The method involves comparing the mortality rate in each gender-age group and combining results to obtain single ratio.

However, comparison of cause-specific mortality requires good quality data. Mortality studies in Thailand raised national concern to improve death registry (DR) data (Rao *et al.*, 2010; Porapakkham *et al.*, 2010). Cause-specific mortality in Thailand is misreported (Mathers *et al.*, 2005). Forty percent of death certificates give the causes of death as ill-defined, which severely limits their public utility. In contrast, less than 4% of deaths in Japan are ill-defined (Mathers *et al.*, 2005; Carmichael, 2011).

Verbal autopsy (VA) is widely used to verify cause of deaths (Setel *et al.*, 2005; Ngo *et al.*, 2010; Rao *et al.*, 2010) in order to get more accurate causes. The latest VA survey in Thailand was carried out in 2005. Mortality based on this study reported that the five leading causes of death among males are stroke, transport accidents, HIV/AIDS, ischemic heart diseases, and chronic obstructive lung diseases. Among females, the leading causes of death are stroke, diabetes, ischemic heart diseases, HIV/AIDS and renal diseases (Rao *et al.*, 2010).

The objective of this study was to assess all cause and cause-specific deaths in Thailand in 2005 compared to Japan in 2006. Before comparisons we adjusted the Thai data in cause groups for misreporting using data from the 2005 VA study.

MATERIALS AND METHODS

Data sources and management

Death data by cause, gender-age group for ages five years or more in Thailand in 2005 (Thai data) and Japan in 2006 (Japan data) were used for comparison. Causes of deaths for ages less than five years have specific causes including pneumonia, prematurity, birth asphyxia and congenital, which are different from causes of death for adults (OECD/WHO, 2012). Therefore, under five year deaths were separated for special attention.

Thai data

The 2005 verbal autopsy (VA) and Thai death registration (DR) data were obtained from the Bureau of Health Policy and Strategy, Ministry of Public Health.

The DR data contains information on province, gender, age in years, cause of death reported on the death certificate, and the location of death (in hospital or outside hospital) of each deceased person. The 10th revision of the International Classification of Diseases (ICD-10) were used for causes of deaths (WHO, 2004b).

The VA data contain information of deceased persons for province of residence, gender, age group, location (death occurrence in or outside hospital), DR reported ICD-10 code and VA-assessed ICD-10 code. It comprised 9,644 (3,316 in-hospital and 6,328 outside-hospital) death records. The VA assessed ICD-10 codes were more reliable than the DR reported ICD-10 codes. The VA-assessed deaths by cause were the basis analysis. Study design, methodologies, and data analysis of the VA study have been published elsewhere (Rao *et al.*, 2010; Porapakkham *et al.*, 2010).

Without appropriate the classification, the DR data in Thailand cannot be compared to Japan data directly due to misclassification of specific causes of death. To allocate causes of death to their correct ICD chapter-block groups, we used the method described in detailed by Chutinantakul *et al.* (2014).

Definition of causes of death

The VA assessed ICD-10 codes were grouped into 21 major causes based on the chapter-block classification of ICD-10 codes (WHO, 2004b). For statistical accuracy, groups need around 200 cases, with septicemia (over-reported) excepted for special attention. The ICD-10 codes with small numbers of deaths were grouped into an “All other” group. Table 1 shows these cause groups and numbers of deaths aged five years or more (9,495 deaths) for the VA data.

Table 1 Definition of 21 major cause groups with ICD-10 codes

Cause groups	Count	Cause groups	Count
1: Transport Accident (V)	536	11: Genitourinary (N)	412
2: Other Injury (W, X0-59)	327	12: Ischemic (I20-25)	617
3: Suicide (X60-84)	158	13: Stroke (I60-69)	1076
4: All other	358	14: Other CVD (I$^-$)	540
5: TB (A15-19)	195	15: Liver Cancer (C22)	500
6: HIV (B20-24)	512	16: Lung Cancer$^+$(C30-39)	320
7: Other Infectious (A, B)$^-$	219	17: Other Digestive (C15-26$^-$)	290
8: Endocrine (E)	647	18: Other Cancer (C$^-$, D0-48)	697
9: Mental, Nervous (F, G)	223	19: Respiratory (J)	801
10: Digestive (K)	489	20: Septicemia (A40-41)	77
		21: Ill-defined (R)	501

$^+$ Respiratory/thoracic, $^-$ exclude above

Numbers of deaths by cause adjusted for misclassification

Numbers of deaths from the Thai DR database were adjusted for misclassification using logistic regression based on the VA data. Each cause group was considered as a binary outcome (deaths from this cause or other). Different models were fitted separately for each cause group with province, demographic and medical factors.

The province factor has nine levels corresponding to the nine provinces in the VA study. Gender and age group were combined with number of levels depending on age distribution of deaths from the selected cause group outcome. Cross tabulation of the 21 VA and DR cause groups suggested the reported cause groups that were used as determinants. The interaction term between reported cause group and location of death (in or outside hospital) has number of levels depending on the reported cause groups that affect the selected cause outcome. The models gave estimated probabilities of deaths for each cause group for specified determinants. The goodness of fit of the models was assessed using a Receiver Operating Characteristic (ROC) curve. The ROC plots sensitivity against false positive rate (1-specificity) (Fan *et al.*, 2006; Sakar & Midi, 2010). It shows how well a model predicts a binary outcome.

For each cause group, the model gives coefficients for the nine provinces sampled in the VA study. To use the model to predict the coefficient of the remaining provinces of Thailand, we used a spatial triangulation method based on the latitude and longitude of the province. The remaining coefficients in each triangle were obtained by solving three linear equations. The estimated probabilities from the model for each cause group were applied to the DR data in 2005 giving numbers of VA-assessed deaths for each cause. Thus, total unscaled VA-assessed deaths in 2005, which are slightly less than those reported, were obtained.

Scaled VA-assessed deaths by causes

The unscaled VA-assessed deaths were categorized into eight age groups (5-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, and 80+) for every cause group. The

VA-assessed deaths counts were scaled to match the DR reported counts using the formula

$$vas_{ijk} = va_{ijk} \times \left(\frac{drTot_{jk}}{vaTot_{jk}} \right),$$

In this formula vas_{ijk} and va_{ijk} are the scaled and unscaled VA-assessed deaths classified by cause group *i*, gender *j*, and age-group *k* ($i = 1, 2 \ldots, 21$, $j = 1, 2$ and $k = 1, 2, \ldots, 8$). The factors $drTot_{jk}$ and $vaTot_{jk}$ are numbers of deaths from all causes for DR reported and for VA-assessed deaths by gender *j* and age-group *k*. Thus aggregated scaled VA-assessed deaths by gender-age group for each 21 cause groups were obtained.

The 2005 projected populations for Thailand by gender and age group were used. The projection method is based on assumptions about mortality and female fertility.

Japan data

The numbers of deaths for Japan in 2006 by gender, five year age group and ICD-10 codes, with corresponding population at risk, are available from the World Health Organization (WHO, 2011).

The ICD-10 codes were grouped into 21 major causes based on the chapter block classification of ICD-10 codes (WHO, 2004b). The number of deaths were categorized in to eight age-groups (5-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, and 80+) for each 21 cause group.

Comparison analysis with Japan in 2006

Age distributions of Thai populations for each gender were compared to Japan using bar plots. To explore deaths in Thailand and Japan, gender age-specific death rates were computed as the numbers of deaths classified by gender and age group divided by the corresponding population at risk.

To compare numbers of deaths in populations with different age distributions, age-standardized rates were used for such comparison by matching the Japan population to the Thai population. To do this, we applied Japan's gender age-specific death rates for each major cause group to the Thai population using the formula

$$n'_{ijk} = \left(\frac{n_{ijk}}{PJ_{jk}} \right) \times PT_{jk} .$$

In this formula n'_{ijk} are numbers of age-standardized deaths by cause group *i*, gender *j*, and age-group *k* and n_{ijk} are numbers of observed deaths in Japan. PJ_{jk} and PT_{jk} are the corresponding population sizes for Japan and Thailand, respectively. Thus numbers of age-standardized deaths for Thailand were obtained.

Gender excess deaths were calculated as the ratio of the numbers of VA assessed deaths in Thailand to the numbers of age-standardized deaths in Japan separated by gender.

Cause-specific excess deaths were calculated as the difference between the numbers of VA assessed deaths in Thailand to the numbers of age-standardized deaths in Japan separated by gender.

Doughnut plot

Doughnut plots were used to display results of cause-specific excess deaths. This plot is a useful method for visually comparison data from two countries or regions. It comprises inner and outer coloured circles. The areas of yellow inner circles represent numbers of deaths by cause groups in the country with the lower age-standardized death rate. The outer rings, with areas denoting the difference between inner and outer circles, represent the magnitude of excess deaths in the other country (using red for Thailand and blue for Japan) for each cause.

RESULTS

Preliminary results

Figure 1 illustrates the differences in overall population structure for ages five years or more. Thailand and Japan have different age distributions. Japan has a much greater proportion of persons aged 60 years and older (29%) than Thailand (12%) as shown in Figure 1.

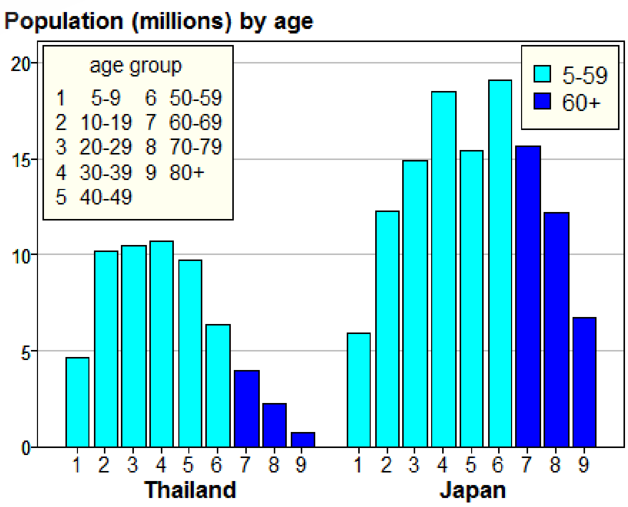


Fig 1 Population age distributions (millions) for Thailand (2005) and Japan (2006). Note that the first age group (5-9) is half the width of the others to ensure that all

areas of bars accurately denote populations and the vertical scale denotes population density.

Fig 2 shows that both males and females population age distributions in Thailand are different from those in Japan. Women live longer than men, especially in Japan, where 31.3% of females were aged 60 years or more compared with 25.8% of males, and 7.3% of females were aged 80 years or more compared to only 3.7% of males.

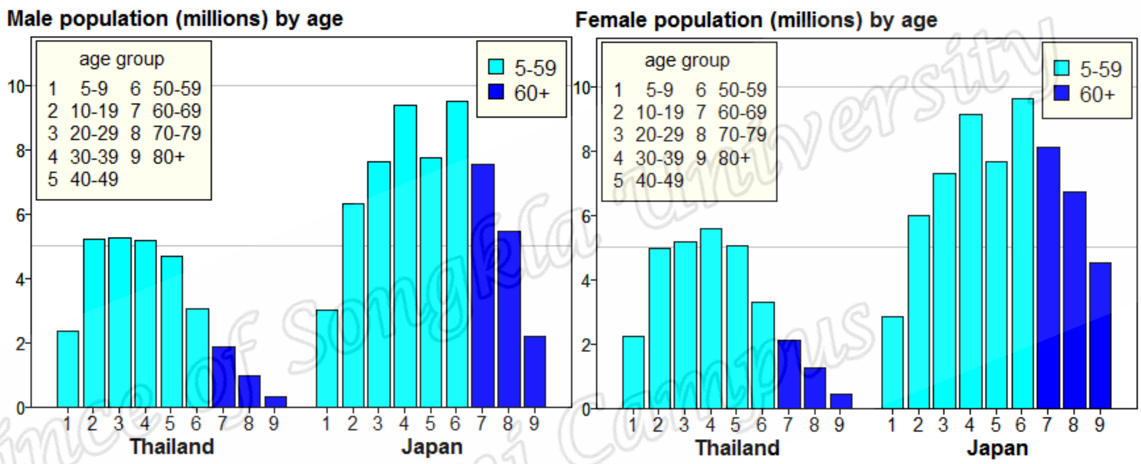


Fig 2 Population age distributions (millions) by gender between Thailand (2005) and Japan (2006)

Fig 3 shows gender age-specific death rates for Thailand in 2005 and Japan in 2006. Japan has lower age-specific death rate at every age group. The excess mortality for Thailand in 2005 over Japan in 2006 is highlighted in blue shading in the bar strips. For clarity, a cube-root *y* axis scale is used, so that graphs do not accurately depict the extra burden of death in Thailand compared to Japan.

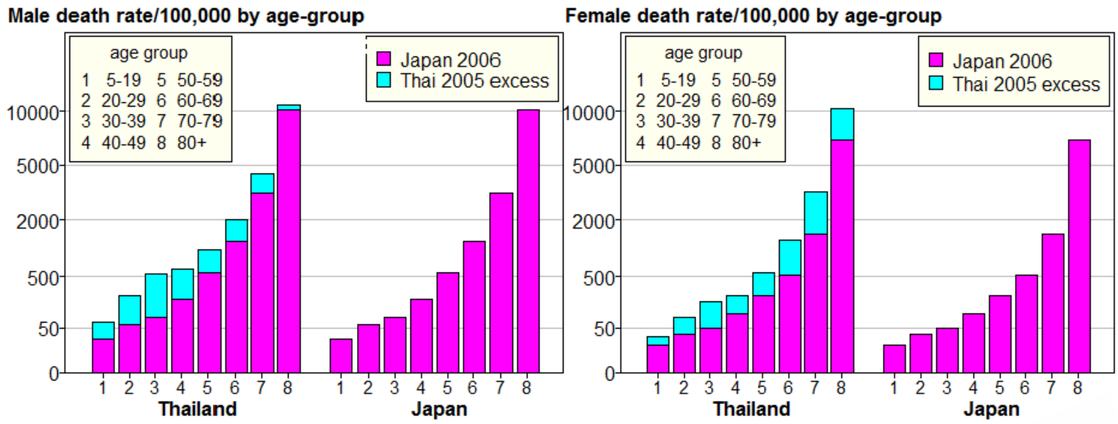


Fig 3 Age-specific death rates by gender between Thailand (2005) and Japan (2006) showing excess deaths

Age-standardized mortality comparison

Applying Japan's gender age-specific death rates to the Thai population, the total number of age-standardized deaths for Japan would be only 203,133 (122,496 for males and 80,637 for females). This is 53% of the number of deaths in Thailand (384,689) in 2005 as shown in Fig 4. For gender excess deaths, Thailand has 2.05 times as many female age-standardized deaths in Japan compared with 1.79 times as many male age-standardized deaths.

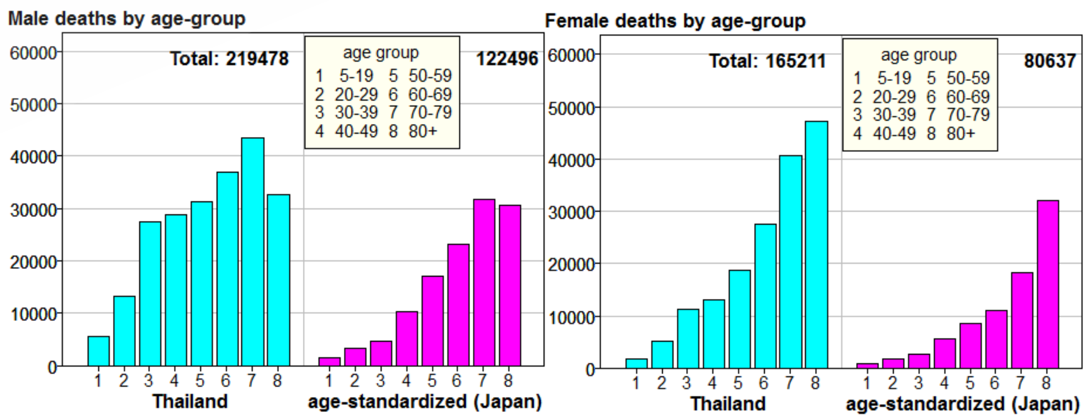


Fig 4 Overall numbers of deaths in Thailand and numbers of age-standardized deaths for Japan by gender and age group

Table 2 shows scaled VA-assessed deaths for Thailand in 2005 and age-standardized deaths for Japan in 2006 by cause group and gender. Age-standardized deaths for Japan were computed after excluding causes with OPQ ICD-10 codes that are causes for child mortality. The total number of age-standardized deaths for Japan is only 202,721 (122,305 for males and 80,416 for females). Deaths caused by injuries, infectious diseases, internal causes (especially female endocrine), stroke and liver cancer are all much higher in Thailand, whereas deaths from other digestive cancer (mainly stomach cancer), and suicide are higher in Japan.

Table 2 Scaled numbers of deaths in Thailand in 2005 and age-standardized deaths in Japan in 2006 by 21 major cause groups and gender

Cause groups	Male		Female	
	Thai	Japan	Thai	Japan
Transport Accident	18475	2408	4335	796
Other Injury	9360	4117	4545	2143
Suicide	4954	8798	1232	3432
All Other	8726	1734	6317	1588
TB	5609	298	2733	108
HIV	14805	22	8006	1
Other Infectious	4448	1372	3535	1076
Endocrine	8780	2183	15463	1495
Mental and Nervous	5751	1998	3510	1745
Digestive	12594	5395	7420	2954
Genitourinary	7352	2123	8324	2026
IHD	13478	8684	11720	4905
Stroke	24475	11989	22371	9406
Other CVD	9483	10519	10619	9035
liver Cancer	14137	4872	6513	1943
lung Cancer	8581	9542	3492	3243
Other Digestive Cancer	6879	18244	5269	10674
Other Cancer	11806	9864	16167	11310
Respiratory	20340	14933	10734	8996
Septicemia	1172	837	1699	692
Ill Defined	8275	2373	11206	2848
Total	219478	122305	165211	80416

The doughnut plots in Fig 5 clearly show cause-specific excess deaths. Larger numbers of Thai excess deaths for both sexes were observed for transport accidents, infectious diseases, internal causes (especially female endocrine), stroke and liver cancer. In contrast, excess deaths were observed for suicide and other digestive cancers in Japan.

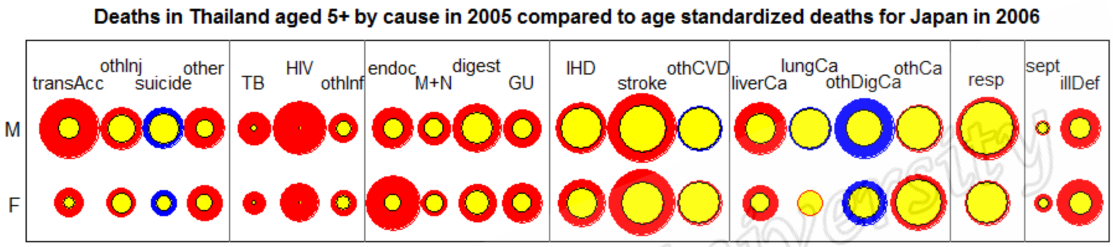


Fig 5 Doughnut plots with deaths in Thailand in 2005 by cause group and gender compared to age-standardized deaths for Japan in 2006**,** the outer rings represent the magnitude of excess deaths in the other country (using red for Thailand and blue for Japan) for each cause.

DISCUSSION

This study compared mortality aged five years or more between Thailand in 2005 and Japan in 2006. Thailand and Japan have different population age-distribution so an age-standardization rate was used to make the comparisons possible by applying gender-age specific death rates from Japan to the Thai population. The gender excess deaths for all-cause were computed and cause-specific excess deaths were presented using a doughnut plot. The total number of age-standardized deaths for Japan was 203,133 cases, that is, 53% of the number of deaths in Thailand in 2005 (384,689). Gender excess deaths were observed with Thailand having 2.05 times as many female deaths as Japan, compared with 1.79 times as many male deaths. In

Thailand, there were excess deaths for causes related to injuries (especially, transport accident for male), infectious diseases, internal causes (especially endocrine for female), stroke and liver cancer. In contrast, there were excess deaths for causes related to suicide and other digestive cancer for Japan.

Thailand suffered 47% more deaths (181,556 deaths) than would be expected if they had experienced the same death rates as Japan. This is in agreement with WHO reports which give age-standardized death rates for Thailand double that of Japan (OECD/WHO, 2012). This could be due to the health disparity between Thailand and Japan (Yiengprugsawan *et al.*, 2010; OECD/WHO, 2012).

Thailand is a middle income country, socio-economic inequality exist in Thailand that may effect to deficiency of access to health care services, for example poor people have further to see doctor and that after adjusting for different levels of needs, the rich are more likely to see a specialist and high quality of care than the poor. Although, Thailand adopted a Universal Coverage Scheme (USC) in 2001-2002, the USC employs a minimal co-payment of 30 Baht (0.75 USD) per hospital visit (Naranong & Naranong, 2006; Prakongsai *et al.*, 2002). In 2006, the government abolished the co-payment and made it free. However, poor people still use health services less when they ill (Yiengprugsawan *et al.*, 2010). Moreover, most of health care services and high quality of care in Thailand are mainly located in Bangkok and others major cities.

In contrasts, Japan is a high income country and has advantages in terms of equality of socio-economic condition that may lead to more equal access to health care services and high quality of care than Thailand. Since, Japan implemented universal health care in 1961 with compulsory social insurance schemes, and achieved a high level of performance. Equal access to health care is promoted by relatively high share of public health spending and fairly equal geographic distribution of doctors. In addition, the quality of health care is good in many areas, including cancer care and stroke (OECD/WHO, 2012; Fukuda *et al.*, 2004).

A large number of excess injury deaths especially transport accident for males were also consistent with WHO reports on mortality rates of 106 deaths per 100,000 injuries mortality rates for Thailand and 36 for Japan. Most injury deaths were self inflected in Japan whereas in Thailand 40% or more were due to transport accidents (OECD/WHO, 2012). This might be partly due to most of motorcyclists not following motorcycle helmet laws; only 32% of motorcyclists wore helmets during the daytime and 9% at night, and only 30% of helmets comply with the Thai standard (Tanaboriboon & Satiennam, 2005; Hyder *et al.*, 2007).

Over the past three decades, Thailand has faced social and economic transitions. These shifts have changed in Thai society through globalization and rapid economic growth, as well as the development of the Thai health care system. These changes are evidenced by an increased in total life expectancy at birth from 67 to 74 years from 1990 to 2011, (WHO, 2013; Kaufman *et al.*, 2011) infant mortality rate (IMR per 1000 live births) decline from 84.3 to 11.3, and maternal mortality rate (MMR per 1000 live births) decline from 317.6 to 9.8 (Bureau of Policy and Strategy, 2008).

However, excess deaths for infectious diseases in this study suggest that they are still high in Thailand. This finding is consistent with studies reporting that although deaths from HIV/AIDS and TB declined during 2004-2009; they were still a major health problem and responsible for a considerable number of deaths in Thailand (Chasombat *et al.*, 2006; Aungkulanon *et al.*, 2012). This decline may account for the concurrent leveling-off of infectious disease related mortality.

There were a larger number of excess deaths for causes related to non-communicable diseases (especially, stroke, liver cancer, female endocrine). This result is consistent with studies reporting that currently, death rates in Thailand from non-communicable diseases have increased dramatically (Kaufman *et al.*, 2011). Rao *et al.* (2010) reported that the leading cause of deaths in Thailand among males and females is stroke. This might be linked to changing in Thai lifestyle, along with the development of the national economy. Significant changes in diet, particularly increased consumption of fat and sugar, along with reduction in physical activity, have changed the disease pattern in Thailand (Kaufman *et al.*, 2011; Kosulwat, 2002). Burden of disease in Thailand is shifting increasingly to non-communicable diseases such as heart disease and stroke, cancer, chronic obstructive pulmonary disease and diabetes, as well as mental health disorders (Hwang *et al.*, 2012). Liver cancer is also a significant concern in Thailand. Age-standardized liver cancer mortality in Thailand in 2004 was 31.0 per 100,000 whereas it was 13.0 for Japan (WHO, 2008).

In Japan, there were excess suicide deaths. This finding is consistent with previous studies. For suicide mortality, Japan is ranked second highest among OECD (Organization for Economic Co-operation Development) countries, and the eight largest in the world (Chang *et al.*, 2009; Naranong & Naranong, 2006). Japan also had excess mortality over Thailand for other digestive cancers. This finding is consistent with a study reporting that stomach cancer remains a leading cause of death in Japanese population (Qui *et al.*, 2009; Pham *et al.*, 2010).

This study has some strength. The doughnut plots clearly show deaths comparison between Thailand in 2005 and Japan in 2006. This useful method can be applied to other available mortality data for comparison from two countries or regions.

However, the doughnut plot is lack of information of age group and year which may be useful for getting additional information. Further studies need to include these factors to assess the burden of death by region in Thailand compared to Japan using doughnut plot.

CONCLUSION

This study reveals that if the gender-age specific death rate in Thailand were the same as in Japan, the total number of deaths would be only 53% of the number of deaths in Thailand in 2005. Thus, the cause-specific excess deaths remain in Thailand, and this imposes a significant public health burden that demands attention and action by health policy-makers.

ACKNOWLEDGEMENTS

This research was supported by grant funds from the program Strategic Scholarships Fellowships Frontier Research Networks (Specific for Southern region) for the Ph.D. Program Thai Doctoral degree from Office of the Higher Education Commission, Thailand. We would like to thank the Ministry of Public Health and Dr.

Kanitta Bundhamcharoen for providing us data. We also would like to express our thanks to Prof Don McNeil for assistance and helpful guidance.

REFERENCES

Aungkulanon, S., McCarron, M., Lertiendumrong, J., Olsen, S. J., & Bundhamcharoen, K. (2012). Infectious disease mortality rates, Thailand, 1958-2009. *Emerging Infectious Diseases*, *18*(11), 1794-1801.

Bureau of Policy and Strategy. (2008). *Thailand Health Profile 2005-2007*. Bangkok: Ministry of Public Health of Thailand.

Carmichael, G. (2011). Exploring Thailand's mortality transition with the aid of life tables. *Asia Pacific Viewpoints*, *52*(1), 85-105.

Chang, S. S., Gunnell, D., Sterne, J. A., Lu, T. H., & Chenq, A. T. (2009). Was the economic crisis 1997-1998 responsible for rising suicide rate in East/Southeast Asia? A time-trend analyses for Japan, HongKong, SouthKorea, Taiwan, Singapore and Thailand. *Social Science & Medicine*, *68*, 1322-1331.

Chasombat, S., Lertpiriyasuwat, C., Thanprasertsuk, S., Suebsaeng, L., & Lo, Y. R. (2006). National access to antiretroviral program for PHA (NAPHA) in Thailand. *The Southeast Asian Journal of Tropical Medicine and Public Health*, *37*, 704-715.

Chutinantakul, A., Tongkumchum, P., Bundhacharoen, K., & Chongsuvivatwong, V. (2014). Correcting and Estimating HIV mortality in Thailand based on 2005 verbal autopsy data focusing on demographic factors, 1996-2009. *Population Health Metrics*, *12*(25). http://dx.doi.org/10.1186/s12963-014-0025-x

Fan, J., Upadhye, S., & Worster, A. (2006). Understanding receiver operating characteristic (ROC) curves. *Canadian Journal of Emergency Medicine*, *8*(1), 19-20.

Hwang, G. S., Choi, J. W., Choi, S. H., Lee, S. G., Kim, K. H., Cho, Y. M., & Yoon, C. (2012). Effects of tailored health promotion program to reduce cardiovascular disease risk factors among middle-aged and advanced-aged bus drivers. *Asia Pacific Journal of Public Health, 24*(1), 117-127.

Hyder, A. A., Waters, H., Phillips, T., & Rehwinkel, J. (2007). Exploring the economics of motorcycle helmet laws-implications for low and middle-income countries. *Asia Pacific Journal of Public Health*, *19*(2), 16-22.

Kaufman, N. D., Chasombat, S., Tanomsingh, S., Rajataramya, B., & Potempa, K. (2011). Public Health in Thailand: emerging focus on non-communicable diseases. *The International Journal of Health Planning and Management*, *26*, e197-e212.

Kosulwat, V. (2002). The nutrition and health transition in Thailand. *Public Health Nutrition*, *5*(1A), 183-189.

Mathers, C. D., Fat, D. M., Inoue, M., Rao, C., & Lopez, A. D. (2005). Counting the dead and what they die from: an assessment of the global status of cause of death data. *Bulletin of the World Health Organization*, *83*(3), 171-177.

Naing, N. N. (2000). Easy way to learn standardization: direct and indirect methods. *Malaysian Journal of Medical Sciences, 7*(1), 10-15.

Naranong, V., & Naranong, A. (2006). Universal Health Care Coverage: Impacts of the 30-Bath Health-Care Scheme on the Poor in Thailand. *Thailand Development Research Institute*, *21*, 3-10.

Ngo, A. D., Rao, C., Hoa, N. P., Adair, T., & Chuc, N. T. (2010). Mortality patterns in Vietnam, 2006: Findings from a national verbal autopsy survey. *BioMed Central*, *3*(78). http://dx.doi.org/10.1186/1756-0500-3-78.

OECD/World Health Organization. (2012). *Health at a Glance: Asia/Pacific 2012*. http://dx.doi.org/10.1787/97892 6418 3902-en

Pham, T. M., Fujino, Y., Kikuchi, S., Tamakoshi, A., Matsuda, S., & Yoshimura, T. (2010). Dietary Patterns and Risk of Stomach Cancer Mortality: The Japan Collaborative Cohort Study. *Annals of Epidemiology*, *20*, 356-363.

Porapakkham, Y., Rao, C., Pattaraarchachai, J., Polprasert, W., Vos, T., & lopez, A. D. (2010). Estimated causes of death in Thailand, 2005: implications for health policy. *Population Health Metrics*, *8*(14). http://dx.doi.org/10.1186/1478-7954-8-14

Prakongsai, P., Patcharanarumol, W., Tisayaticom, K., & Tangcharoensathien, V. (2002). Capitation rate of the Universal Health Care coverage for the fiscal year 2546 (2003AD). *Journal of Health Science* [in Thai], *11*(5), 599-613.

Qui, D., Katanoda, K., Mrugame, T., & Sobue, T. (2009). Join point regression analysis of long-term trends in cancer mortality in Japan (1958-2004). *International Journal of Cancer*, *124*, 443-448.

Rao, C., Porapakkham, Y., Pattaraarchachai, J., Polprasert, W., Swampunyalert, N., & Lopez, A. D. (2010). Verifying causes of death in Thailand: rationale and methods for empirical investigation. *Population Health Metrics*, *8*(11). http://dx.doi.org/ 10.1186/1478-7954-8-11

Robson, B., Purdie, G., Cram, F., & Simmonds, S. (2007). Age standardization-an indigenous standard?. *Emerging Themes in Epidemiology*, *4*(3). http://dx.doi.org/ 10.1186/1742-7622-4-3

Sakar, S., & Midi, H. (2010). Importance of assessing the model adequacy of binary logistic regression. *Journal of Applied Sciences*, *10*(6), 479-486.

Setel, P. W., Sankoh, O., Velkoff, V. A., Mathers, C., Gonghuan, Y., Hemed, Y., Jha, P., & Lopez, A. D. (2005). Sample registration of vital events with verbal autopsy: A renewed commitment to measuring and monitoring vital statistics. *Bulletin of the World Health Organization*, *83*, 611-617.

Tanaboriboon, Y., & Satiennam, T. (2005). Traffic accidents in Thailand. *International Association of Traffic and Safety Sciences*, *29*, 88-100.

Torrey, B. B., & Haub, C. (2004). A comparison of US and Canada mortality in 1998. *Population and Development Review*, *30*(3), 519-530.

World Health Organization. (2004a). *Health statistics and information systems: Disease and injury estimates, 2000-2002. Death and DALY estimates for 2002 by cause for WHO Member States.* Geneva, Switzerland: World Health Organization. Retrieved from http://www.who.int/healthinfo/global_ burden_disease/ estimates_2000_2002/en/

World Health Organization. (2004b). *Special tabulation lists for mortality and morbidity: mortality tabulation list 1. In International Statistical Classification of Diseases and Health Related to Problems-Tenth Revision (Vol 1) (p. 1163-1175).* Geneva, Switzerland: World Health Organization.

World Health Organization. (2008). *The Global Burden of Disease 2004 update.* Geneva, Switzeland: World Health Organization.

World Health Organization. (2011*). Health statistics and information systems. Download the raw data files of the WHO mortality database.* Geneva, Switzerland: World Health Organization. Retrieved from http://www.who.int/healthinfo/statistics/mortality_rawdata/en/

World Health Organization. (2013). *World Health Statistics 2013*. Geneva, Switzerland: World Health Organization.

Yiengprugsawan, V., Carmichael, G. A., Lim, L. L-Y., Seubsman, S., & Sleigh, A. C. (2010). Has universal health insurance reduced socioeconomic inequalities in urban and rural health service use in Thailand?. *Health & Place*, *16*, 1030-1037.

3.2.2 Statistical methods for comparing stroke death rates in Suphan Buri with Bangkok in 2005

Manuscript was submitted to the *Journal of Research in Health Sciences.*

Arinda Ma-a-lee[1], Phattrawan Tongkumchum[1], and Nattakit Pipatjaturon[2]

[1]Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Thailand

[2]The office of Diseases Prevention and control 2th Phitsanulok, Phitsanulok, Thailand

Abstract

Background: Stroke is a major health problem in Thailand. Several studies compared stroke deaths between populations and always provide the results based on Standardized Mortality Ratio (SMR). This study aims to compare stroke death rates between two populations using Poisson regression model. The model is directly provide 95% CI and p-value.

Methods: The stroke deaths for Suphan Buri and Bangkok in 2005 were analyzed after adjusting for misclassification of cause of death. Age-standardization was used to adjust for population differences in gender and age group. SMR was used to explore deaths. A Poisson model was used to compare stroke death rates after adjusting for demographic factors.

Results: SMRs were greater than 1 for both males and females. These SMRs indicated that there were excess stroke deaths for both males and females in Suphan Buri compared to Bangkok. The Poisson model was fitted well with the stroke data. The stroke death rates (per 100,000 population) was statistically significant higher in

Suphan Buri (232, [95%CI 218, 248]) than those in Bangkok (193) after adjusting for gender-age-group. The rates were higher for males than for females in all age groups under 80 years.

Conclusions: Increasing efforts in health care and stroke prevention should target at males in suburban areas. The statistical methods used in this study can be applied to any regions for comparing cause-specific mortality between populations.

Keywords: Bangkok, Poisson regression, standardization, stroke, Suphan Buri

Introduction

Stroke is the leading cause of death and the leading cause of adult disability worldwide.[1] The number of stroke deaths, particularly in developing countries, is expected to increase in future due to ageing populations, the effects of smoking, and the rising burden of diabetes and hypertension.[2] In Thailand, stroke is a major health problem. It is also a leading cause of deaths and a leading cause of disability among persons ages more than 45 years old.[2-3] The geographical variation of stroke deaths between the five regions were observed with the highest prevalence in the central region (2.41%), followed by the southern (2.29%), northern (1.46%), and north-eastern regions (1.09%).[4] Understanding the geographical variation of stroke is essential for health care planning and resource allocation.

Developed countries have systems in place for assessing causes of death in the population and their mortality statistics are reliable. Most developing countries do not have such systems. Although some countries including Thailand have such systems, causes of death are often misclassified.[5] The poor quality data in Thailand because

40% deaths are ill-defined.[5] Therefore, it needs adjustment to provide national deaths statistics. Methods for adjustments based on verbal autopsy (VA) data in 2005 using cross referencing method [6-8] and logistic regression model[9-10] have been published.

High stroke prevalence was reported in central region. This region comprises 11 provinces, only Suphan Buri and Bangkok were selected in the VA study.[7] Suphan Buri is a suburban area which is located northwest of Bangkok, and has population density in the range 100-250 persons per square kilometer, whereas Bangkok is the capital city of Thailand, and has population density in the range 400-4,060 persons per square kilometer .[11] In addition, health care systems are located to a greater extent in Bangkok than in Suphan Buri and other places in Thailand. For these reasons our study assessed this geographical health disparity by comparing stroke mortality between Suphan Buri and Bangkok.

There have been several studies that compared deaths including stroke between populations[12-15] and they used age-standardization methods to adjust the population difference in age distribution, and always provide the results based on the SMR. However, these analysis methods were not based on the statistical significant test. When using SMR, its precision should be calculated.[16-17] Since this method does not directly provide 95%CI and p-values an appropriate statistical model is needed.

This study uses SMR and Poisson regression to analyze stroke mortality data of Supan Buri and Bangkok. A Poisson model is generally applied in situation with the outcome variable is count.[18] Generally, mortality counts data has Poisson distribution. There are some situations that over dispersion occur. In this case negative binomial model can be applied.[18-19] The objective of the analysis is to compare stroke death rates for persons aged 30 years or more between two populations.

Methods

Data sources and management

This study uses the 2005 verbal autopsy (VA), Thai death registration (DR), and projected population in 2005. The 2005 VA and DR data were obtained from the Bureau of Health Policy and Strategy, Ministry of Public Health. The VA data was used to adjust for numbers of stroke death in the DR database. The detailed of adjustment methods were provided elsewhere.[9-10] These methods were summarized in Figure 1.

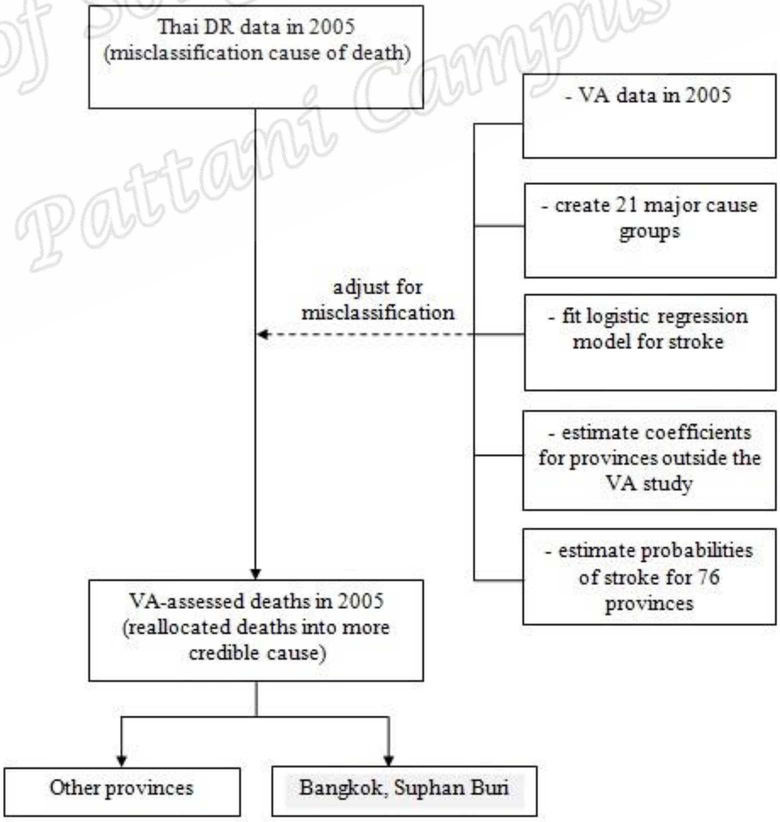


Figure 1: Diagram for estimating stroke deaths.

We began with analysis of the VA data used chapter-block classification of ICD-10 codes[20], where blocks are classified mainly by human organs, creating 21 major cause groups with VA counts. A logistic regression model was fitted to stroke cause with demographic (province, gender, age-group) and medical (reported cause group and location) factors as determinants. A Receiver Operating Characteristic (ROC) was used to assess the model.[21] A spatial triangulation method was used to estimate coefficients for the provinces outside the VA study. The estimated probabilities from the model were applied to the DR data giving the VA-assessed stroke deaths for the whole of Thailand. Then VA-assessed stroke deaths in Suphan Buri and Bangkok were chosen for comparison.

The 2005 projected populations for Suphan Buri and Bangkok were used. The projection method is based on the assumptions about mortality and female fertility.[22-23] We developed a mathematical model for projection, which is formulated in terms of the size of the population, the mortality and fertility rate per unit population. To avoid the estimation of the age-specific fertility, we assume that births decrease by 2% annually. From the model we start with gender-specific population census age distributions in 2000, and we used reported death rates by age and gender from 2000 to 2009. Thus projected populations for the whole of Thailand from 2000 to 2009 were obtained. Finally, Population in 2005 for Suphan Buri and Bangkok were chosen for comparison.

Preliminary analysis

Suphan Buri and Bangkok have different age distributions. We applied Bangkok's gender age-specific stroke death rates to the Suphan Buri population. This calculation provides the expected number of stroke deaths by gender (male and female) and age group (30-39, 40-49, 50-59, 60-69, 70-79, 80+) for Suphan Buri, based on corresponding rates for Bangkok.

For each gender, the SMR was then calculated by dividing the total number of observed deaths by the total number of expected deaths. If SMR is greater than 1, there is said to be "excess deaths" in the study population.

Poisson regression model

The stroke counts data was used. Poisson generalized linear model was used for count data. Suppose that n_{ij} is the number of observed stroke deaths in province *i*, gender-age group *j*, and that P_{ij} is the population at risk per 100,000 for each data cell. Denoting the corresponding mean death rate by λ_{ij}, where

$$\ln(\lambda_{ij} / P_{ij}) = \mu + \alpha_i + \beta_j . \quad (1)$$

In this model, μ is a constant, α_i and β_j are province and gender-age group, respectively. The generalized linear model based on Poisson distribution with mean λ_{ij} and P_{ij} is and offset.[19] The Poisson distribution has only the single parameter λ, and the variance is also λ.

Plot of residuals against normal quantiles is used to assess how well a model fits the data. For Poisson model deviance residuals based on the likelihood are recommended. The plotted residuals should follow a straight line on the plot corresponding to the normal quantiles.

The model results were presented as a graph of adjusted death rates with their corresponding 95% confidence intervals. The plots can be used to display the pattern of death rates for each factor adjusted for other factor.

R program version 3.2.1[24] was used for data management, statistical analysis and graphical displays.

Results

Projected population for Suphan Buri is approximately 481,868 (227,151 for males and 254,717 for females) and for Bangkok is approximately 3,797,176 (1,746,299 for males and 2,050,877 for females).

VA-assessed stroke deaths for Suphan Buri is 1,119 (559 for males and 560 for females) and for Bangkok is 4,724 (2,523 for males and 2,201 for females)

Age distribution for Suphan Buri and Bangkok

Figure 2 shows population's age distribution for Suphan Buri and Bangkok by gender. The left panel shows age distribution for males and the right panel for females. Bar strips represent the proportions of population for each province by age-group. Suphan Buri and Bangkok have different age distributions. Bangkok has a greater proportion of population aged 30-49 years for both males and females than Suphan Buri whereas Suphan Buri has a higher proportion of females aged 60 years or more.

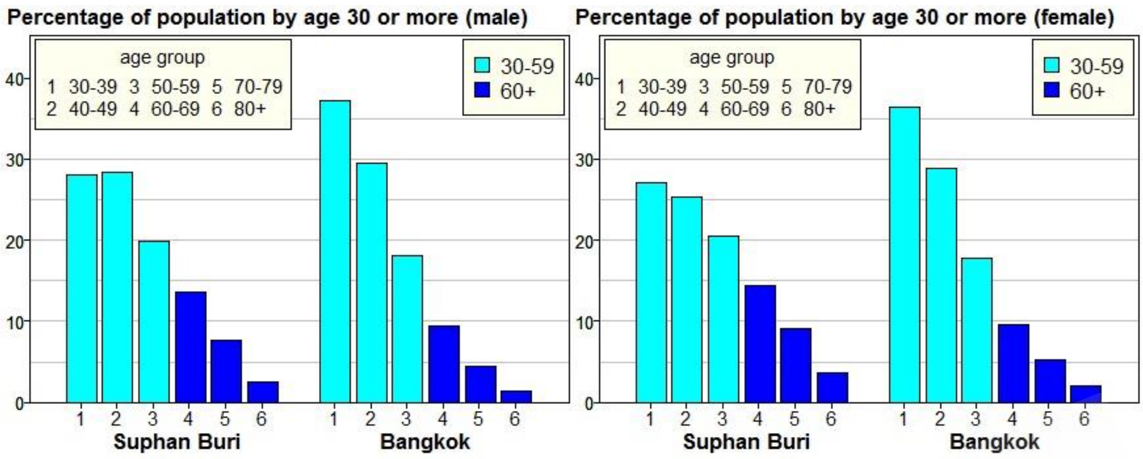


Figure 2: Percentage of population age distribution for Suphan Buri and Bangkok by gender for population aged 30 or more.

Age-specific stroke death rates

Figure 3 shows age-specific stroke death rates by gender. Blue strips represent the rates for Bankok and the superimposed magenta strips represent rates for Suphan Buri. For clarity, a cube root *y*-axis scale is used. Bangkok population has lower stroke death rates than those in Suphan Buri in every age group. The rates for males were greater than those for females in all age group with the exception for female aged 80 or more for both Suphan Buri and Bangkok. Males aged 80 or more and female aged 60-69 has similar age-specific death rates between Suphan Buri and Bangkok. However, this graph does not accurately depict the extra burden of death in Suphan Buri compared to Bangkok because Suphan Buri and Bangkok have different population size and age distribution.

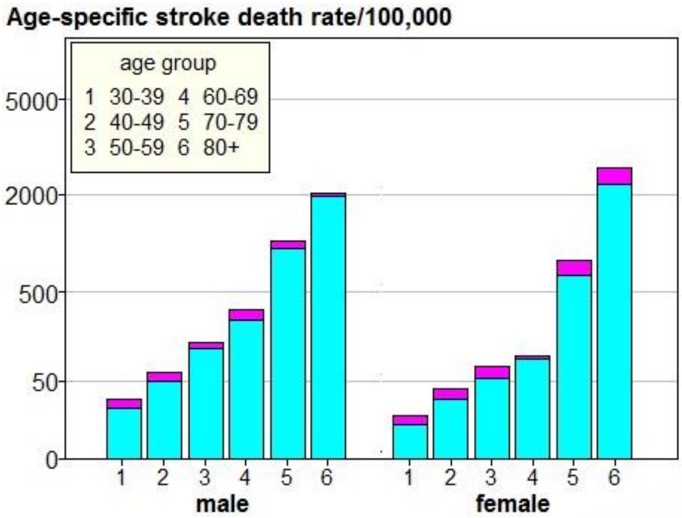


Figure 3: Age-specific stroke death rates for Suphan Buri and Bangkok for population aged 30 or more.

Age-standardized stroke deaths

Applying Bangkok gender age-specific stroke death rates to the Suphan Buri population, the total number of stroke deaths for Bangkok (929) would be 83% of the number of stroke deaths in Suphan Buri (1,119) as shown in Figure 4. The excess stroke deaths in Suphan Buri are only 16% for males although female deaths are 26% higher. The graph shows that Suphan Buri excess stroke death rates are higher than those for Bangkok in every age group. To detect statistical significant excess an appropriate statistical model is needed.

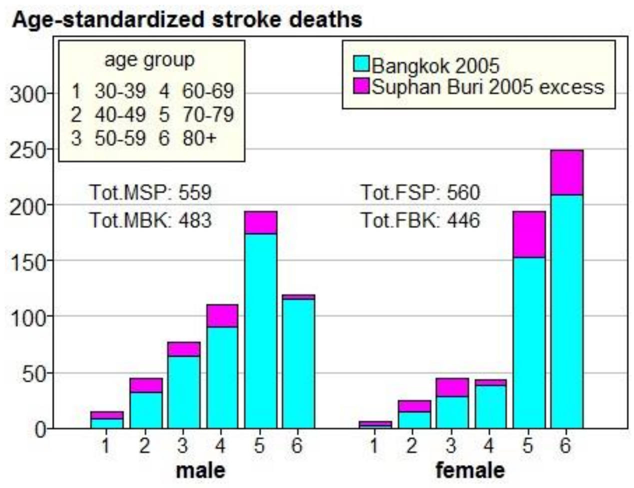


Figure 4: Age-standardized stroke deaths for Suphan Buri and Bangkok by gender-age group for population aged 30 or more.

Statistical model results

The results from Poisson regression model are presented in Figure 5. The above panel shows residual plot. The model was fitted well with the data, and the chi-squared test (p-value= 0.4087) indicates that the fit is statistically acceptable. The lower panel shows the adjusted stroke death rates with 95% confidence intervals. These confidence intervals compare death rate in Suphan Buri with Bangkok. Bangkok is a reference group for province and males aged 70-79 is a reference group for gender-age group. The plus and dot symbols denoted the unadjusted ("crude") and adjusted death rates, respectively. There is a little difference between the crude and adjusted death rates, so there is little evidence of confounding.

The stroke death rates (per 100,000 population) was statistically significant higher in Suphan Buri (232, [95%CI: 218, 248]) than those in Bangkok (193) after adjusting for gender-age group. Rates were higher in males than in females in every age group under 80 years.

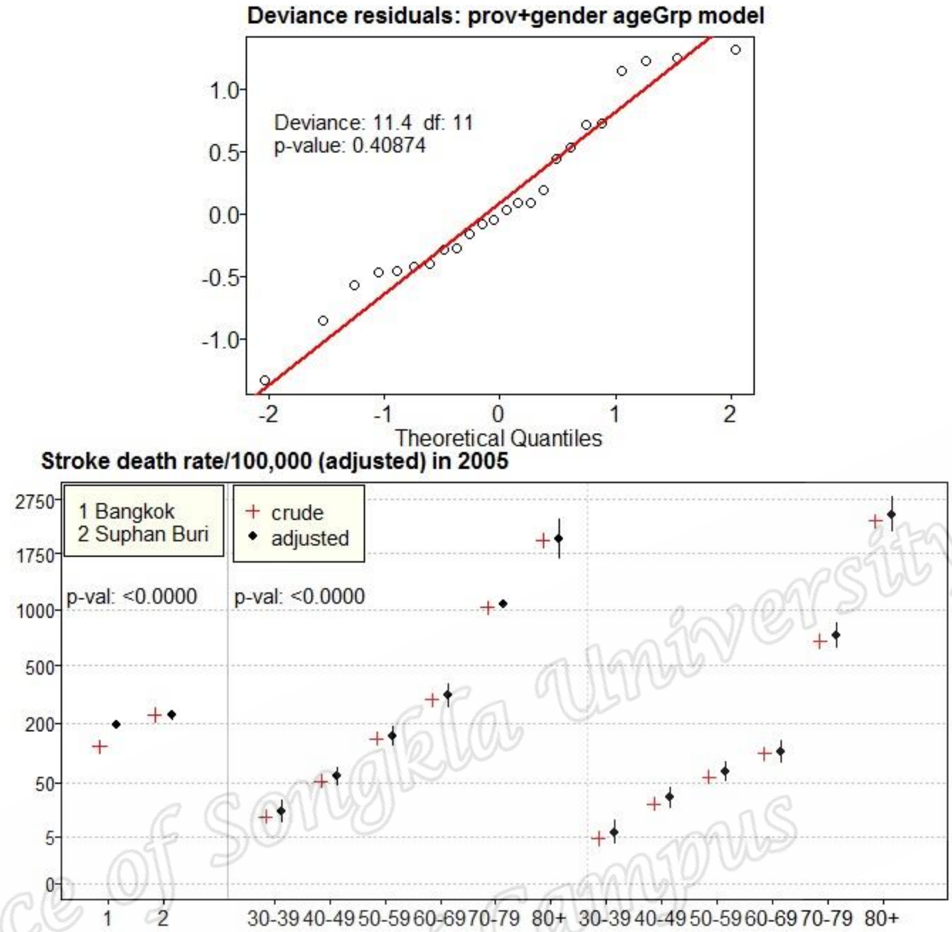


Figure 5: Poisson regression analysis for comparing stroke death rates between Suphan Buri and Bangkok after adjusting for gender-age group, with these factors unadjusted (denoted by + symbols) and after adjusting for other factor (denoted by • symbol): *Bangkok as a reference group for province and male aged 70-79 as a reference group for gender-age group.*

Discussion

This study used appropriate methods for comparing stroke death rates between two populations in suburban (Suphan Buri) and urban city (Bangkok). The methods involved SMR and Poisson regression model.

The SMRs were greater than 1 for both males and females. These SMRs indicates that there were excess stroke deaths in Suphan Buri for both males and females. Several studies have used SMRs to measure all cause and cause-specific excess deaths including stroke.[13, 25-26] However, the main objectives of these studies was not compare death rates based on the statistical model.

Our study used a Poisson regression model to compare stroke death rates between populations. The model provided a good fit to stroke mortality data. This model gives death rates for a factor (province) adjusted for other factor (gender-age group). A Poisson model has the advantage over the SMR method as it directly provides a 95% confidence interval and p-value for adjusted death rates. Moreover, it can include several factors for analysis and can also extend to negative binomial model to allow for over dispersion. Therefore it helps to facilitate comparison of death rates between populations.

The adjusted stroke death rate from the model was much higher in Suphan Buri than in Bangkok. This mortality disparity may partly be due to differences in access to quality stroke care and the prevalence of stroke risk factors among these areas.[27-29] An important understanding of these situations will assist in efforts to reduce health disparities.

Suphan Buri was found to be high in prevalence of hypertension which is a major risk factor of stroke.[15] A study in China showed that the overall relative risk of stroke associated with hypertension was 5.43.[29] This evidence led to more emphasis in its prevention, and therefore more emphasis given towards rising hypertension, especially in population who live in suburban areas such as Suphan Buri.[15]

Our finding also found that stroke death rates for males were greater than those for females with an exception for females aged 80 years or more. Excess female stroke mortality needs to be investigated further.

Using estimated deaths based on more accurate ICD-10 diagnosis from the 2005 VA data has the obvious advantage that it is reporting the real situation within a population. It gives a sense of anchorage in understanding prevailing epidemiological conditions, and, in doing so, enables the development of context-specific program for improving the health of that population.[2, 10]

Further investigation will be needed to investigate the differences in health care resources, lifestyle behavior, and other risk factors in suburban and urban areas that might explain the suburban/urban difference in stroke death rates.

Conclusion

This study reveals that age-standardization must be used to adjust for population differences in age distribution. SMR was used to explore excess deaths and Poisson model was used to compare death rates after adjusting for demographic factors. Stroke is still the main threat to the health of population in Suphan Buri (suburban area). The statistical methods used in this study can be applied to other regions for comparing death rates between populations.

Acknowledgements

This study was funded by the Office of the Higher Education Commission of Thailand and also by the Centre of Excellence in Mathematics, the Commission on Higher Education, Thailand. We would like to thank the Ministry of Public Health and Dr. Kanitta Bundhamcharoen for providing us data. We also would like to express our thanks to Prof Don McNeil for assistance and helpful guidance.

References

1. Mathers CD, Loncar D. Projections of global mortality burden of disease from 2002 to 2030. *PLos Medicine*. 2006; 3: 2011-30.
2. Hoy DG, Rao C, Hoa NP, Suhardi S, Lwin AMM. Stroke mortality variations in South-East Asia: empirical evidence from the field. *Int J of Stroke*. 2013; 8: 21-27.
3. Jalayondeja C, Kaewkungwa J, Sullivan PE, Nidhinandana S, Pichaiyongwongdee S, Jareinpituk S. Factors related to community participation by stroke victims six month post-stroke. *Southeast Asian J Trop Med* Public Health. 2011; 42(4): 1005-1013.
4. Suwanwela N. Stroke epidemiology in Thailand. *JoS*. 2014; 16(1). doi: 10.5853/ jos.2014.16.1.1.
5. Mathers CD, Fat DM, Inoue M, Rao C, Lopez D. Counting the dead and what they die from: an assessment of the global status of cause of death data. *Bull World Health Organ*. 2005; 83: 171-177.
6. Rao C, Porapakkham Y, Pattaraarchachai J, Polprasert W, Swanpunyalert N, Lopez AD. Verifying causes of death in Thailand: rationale and methods for empirical investigation. *Popul Health Metr*. 2010; 8(11). doi: 10.1186/1478-7954-8-11.

7. Porapakkham Y, Rao C, Pattaraachachai J, Polprasert W, Vos T, Adair T, et al. Estimated causes of death in Thailand, 2005: implications for health policy. *Popul Health Metr*. 2010; *8*(14). doi: 10.1186/1478-7954-8-13.

8. Pattaraachachai J, Rao C, Polprasert W, Porapakkham Y, Pao-in W, Singwerathum N, et al. Cause-specific mortality patterns among hospital deaths in Thailand: validating routine death certification. *Popul Health Metr*. 2010; *8*(12). doi: 10.1186/1478-7954-8-12.

9. Chutinantakul A, Tongkumchum P, Bundhamcharoen K, Chongsuvivatwong V. Correcting and eestimating HIV mortality in Thailand based on 2005 verbal autopsy data focusing on demographic factors, 1996-2009. *Popul Health Metr*. 2014; 12(25). doi: 10.1186/s12963-014-0025-x.

10. Kinjun N, Lim A, Bundhamcharoen K. A logistic regression model for estimating transport accident deaths using verbal autopsy data. *Asia Pac J Public Health.* 2015; 27(3): 286-292. doi: 10.1177/1010539514529810.

11. National Statistical Office. The 2000 Population and Housing Census. The Ministry of Information and Communication Technology, Thailand. http://web.nso.go.th/en/census/poph/prelim_e.htm. Accessed August 20, 2014.

12. Brønnum-Hansen H, Davidsen M, Thorvaldsen P. Long-Term Survival and Causes of Death after Stroke. *J Am Hearth Assoc*. 2001; 32: 2131-2136. doi: 10.1161/hs0901.094253

13. Torrey B, Haub C. A comparison of US and Canadian Mortality in 1998. *Popul Dev Rev.* 2004; 30: 519-530.

14. Sather D, Fryer GE, Jr GE, McCann J, Troutman A, Woolf SH, et al . What if we were equal? A comparison of the Black-White mortality gap in 1960 and 2000. *Health Aff*. 2005; 24: 459-464

15. Chongsuvivatwong V, Yipintsoi T, Suriyawongpaisal P, Cheepudomwit S, Aekplakorn W, Faramnuayphol P, et al. Comparison of cardiovascular risk factors in five regions of Thailand: Inter ASIA data. *J Med Assoc Thai*. 2010; 93: 17-26.

16. Breslow NE, Day NE. Indirect standardization and multiplicative models f rates, with reference the age adjustment of cancer and relative frequency data. *J Chronic Dis*. 1975; 28: 289-303.

17. Court, B.V. and Cheng, K.K. Pros and cons of standardized mortality ratios. *Lancet*. 1995; 346(25), 1432.

18. McCullagh P, Nelder JA. *Generalized linear models*. London: Chapman and Hall; 1989.

19. Venables WN, Ripley BD. *Modern applied statistics with S, 4th Ed.* New York: Springer-Verlag; 2002.

20. World Health Organization. *ICD-10 International Statistical Classification of Diseases and Related Health Problems.* Geneva: Switzerland, WHO; 2004.

21. Sakar S, Midi H. Importance of assessing the model adequacy of binary logistic regression. *J Appl Sci*. 2010; 10: 479-486.

22. Vapattanawong V, Prasartkul P. Under-registration of deaths in Thailand in 2005–2006: results of cross-matching data from two sources. *Bull World Health Organ*. 2011; 89: 806-812. doi:10.2471/BLT.10.083931.

23. Health Research System Institute, Institute for Population and Social Research, Mahidol University. *Population Projection for Thailand, 2000-2025*. Bangkok, Thailand; 2003.

24. The R Core Team. *R: A language and environment for statistical computing.* https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf. Accessed by March 30, 2014.

25. Fukuda Y, Nakamura K, Takano T. Increased excess deaths in urban areas: quantification of geographical variation in mortality in Japan, 1973-1998. *Health Policy*. 2004; 68: 233-44.

26. Faramnuayphol P, Chongsuvivatwong V, Pannarunothai S. Geographical variation of mortality in Thailand. *J Med Assoc Thai*. 2008; 91: 1455-1460.

27. Tideman P, Taylor AW, Janus E, Philpot B, Clark R, Peach E. et al. A comparison of Australia rural and metropolitan cardiovascular risk and mortality: the Greater Green Triangle and North West Adelaide population surveys. *BMJ Open*. 2013; 3. doi:10.1136/bmjopen-2013-004724.

28. Yiengprugsawan V, Carmichael GA, Lim LL, Seubsman S, Sleigh AC. Has universal health insurance reduced socioeconomic inequalities in urban and rural health service use in Thailand?. *Health Place*. 2010; 16: 1030-1037.

29. OECD/World Health Organization. *Health at a Glance: Asia/Pacific, 2014: Measuring Progress towards Universal Health Coverage*. Geneva, Switzerland: OECD/WHO; 2014.

30. Wang X, Jiang G, Choi BC, Wang D, Wu T, Pan Y, et al. Surveillance of trend and distribution of stroke mortality by subtype, age, gender, and geographical areas in Tianjin, China, 1999-2006. *Int J Stroke*. 2009; 4: 169-174.

3.2.3 A Statistical methods for correcting misreported multinomial outcome data with application to stroke mortality in Thailand

Manuscript was submitted to the *Walailak Journal of Science and Technology*

Arinda Ma-a-lee[1], Nattakit Pipatjaturon[2] and Phattrawan Tongkumchum[1]

[1]Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Thailand

[2]The office of Diseases Prevention and control 2th Phitsanulok, Phitsanulok, Thailand

Abstract

Causes of deaths in Thailand are misreported because about 40% of deaths have been recorded as ill-defined. This study aims to offer statistical methods to correct misreported multinomial outcome. The methods involve analysis of verbal autopsy (VA) data. Since the outcome is a nominal variable with 21 levels, the appropriate model for systematic analysis of death by ICD-10 code is multinomial regression. However, it is simpler and more informative to separately fit logistic regression models to the 21 outcome cause groups, and then rescale the results to ensure that the total number of estimated deaths for each group match those reported in the corresponding populations. This method also gives confidence intervals for percentages of deaths in cause groups for levels of each risk factor adjusted for other risk factors. These confidence intervals are compared with bar charts of sample percentages to assess evidence of confounding bias. Area plots are used to show results by gender, age group and year. The methods were illustrated using stroke deaths.

Keyword: logistic regression, misreported deaths, multinomial regression, stroke deaths, Thailand

Introduction

Mortality statistics are important for measuring health status at global and national levels. The World Health Organization (WHO) has used mortality statistics as indicators for global health [1]. National mortality statistics reflect the quality of life of people and provide useful information for priority setting and resource allocation in a country.

Data quality is a main issue to provide accurate mortality statistics especially in developing countries. In Thailand, under-registration and ill-defined causes of death are reported [2-4]. These lead to some degree of misclassification of causes of death for example, septicemia is over-reported but stroke and others are under-reported [5].

A verbal autopsy (VA) study has been widely used for the assessment of cause of death in countries where death registration (DR) systems are fragile and most people die at home without medical certification of cause of death. The VA method determines the cause of death from data collected about the symptoms and signs of illness and the events preceding death. It has procedures to ensure that the causes of death collected are of high quality [6-9].

Estimated causes of death based on the VA data have been published [7-8, 18, 22-23]. They are using and not using statistical model. A benefit of using statistical models is that several factors can be investigated and the effect of one factor can be adjusted for other factors. Moreover, a statistical model can detect confounding that biases the results.

This study offers methods for correcting multinomial outcome data with application to stroke mortality in Thailand. The methods involve analysis of 2005 VA sample. The multinomial outcome is cause of death comprising 21 categories. Demographic factors, province of residence, location of death, and registered cause of death are determinants. Since the outcome is nominal variable with 21 levels, the appropriate model for systematic analysis of death by ICD-10 code is multinomial regression. However, it is simpler and more informative to separately fit logistic regression models to 21 outcome cause groups, and rescale the results to ensure that the total numbers of estimated deaths for each group match those reported in the corresponding populations. The methods were illustrated using stroke mortality. A previous cross-reference analysis of the 2005 VA data reported that stroke was a leading cause of death in Thailand, causing 10.7% of deaths [8]. Moreover, stroke was reported as a leading cause of disability and death among people aged 45 years and older [10]. Stroke prevalence also varies with demographic and geographic factors [11].

Materials and Methods

The VA data and death registration (DR) data were used. These were obtained from the Bureau of Health Policy and Strategy, Ministry of Public Health. The VA survey was carried out in 2005 by the Setting Priorities using Information on Cost-Effectiveness (SPICE) analysis project to assess causes of death based on a sample of 9,644 deaths (3,316 in-hospital and 6,328 outside-hospital deaths). A clustered sample was taken from 28 selected districts in nine provinces. Districts were selected by two-stage stratification where Bangkok and pairs of provinces from four regions were first randomly selected. Stratification was based on the number of deaths in the regional province or district. Then, a number of death certificates to be assessed were randomly selected form the 28 districts using Probability Proportional to Size (PPS) method.

For this study, the sample was reduced to 9,495 deaths aged five years and older. We used the chapter-block classification of ICD-10 codes based on mortality tabulation [12], creating 21 major cause groups. These groups required adequate sample sizes enough for statistical analysis. The groups with small counts were combined into a larger group as "All other". The sample sizes vary from 77 (0.8%) for septicemia (ICD code A40-A41) to 1,076 (11.3%) for stroke (ICD code I60-I69).

The outcome is cause of death and a nominal variable with 21 levels as shown in **Table 1**. It can be analyzed using a multinomial regression model. However, logistic regression model can also be used. Since logistic regression is appropriate for binary outcome, the cause of death was considered as 21 binary variables. We create only two cause groups where one group is the cause of interest and the other group aggregates deaths from all other causes. For example, the outcome of interest is TB. There are 195 deaths due to TB coded as 1 and 9,300 deaths due to all other causes coded as 0. The outcome of interest is septicemia. There are 77 deaths due to septicemia coded as 1 and 9,418 deaths due to all other causes coded as 0. The outcome of interest is stroke with 1,076 cases coded as 1 and 8,419 deaths due to all other causes coded as 0 and so on.

Table 1 Cause of deaths comprising 21 major cause groups

Cause of death	Number of deaths	Percent
1:TB (A15-A19)	195	2.1
2:Septicemia (A40-A41)	77	0.8
3:HIV (B20-B24)	512	5.4
4:Other Infectious (A, B)[-]	219	2.3
5:Liver Cancer (C22)	500	5.3
6:Lung Cancer[+] (C30-C39)	320	3.4
7:Other Digestive Cancer (C15-C26)[-]	290	3.1
8:Other Cancer (C[-], D00-D48)	697	7.3
9:Endocrine (E00-E99)	647	6.8
10:Mental, Nervous (F00-F99, G00-G99)	223	2.3
11:Ischemic (I20-I25)	617	6.5
12:Stroke (I60-I69)	1,076	11.3
13:Other CVD (I[-])	540	5.7
14:Respiratory (J00-J99)	801	8.4
15:Digestive (K00-K93)	489	5.2
16:GenitoUrinary (N00-N99)	412	4.3
17:Ill-defined (R00-R99)	501	5.3
18:Transport Accident (V00- V99)	536	5.6
19:Other injury (W00-W99, X00-X59)	327	3.4
20:Suicide (X60-X84)	158	1.7
21:All other	358	3.8
Total	9,495	100.0

[+] *Respiratory/thoracic*, [-]*exclude above*

The determinants were gender-age group, location of death, province, and registered cause on the death certificates for a particular outcome. Gender and age group were combined. The number of levels in the gender-age group factor depends on the age distribution of deaths from the selected outcome. For stroke we chose 12 levels of a gender-age group factor with the two gender and six age groups (5-39, 40-49, 50-59, 60-69, 70-79 and 80+).

Similarly the number of levels in the DR-cause location factor depends on the number of such reported cause groups that affect the outcome cause group. For stroke we chose 18 levels of a DR-cause location factor with the two locations (in or outside hospitals) and nine DR-reported ICD-10 coded cause groups most likely to be stroke (stroke, ill-defined, septicemia, respiratory disease, mental and nervous, other cardiovascular diseases, endocrine all other, and other groups).

The province factor has nine levels corresponding to the nine provinces in the VA sample (ChiangRai, Phayao, Loei, UbonRatchathani, Supan Buri, Nakhonnayok, Bangkok, Chumpon, and Songkla).

The logistic regression model [11] formulates the logit of the probability of death due to the specified cause group as an additive linear function of the three determinant factors as follows.

$$\text{logit}(p_{ijk}) = \log(p_{ijk}/(1-p_{ijk})) = \mu + \alpha_i + \beta_j + \gamma_k \quad (1)$$

where p_{ijk} is the probability of the specified cause group in each of the i, j and k groups of determinant factors, μ is a constant, α_i, β_j, and γ_k refer to effects of the province, gender-age group and DR-cause location, respectively.

This model also gives confidence intervals for percentages of stroke deaths for levels of each risk factor adjusted for other risk factors using sum contrasts methods [13-16].

The confidence intervals based on sum contrasts has an advantage that they provide a simple criterion for classifying levels of the factor into three groups according to whether each corresponding confidence interval exceeds, crosses, or is below the overall percentage. They are more appropriate compared to the corresponding confidence intervals based on the treatment contrasts. The confidence intervals compared percentage of the specified cause group in each category with the overall percentage. They applied equitably to each category, whereas the commonly used confidence intervals based on treatment contrasts measured the difference from a reference group that is taken to be fixed and thus does not have a confidence interval. These confidence intervals are compared with bar charts of sample percentages to assess evidence of confounding bias.

Receiver Operating Characteristic (ROC) curve is used to assess the model's ability to distinguish between stroke deaths and others [17]. The ROC curve gives error rate. It plots sensitivity against the false positive rate to show how well a model predicts a binary outcome.

Since only nine of 76 provinces were in the VA survey, a spatial triangulation method [18-19] was used to interpolate the province effects outside the VA study.

The model results were applied to number of death in the DR data and thus the estimated numbers of stroke death were obtained. The VA/DR inflation factors (IF) greater than one reflects under-reported. Area plots are used to show results by gender, age group and year. Estimating number of deaths for the whole of Thailand is as follows.

Step 1: VA data in 2005 with 9,495 deaths aged five years and older

Step 2: There are 21 major cause groups (21 binary outcomes)

Step 3: Cross tabulation between VA and DR cause group

Step 4: Separately fit logistic regression models to the 21 outcome cause groups

Step 5: Estimate coefficients for provinces outside the VA study

Step 6: Estimate probabilities of the specified cause group for 76 provinces

Step 7: Applied the probabilities to the DR data in 1996-2009

Step 8: Obtain VA-assessed deaths in 1996-2009

To illustrate these methods we applied them to stroke deaths. The graphical displays and statistical analyses were performed using R program [20].

Results and discussion

Results

The VA data comprise 9,495 deaths aged five year and older. Of the 9,495 deaths, 1,076 deaths (386 in-hospital and 690 outside hospital deaths) were verified cause of death as stroke. The nine DR-cause groups most likely to be stroke were stroke (267), ill-defined (535), septicemia (60), respiratory (45), mental and nerve (42), other CVD (35), all other (31), endocrine (13), and the rest were aggregated to other group with small number of cases (48).

The p-values based on the logistic regression model were highly statistically significant for DR-cause location, gender-age group and province. **Figure 1** shows bar charts of percentage of stroke deaths superimposed with adjusted percentages and their corresponding 95% confidence intervals. The horizontal red line was the average percentage of stroke deaths (11.3%).

The 95% confidence intervals of percentages of stroke deaths in Bangkok, Suphan Buri and Songkhla were higher than the average. The adjusted percentages of stroke deaths for males in age groups 60-79 years and for females in age groups 70 years and older were above the average.

The misreported causes of stroke outside hospital were mental and nerve, ill-defined, other CVD and all other inside hospital. For outside hospital, 37.1% of mental and nerve, and 25.7% of other cardiovascular diseases were deaths due to stroke. On the other hand, in hospital 22.5% of all other causes were deaths due to stroke.

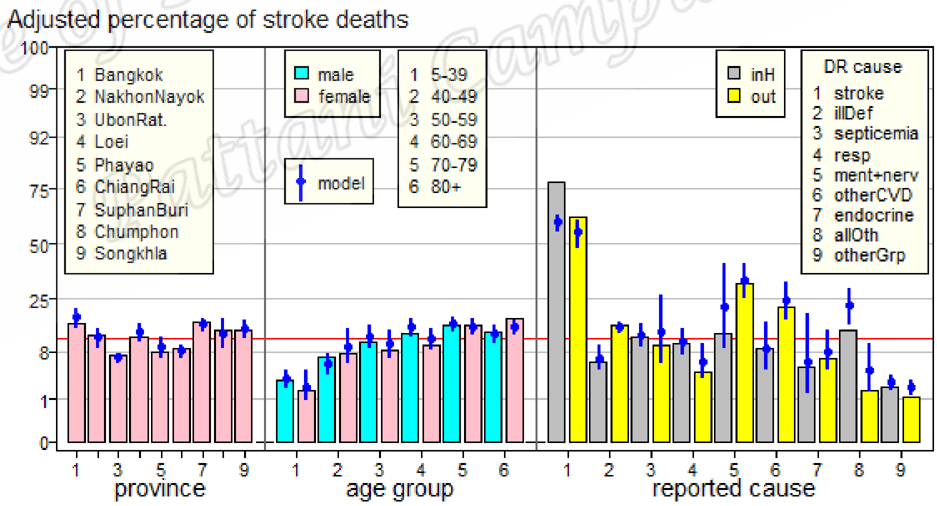


Figure 1 Adjusted percentage of stroke death by province, gender-age and DR-cause.

Figure 2 shows the ROC curve. A cut-off point in the curve, where the predicted number of stroke deaths (1,072) is close to the observed value (1,076) in the VA data. The sensitivity is 41.6%, the specificity is 92.6%. The area under the curve

(AUC) is 0.67, indicating a model performance is moderate predicting. The model estimated the proportion of other deaths (non-stroke) that was classified as stroke (false-positive) 7.4% and the proportion of stroke deaths that were misclassified as non-stroke deaths (false-negative) 58.4%. The ROC curve for the simple model with only DR-cause location group as a determinant was also given. It shows that adding province and age-group into the model decrease error rate by 10%.

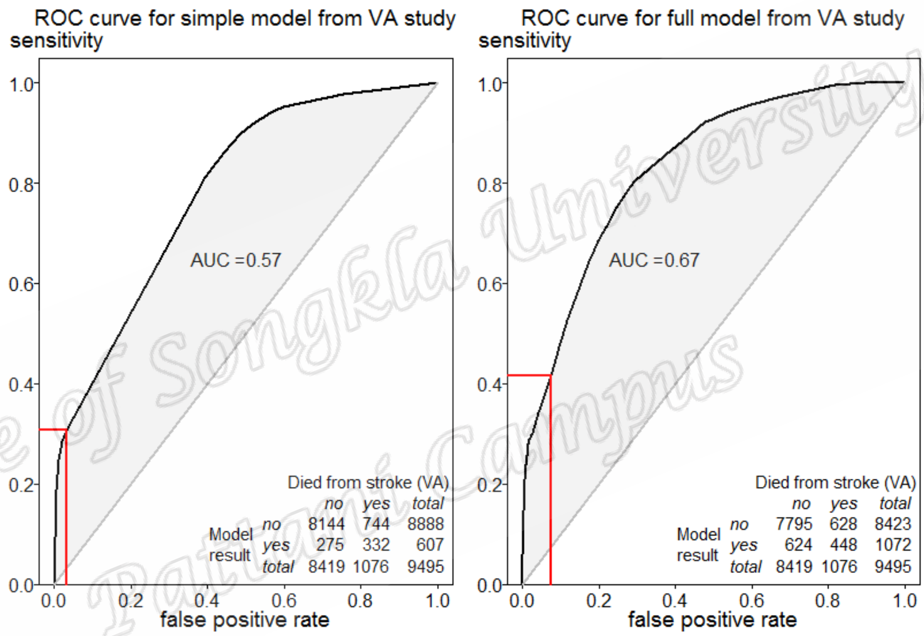


Figure 2 ROC curve of stroke model.

The result from model developed from the sample data was then applied to the DR data to estimate the number of stroke deaths in 2005. **Figure 3** shows estimated percentages of stroke deaths. The percentage of VA-estimated stroke death was high in central region (SingBuri, ChaiNat and SamutSongKhram) and low in upper north (ChiangMai, ChiangRai, MaeHongSon, and Phayao) and north east (UdonThani, SakonNakhon, NakhonPhanom, Kalasin, Mukdahan, MahaSarakham, RoiEt, Yasothon, AmnatCharoen, Surin, SiSaKet and UbonRatchathani).

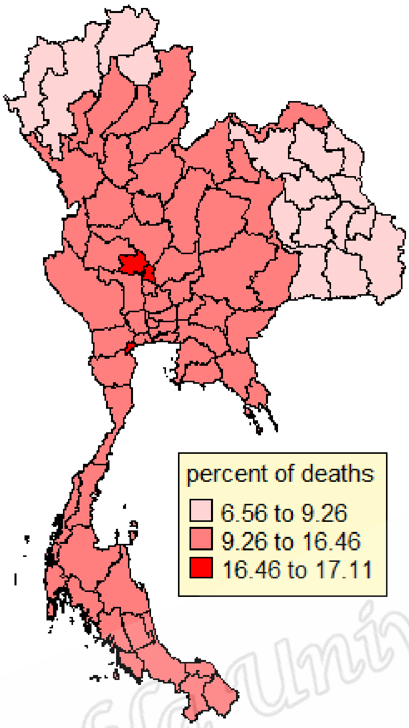


Figure 3 Percentages of VA-estimated strokes in 2005.

Table 2 shows DR-reported and VA-estimated stroke deaths. The model estimated stroke deaths as 24,110 (11.0%) in males and 21,913 (13.3%) in females. The reported stroke deaths were 9,105 (4.2%) in males and 6,470 (3.9%) in females. The inflation factors (IFs) show that deaths in all gender-age groups were substantially under-reported, especially for age group 60 years and older.

Table 2 DR-reported and VA-estimated stroke deaths in 2005

Gender age group	DR stroke	% DR stroke	VA-estimated stroke	% VA-estimated stroke	IF
male	9,105	4.2	24,110	11.0	2.7
m:5-39	1,145	2.5	1,388	3.0	1.2
m:40-49	1,579	5.5	1,878	6.5	1.2
m:50-59	1,964	6.3	3,707	11.8	1.9
m:60-69	1,886	5.1	5,158	13.9	2.7
m:70-79	1,701	3.9	7,107	16.5	4.2
m:80+	830	2.5	4,872	14.8	5.9
female	6,470	3.9	21,913	13.3	3.4
f:5-39	361	2.0	362	2.0	1.0
f:40-49	617	4.7	1,096	8.4	1.8
f:50-59	1,035	5.5	1,821	9.7	1.8
f:60-69	1,339	4.9	3,083	11.2	2.3
f:70-79	1,805	4.5	6,773	16.7	3.8
f:80+	1,313	2.8	8,778	18.6	6.7

Figure 4 shows area plots of DR reported, simple model estimated and full model estimated number of deaths from stroke by gender-age groups in 1996-2009. The area of each colour strip denotes the number of deaths in each age group.

The total numbers of deaths reported for 14 years are 157,537. The estimated total numbers of stroke deaths from the simple and full models are 549,653 and 570,245, respectively. The total number of DR reported stroke deaths were lower than those estimated by simple model and full model by factors of 3.49 and 3.61, respectively.

The simple model gave large proportions of stroke deaths at ages below 40 years, these were reduced when full model was used. For the older age groups, cause of stroke deaths was already improved in accuracy by the simple model. The total number of DR reported stroke deaths were lower than those estimated by simple model and full model by factors of 3.28 and 3.27 for males. For females the total number of DR reported stroke deaths were lower than those estimated by simple model and full model by factors of 3.78 and 4.10.

The area plots for stroke deaths clearly reveal that numbers of deaths were under reported especially for the earlier years. Similar patterns are seen with number of deaths increase in recent years.

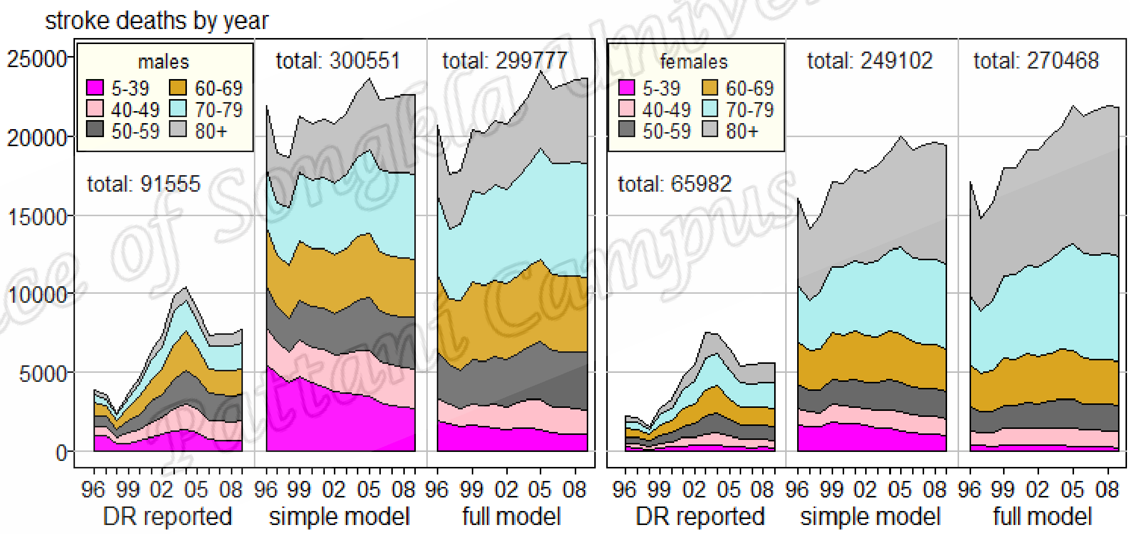


Figure 4 Area plot of DR reported and VA estimated stroke deaths

Discussion

Death due to 21 causes was considered as a multinomial outcome. Since the outcome is a nominal variable with 21 levels, the appropriate model for systematic analysis of death by ICD-10 code is multinomial regression. However, it is simpler and more informative to separately fit logistic regression models to each outcome cause groups, and then rescale the results to ensure that the total numbers of deaths estimated for each group match those reported in the corresponding populations.

This study described the methods based on logistic regression for correcting misreported cause of death and illustrated the methods using death from stroke. The data were from 2005 VA sample. Separate logistic regression model of VA-assessed causes of deaths with demographic factors is found to be appropriate to use. It allowing for gender-age group, province, and DR-cause location predicted causes specific deaths with higher sensitivity and specificity compared to those derived from simple model (simple cross-referencing method) [8]. The models do not ensure that adjusted death counts in each year match reported totals because they aggregate results from separate logistic regression models.

Although multinomial model ensures that adjusted and reported totals match, simply scaling totals from separate logistic models gives similar results. Advantage of using logistic regression model to analyse data is that it can handle general determinants. The logistic function has many desirable properties. Its range is between 0 and 1 when the independent variable varies from -∞ to ∞, so the logistic regression model can be used to model the probability of an individual death. In addition, logistic regression can control confounding and assess interaction very effectively when there are several confounders or the confounder is a continuous variable [21]. Moreover, it can be used to calculate an odds ratio and its confidence interval directly, so that the results can be interpreted easily. The probability of a given subject death from a specific disease can also be calculated.

The model can be simply extended to the larger target population comprising all deaths in Thailand for longer periods of time and other populations in similar context.

To reduce costs from conducting VA study for the whole country, we proposed an analysis of the VA data using appropriate statistical methods to a large-scale VA study, for example, in case of HIV [18], transport accident [22] and liver cancer [23].

The importance of evaluating the reliability and validity of causes of death in mortality statistics has long been recognized in public health [24]. Periodic validation of the quality of diagnostic information ensures that countries have a more confident basis on which to develop their policies and guide health planning [25]. VA survey is generally the most reliable method to determine causes of death [7, 26]. However, conducting a survey is expensive and time consuming. It is important for public authorizes to pay attention on quality of death registry rather than conducting verbal autopsy.

This method enables public health researchers to estimate percentages of specific causes of deaths in countries where there is low quality for recorded cause of deaths but reliable sample data such as a VA study should be available.

The logistic regression was used to estimate number of stroke deaths with province, gender-age group and cause-location from the VA data, which assessed the true causes of death. All variables were statistically significant and the AUC indicated model moderate predicting [27]. The model has ability to allocate misclassification of stroke deaths by gender-age group, DR-cause location and province.

The estimated numbers of death were higher than the DR reported. The percent of deaths were higher than those from the previous publication based on the same data set but use cross-referencing method based on proportionate mortality distributions [8]. A model-based method allowing for gender-age group and province effects predicted stroke death when compared with those stroke death derived from cross-referencing from simple tabulation.

The stroke deaths were misclassified as nervous cause, and other cardiovascular diseases. It agrees with a previous study which reported that some under reported stroke deaths due to coding error was diagnosed as other cardiovascular diseases [5].

Stroke is a cause of death among elderly. It is not surprise that high percent of stroke death were observed among people aged 60 and older. This may be related to high prevalence of stoke among adults aged 75 years and older [11]. Another study also supported that the average age of stroke onset was 65 years [28].

Geographical distribution of stroke deaths was higher in central region and lower north-eastern region. The geographical variation on stroke deaths in 2000 has been reported of high Standardized Mortality Ratio (SMR) in Bangkok and lowest in upper north-eastern region [29]. In addition, Thailand Epidemiology Stroke found stroke prevalence was highest in Bangkok and lowest in north-eastern region [11].

Our analysis has few limitations. The survey design did not stratify by strong a predictor of the outcome such as registered cause. Thus the study sample may not adequately cover the population at risk. For example, the Muslim majority districts were not been selected.

Conclusions

The methods enable health professionals to estimate any specific causes of deaths in countries where causes of death are of low quality and reliable cause of death from other sources are available. In addition, there is still a substantial misclassification of stroke mortality according to our model.

Acknowledgements

This study was funded by the Program Strategic Scholarships Fellowships Frontier Research Networks (Specific for Southern region) for the Ph.D. Program Thai Doctoral degree from Office of the Higher Education Commission, Thailand and also by the Centre of Excellence in Mathematics, the Commission on Higher Education, Thailand We would like to thank Professor Don McNeil for his helpful guidance and Dr. Kanitta Bundhamcharoen from Thai Ministry of Public Health for providing us the data.

References

[1] World Health Organization. World Health Statistics 2012. Geneva, Switzerland. 2012.

[2] CD Mathers, DM Fat, M Inoue, C Rao and AD Lopez. Counting the dead and what they died from: and assessment of the global status of cause of death data. *Bull World Health Organ*. 2005; **83**, 171-7.

[3] V Tangcharoensathien, P Faramnuayphol, W Teokul, K Bundhamcharoen and S Wibulpholprasert. A critical assessment of mortality statistics in Thailand: potential for improvements. *Bull World Health Organ*. 2006; **84**, 233-9.

[4] P Vapattanawong and P Prasartkul. Under-registration of deaths in Thailand in 2005-2006: results of cross-matching data from two sources. *Bull World Health Organ*. 2011; **89**, 806-12.

[5] DL Brown, F Al-Senani, LD Lisabeth, MA Farnie, LA Colletti, KM Langa, AM Fendrik, NM Garcia, MA Smith and LM Morgenrtern. Defining cause of death in stroke patients: The brain attack surveillance in Corpus Christi project. *Am J Epidemiol.* 2007; **165**, 591-6.

[6] J Pattaraarchachai, C Rao, W Polprasert, Y Porapakkham, W Poa-in, S Nophcha and AD Lopez. Cause-specific mortality patterns among hospital deaths in Thailand: validating routine death certification. *Popul Health Metr.* 2010; **8**. DOI: 10.1186/1478-7954-8-12.

[7] C Rao, Y Porapakkham, J Pattaraarchachai, W Polprasert, N Swampunyalert and AD Lopez. Verifying causes of death in Thailand: rationale and methods for empirical investigation. *Popul Health Metr*. 2010; **8**. DOI: 10.1186/1478-7954-8-11.

[8] Y Porapakkham, C Rao, J Pattaraachachai, W Polprasert, T Vos, T Adair and AD Lopez. Estimated causes of death in Thailand, 2005: implications for health policy. *Popul Health Metr*. 2010; **8**. DOI: 10.1186/1478-7954-8-13.

[9] W Polprasert, C Rao, T Adair, J Pattaraachachai, Y Porapakkham and AD Lopez. Cause-of-death ascertainment for deaths that occur outside hospitals in Thailand: application of verbal autopsy methods. *Popul Health Metr*. 2010; 8. DOI: 10.1186/1478-7954-8-13.

[10] C Jalayondeja, J kaewkungwal, PE Sullivan, S Nidhinandana and S Pichaiyongwongdee. Factors related to community participation by stroke victims six month post-stroke. *Southeast Asian J Trop Med Public Health.* 2011; **42**, 1005-13.

[11] S Hanchaiphiboolkul, N Poungvarin, S Nidhinandana, NC Suwanwela, P Puthkhao, S Towanabut, J Suwantamee and M Samsen. Prevalence of Stroke and Stroke Risk Factors in Thailand: Thai Epidemiologic Stroke (TES) Study. *J Med Assoc Thai.* 2011; **94**, 427-36.

[12] World Health Organization. *ICD-10 International Statistical Classification of Diseases and Related Health Problems.* Geneva, Switzerland. 2004.

[13] WN Venables and BD Ripley. *Modern Applied Statistics with S*. 4th Edn., New York: Springer-Verlag. 2002.

[14] P Tongkumchum and D McNeil. Confidence interval using contrasts for regression model. *Songklanakarin J Sci Technol*. 2009; **31**, 151-6.

[15] N Kongchouy and U Sampantarak. Confidence intervals for adjusted proportions using logistic regression. *Mod Appl Sci*. 2010; **4**, 2–7.

[16] U Sampantarak, N Kongchouy and M Kuning. Democratic confidence intervals for adjusted means and incidence rates. *AIJCR*. 2011; **1**, 38-43.

[17] SK Sarkar and H Midi. Importance of Assessing the Model Adequacy of Binary Logistic Regression. *J Appl Sci*. 2010; **10**, 479-86.

[18] A Chutinantakul, P Tongkumchum, K Bundhamcharoe and V Chongsuvivatwong. Correcting and estimating HIV mortality in Thailand based on 2005 verbal autopsy data focusing on demographic factors, 1996-2009. *Popul Health Metr.* 2014; **12**. DOI**:** 10.1186/s12963-014-0025-x.

[19] GF Bonham-Carter. *Geographic information systems for geoscientists: Modelling with GIS.* Pergamon, Oxford. 1994.

[20] R Core Team. *A language and environment for statistical computing*, Available from: http://www.R-project.org/, Accessed August, 2014.

[21] DW Hosmer and S Lemeshow. *Applied Logistic Regression*. 2nd Edn., Hoboken, NJ: Wiley. 2000.

[22] N Kinjun, A Lim and K Bundhamcharoen. A logistic regression model for estimating transport accident deaths using verbal autopsy data. *Asia Pac J Public Health*. 2015; **27**, 286-92. DOI: 10.1177/1010539514529810

[23] S Waeto, N Pipatjaturon, P Tongkumchum, C Choonpradub and R Saelim. Estimating liver cancer deaths in Thailand based on verbal autopsy study. *J Res Health Sci*. 2014; **14**, 18-22.

[24] IM Moriyama. Problems in measurement of accuracy of cause-of-death statistics. *Am J Public Health*. 1989; **79**, 1349-50.

[25] A Khosravi, C Rao, M Naghavi, R Taylor and N Jafari. Impact of misclassification on measure of cardiovascular disease mortality in the Islamic

Republic of Iran: a cross-sectional study. *Bull World Health Organ*. 2008; **86**, 688-96.

[26] RA Lahti and A Penttilä. The validity of death certificates: routine validation of death certification and its effects on mortality statistics. *Forensic Sci Int.* 2001; **115**, 15-32.

[27] G Vanagas. Receiver operating characteristic curves and comparison of cardiac surgery risk stratification systems. *Interact Cardiovasc Thorac Surg.* 2004; **3**, 319-22.

[28] NC Suwanwela. Stroke epidemiology in Thailand. *JOS*. 2014; **16**, 1-7.

[29] P Faramnuayphol, V Chongsuvivatwong and S Panarunothai. Geographical variation of mortality in Thailand. *J Med Assoc Thai*. 2008; **91**, 1455-60.

CHAPTER 4

Discussion and Conclusions

The main results are summarized in the first section of this chapter. The implications and limitations are discussed in the later sections. Finally, the conclusion is made in the last section.

4.1 Findings

This thesis provided methods for comparing two populations. The comparisons were made on (1) Thai mortality data in 2005 and Japan in 2006, (2) stroke. Age-standardization was used to adjust for population differences in gender and age group. The population with lower age-standardized death rates was chose to be a standard population. The methods to compare stroke death rates in Suphan Buri with Bangkok in 2005 were also described. The methods involve SMR and Poisson regression model. Since death due to 21 cause group was presented as a multinomial outcome, multinomial regression model is suitable for analysis. However, it is simpler and more informative to separately fit logistic regression models to each cause groups.

For study I, Thailand suffered 47% more deaths than would be expected if they had experienced the same death rates as Japan. Of these, the first three larger excess deaths were stroke, transport accident for males, and endocrine for female. In contrast, in Japan, there were excess deaths for cause related to suicide and other digestive cancer. Other CVD and lung cancer for males in Japan were marginally higher than those in Thailand. This might be partly due to health disparity, especially for poorer people in Thailand, they still use health services less when they ill. This is because of financial barriers (OECD/WHO, 2014). In contrast Japan has the advantage that most people access to health care services and higher quality care than Thailand such as care of stroke and cancer (OECD, 2013; Fukuda *et al*, 2004).

For study II, SMRs were higher in Suphan Buri for both males and females. Using Poison regression model to compare stroke death rates between populations, the model provided a good fit to the data and gives death rates for two factors (province and gender age group) adjusted for other factor. A Poisson model has the advantage over the SMR method as it directly provides 95% confidence interval and p-value. We found that the rate in Suphan Buri was higher than those in Bangkok. The rates were higher for males than for females in all age groups under 80 years.

For study III, death due to 21 major causes was presented as a multinomial outcome. Multinomial regression model is appropriate for systematic analysis of death by ICD-10 code. However, it is simpler and more informative to separately fit logistic regression models to each cause group, and then rescale the results to ensure that the totals numbers of estimated deaths match those reported deaths in the corresponding populations.

Separate logistic regression model of VA-assessed causes of deaths with demographic factors is found to be appropriate to use. It allowing for gender-age group, province, and DR-cause location predicted causes specific deaths with higher sensitivity and specificity compared to those derived from simple model (simple cross-referencing method) (Porapakkham *et al.*, 2010).

Advantage of using logistic regression model to analyze data is that it can handle general determinants. In addition, logistic regression can control confounding and assess interaction very effectively when there are several confounders or the confounder is a continuous variable (Hosmer and Lemeshow, 2010). Moreover, it can be used to calculate an odds ratio and its confidence interval directly, so that the results can be interpreted easily. The probability of a given subject death from a specific disease can also be calculated.

4.2 Implications

The larger cause-specific excess deaths in Thailand provide a better picture for health authorities to help design prevention program in which cause of death is relatively high.

Stroke is the main threat to the health of population in Suphan Buri or suburban areas. Increasing efforts in health care and stroke prevention should target suburban areas. Understanding these relationships illuminates areas where resources can be targeted toward strategies that will extend years of life.

4.3 Limitations and Further study

In the study I we used doughnut plots to represent the 21 cause-specific excess death by gender more clearly. However, it is lack of information of age group and year which may help us in getting additional information. Further studies need to include these factors to assess burden of deaths by region in Thailand compared to Japan using doughnut plot.

A multinomial model may be needed to identify significant excess. However, the multinomial model of outcome with several categories (21 cause groups) provided less interpretable results.

In the study II, further investigation will be also needed to investigate the differences in health care resources, lifestyle behavior, and other risk factors in suburban and urban areas that might explain the suburban/urban difference in stroke death rates.

In the study III, The survey design did not stratify by strong a predictor of the outcome such as registered cause. Thus the study sample may not adequately cover the population at risk. For example, the Muslim majority districts were not been selected.

4.4 Conclusions

In the study I, we compared 21 cause-specific deaths in Thailand in 2005 with Japan in 2006. It shows that if the age-specific death rates in Thailand were the same as in Japan (standard population), the total number of deaths would be only 53% of the number of deaths in Thailand in 2005. Thus health inequality cause-specific excess deaths remain in Thailand, and impose a significant public health burden that demand attention and action by health policy maker.

Doughnut plots is an effectively graph to show cause-specific excess deaths by gender. Stroke is the main excess deaths in Thailand in 2005 whereas suicide is the main excess deaths in Japan.

In the study II, it reveals that age-standardization must be used to adjust for population differences in age distribution. SMR was used to explore deaths and

Poisson model was used to compare death rate after adjusting for demographic factors. The statistical methods used in this study can be applied to other cause and also other regions for comparing death rates between populations.

Appropriate graphs from the statistical results provide useful tools for assessing mortality disparity between populations.

In the study III, the 21 causes group was presented as a multinomial outcome. Multinomial regression is appropriate for analyses this outcome. However, it is simpler and more informative to separately fit logistic regression models to each cause groups, and then rescale the results to ensure that the total numbers of deaths estimated for each group match those reported.

The methods used enable health professionals to estimate any specific causes of deaths in countries where causes of death are of low quality and reliable cause of death from other sources are available.

References

Anderson, R., and Rosenberg, H.M. 1998. Age standardization of death rates: implementation of the year 2000 standard. National Vital Statistics Reports. 47(3), 1-17.

Benach, J., Yasui, Y., Borrell, C., Pasarin, M.I., Martinez, J.M. and Daponte, A. 2003. The public health burden of material deprivation: excess mortality in leading causes of death in Spain. Preventive Medicine. 36, 300-308.

Breslow, N.E. and Day, N.E. 1975. Indirect standardization and multiplicative models for rates, with reference the age adjustment of cancer and relative frequency data. Journal of Chronic Diseases, 28: 289-303.

Carmichael, G. 2011. Exploring Thailand's mortality transition with the aid of life tables. Asia Pacific Viewpoint. 52 (1), 85-105.

Choprapawon, C., Porapakkham, Y., Sablon, O., Panjajaru, R. and Jhantharatat, B. 2005. Thailand’s national death registration reform: verifying the causes of death between July 1997 and December 1999. Asia Pacific Journal of Public Health. 17(2),110-6.

Chutinantakul, A., Tongkumchum, P., Kanitta Bundhamcharoen. and Chongsuvivatwong, V. 2014. Correcting and estimating HIV mortality in Thailand based on 2005 verbal autopsy data focusing on demographic factors, 1996-2009. Population Health Metrics. 12, 25.

Court, B.V. and Cheng, K.K. 1995. Pros and cons of standardized mortality ratios. The Lancet. 346(25),1432.

Curtin, L. and Klein, R. 1995. Direct standardization (Age-adjusted death rates). Department of Health and Human Services. Public Health Service. 6, 1-11.

Davis, R.A. and Wu, R. 2009. Negative binomial model for time series of counts. Biometrika. 96(3), 735-749.

England, P.D. and Haberman, S. 1993. A new approach to modeling excess mortality. Journal of Actuarial Practice. 1(1), 85-117.

England, P.D., and Haberman, S. 1993. A new approach to modeling excess mortality. Journal of Actual Practice. 1(1), 85-117.

Fan, J., Upadhye. S. and Worster, A. 2006. Understanding receiver operating characteristic (ROC) curves. Canadian Journal of Emergency Medicine. 8(1), 19-20.

Faramnuayphol, P., Chongsuvivatwong, V. and Pannarunothai, S. 2008. Geographical variation of mortality in Thailand. Journal of Medical Association of Thailand. 91(9), 1455-1460.

Fottrell, E. and Byass, P. 2010. Verbal autopsy: methods in transition. Epidemiologic Reviews. 32, 38-55.

Fukuda, Y., Nakamura, K. and Takano, T. 2004. Increased excess deaths in urban areas: quantification of geographical variation in mortality in Japan, 1973-1998. Health Policy. 68, 233-244.

Health Research System Institute, Institute for Population and Social Research, Mahidol University. 2003. *Population Projection for Thailand, 2000-2025.* Bangkok, Thailand.

Hosmer, D.W. and S. Lemeshow, 2000. Applied Logistic Regression. 2nd Edn., Hoboken, NJ: Wiley.

Howel, J. 1995. Standardised mortality ratios. The Lancet. 346, 904.

Huong, D.L., Minh, H.V., Janlert, U., Van, D.D. and Byass, P. 2006. Socio-economic status inequality and major causes of death in adults: A 5-year follow-up study in rural Vietnam. Journal of the royal institute of public health. 120, 497-504.

Julious, S.A., Nicholl, J. and George, S. 2001. Why do we continue to use standardized mortality ratios for small area comparison?. Journal of Public Health Medicine. 23(1), 40-46.

Kaewsompak, S., Boonpradit, S., Choonpradab, C. and Chaisuksant, Y. 2005. Mapping acute febrile illness incidence in Yala province. Songklanakarin Medical Journal. 23(6), 455-462 .

King, G. and Lu, Y. 2008. Verbal autopsy methods with multiple causes of death. Statistical Science. 23(1), 78-91.

Kinjun, N., Lim, A., and Bundhamcharoen, K. 2015. A logistic regression model for estimating transport accident deaths using verbal autopsy data. Asia Pacific Journal of Public Health. 27(3), 286-292. doi: 10.1177/1010539514529810.

Li, Y., Tiwari, R.C., and Zou, Z. 2008. An age-stratified Poisson model for comparing trends in cancer rates across overlapping regions. Biometrical Journal. 50(4). 608-619.

Li, Y., Tiwari. and Zou, Z. 2008. An age-stratified Poison model for comparing trends in cancer rates across overlapping regions. Biometrical Journal. 50(4), 608-619. doi:10.1002/bimj.200710430.

Lim, A. and Choonpradub, C. 2007. A statistical method for forecaseting demographic time series counts, with application to HIV/AIDS and other infectious disease mortality in Southern Thailand. Southeast Asian Journal of Tropical Medicine and Public Health, 38(6): 1029-1040.

Mathers, C.D., Fat, D.M., Inoue M, Rao, C. and Lopez, D. 2005. Counting the dead and what they die from: an assessment of the global status of cause of death data. Bulletin of the World Health Organization. 83(3), 171-177.

McCullagh, P. and Nelder, J.A. 1989. Generalized linear models. London: Chapman and Hall.

McNamee, R. 2005. Regression modeling and other methods to control confounding. Occupational and Environmental Medicine. 62, 500-506. doi:10.1136/oem.2002.001115.

Ngo, A.D, Rao, C., Hoa, N.P., Adair, T. and Chuc, N.T.K. 2010. Mortality patterns in Vietnam, 2006: Findings from a national verbal autopsy survey. Bio Medical Central Research Notes. 3(78), 1-8.

Organization for Economic Co-operation Development/World Health Organization. 2013. Health at a Glance 2013: OECD indicators, OECD; 2013.

Organization for Economic Co-operation Development/World Health Organization. 2014. Health at a Glance: Asia/Pacific, 2014: Measuring Progress towards Universal Health Coverage, OECD; 2014.

Pattaraarchachai, J., Rao, C., Polprasert, W., Porapakkham, Y., Pao-in, W., Singwerathum, N. and Lopez, A.D. 2010. Cause-specific mortality patterns among hospital deaths in Thailand: validating routine death certification. Population Health Metrics. 8,12. doi: 10.1186/1478-7954-8-12.

Polprasert, W., Rao, C., Adair, T., Pattaraarchachai, J., Porapakkham, Y. and Lopez, A.D. 2010. Cause-of-death ascertainment for deaths that occur outside hospitals in Thailand: application of verbal autopsy methods. Population Health Metrics. 8,13. doi: 10.1186/1478-7954-8-13.

Porapakkham, Y., Rao, C., Pattaraarchachai, J., Polprasert, W., Vos, T., Adair, T. and Lopez, A.D. 2010. Estimated causes of death in Thailand, 2005: implication for health policy. Population Health Metrics. 8:14. doi: 10.1186/1478-7954-8-14.

Prasartkul, P. and Vapattanawong, P. 2006. The completeness of death registration in Thailand: Evidence from demographic surveillance system of the Kanchanaburi Project. World Health Population. 8, 43-51.

Rao, C., Lopez, A.D., Yang, G.H. 2005. Evaluating national cause of death statistics: Principles and application to the case of China. Bulletin of the World health Organization. 83(8), 618-25.

Rao, C., Porapakkham, Y., Pattaraarchachai, J., Polpraset, W., Swampunyalert, N. and Lopez, A.D. 2010. Verifying causes of death in Thailand: rationale and methods for empirical investigation. Population Health Metrics. 8(11). doi: 10.1186/1478-7954-8-13.

Robson, B., Purdie, G., Cram, F. and Simmonds, S. 2007. Age standardization-an indigenous standard?. Emerging Themes in Epidemiology. 4(3). doi:10.1186/1742-7622-4-3.

Rukumnuaykit, P. 2006. Mortality and causes of death in Thailand: evidence from the survey of population change and registration. Asia-pacific population journal. 21(2), 67-84.

Sakar, S. and Midi, H. 2010. Importance of assessing the model adequacy of binary logistic regression. Journal of Applied Sciences, 10(6): 479-86.

Sather, D., Fryer, G.E., Jr, G.E., McCann, J., Troutman, A., Woolf, S.H. and Rust, G. 2005. What if we were equal? A comparison of the Black-White mortality gap in 1960 and 2000. Health Affair. 24(2), 459-464.

Savesh, K., Shashank, S. and Ashish, A. 2013. Comparison of mortality trends and patterns between India and Japan. Indian Journal of Maternal and Child Health. 15(4), 1-9.

Setel, P.W., Sankoh, O., Rao, C., Velkoff, V.A., Mathers, C., Gonghuan, Y., Hemed, Y., Jha, P. and Lopez, A.D. 2005. Sample registration of vital events with verbal autopsy: a renewed commitment to measuring and monitoring vital statistics. Bulletin of the World Health Organization. 83, 611-617.

Setel, P.W.,Whiting, D.R., Hemed, Y., Chandramohan, D., Wolfson., L.J., Alberti, K. G.M.M. and Lopez, A.D. 2006. Validity of verbal autopsy procedures for determining causes of death in Tanzania. Tropical Medicine and International Health. 11, 681-696

Shah, N.M., Thalib, L. and Al-Ateeqi, A. 2014. Mortality differences according to marital status among men and women in Kuwait. Asian Population Studies. 10(3), 241-251. doi : 10.1080/17441730.2014.942944.

Stewart, J.A., Dundas, R., Howard, R.S., Rudd, A.G. and Wolfe, C.D.A. 1999. Ethnic differences in incidence of stroke: prospective study with stroke register. British Medical Journal. 318, 967-971.

Suwanwela, N. 2014. Stroke epidemiology in Thailand. Journal of Stroke, 16(1): 1-7.

Torrey, B. and Haub, C. 2004. A comparison of US and Canadian Mortality in 1998. Population and Development Review. 30(3), 519-530.

Tsai, S.P. and Wen, C.P. 1986. A review of methodological issues of the standardized mortality ratio (SMR) in occupational cohort studies. International Journal of Epidemiology. 15(1), 1-14.

Vapattanawong, P. and Prasartkul, P. 2011. Under-registration of deaths in Thailand in 2005–2006: results of cross-matching data from two sources. Bulletin of the World Health Organization. 89, 806-812.

Venables, W.N. and Ripley, B.D. Modern applied statistics with S. 4th ed. New York: Springer-Verlag; 2002.

World Health Organization. 2004. Special tabulation lists for mortality and morbidity: mortality tabulation list 1. In International Statistical Classification of Diseases and Health Related to Problems-Tenth Revision (Vol 1). Geneva, Switzerland: World Health Organization; 1163-75.

World Health Organization. 2011. Health statistics and information systems, 2011.

World Health Organization. 2013. World Health Statistics, 2013.

APPENDIX

3RD INTERNATIONAL CONFERENCE ON SOCIAL SCIENCES RESEARCH | ICSSR 2015

Venue: Meliá Hotel Kuala Lumpur, Malaysia | Dates: 8th & 9th June 2015

World Conferences Organized by .net

Certificate of Participation

This is to certify that

ARINDA MA-A-LEE

has participated as a

Presenter

at the 3rd International Conference On Social Sciences Research 2015
(ICSSR 2015)

Organized by
WorldConferences.net

on June 8 & 9, 2015
Meliá Hotel Kuala Lumpur, Malaysia

DR. MOKMIN BASRI
DEPUTY RECTOR (STUDENT DEVELOPMENT & ALUMNI)
INTERNATIONAL ISLAMIC UNIVERSITY COLLEGE SELANGOR
WorldConferences.net

WORLD CONFERENCE RESOURCES

WORLD CONFERENCES.NET WCR

Research Collaborations

Phillip Hannah
HR Administration Officer, HR Service Centre

16 January 2015

Ms Arinda Ma-a-lee
Department of Mathematics and Computer Science
Faculty of Science and Technology,
Prince of Sogkla University Thailand

Private and Confidential

Dear Ms Ma-a-lee,

I would like to congratulate you on the award of the title of Visiting Fellow to pursue research activities in NHMRC Clinical Trials Centre, the University of Sydney Medical School (the **affiliation**).

The title is subject to you continuing to contribute to the activities of the NHMRC Clinical Trials Centre as required by the Director of the Centre, currently Professor John Simes who will negotiate your academic duties with you.

The title is effective from the 22 February 2015 to 22 May 2015. The title may be renewed for further periods provided you continue to contribute to the Centre activities.

When using an honorary tile, the full title must be used.

I would be grateful if you could confirm your acceptance of the affiliation under the terms contained in the attached schedules by signing a copy of this letter and returning it to:

HR Service Centre
Human Resources
Level 2, 1-3 Ross Street (K06)
University of Sydney NSW 2006

Fax: +61 2 8627 1343
Email: hr.servicecentre@sydney.edu.au

The faculty look forward to the commencement of your affiliation.

Yours sincerely,

On behalf of Professor Bruce Robinson
Dean, Sydney Medical School

THE UNIVERSITY OF SYDNEY

Phillip Hannah
HR Administration Officer, HR Service Centre

04 September 2015

Ms Arinda Ma-A-Lee
Department of Mathematics and Computer Sciences
Faculty of Science and Technology
Prince Songkla University
Thailand

Private and Confidential

Dear Ms Ma-A-Lee,

This letter supersedes the letter issued 07 May 2015 due to the delay as a result of visa being granted late.

I would like to congratulate you on the award of the title of Visiting Fellow to pursue research activities in NHMRC Clinical Trials Centre, the Sydney Medical School, the University of Sydney (the **affiliation**).

The title is subject to you continuing to contribute to the activities of the NHMRC Clinical Trials Centre as required by the Director of the Centre, currently title/name who will negotiate your academic duties with you.

The title is effective from the 27 July 2015 to 22 October 2015. The title may be renewed for further periods provided you continue to contribute to the School activities.

When using an honorary tile, the full title must be used.

I would be grateful if you could confirm your acceptance of the affiliation under the terms contained in the attached schedules by signing a copy of this letter and returning it to:

> HR Service Centre
> Human Resources
> Level 2, 1-3 Ross Street (K06)
> University of Sydney NSW 2006
>
> Fax: +61 2 8627 1343
> Email: hr.servicecentre@sydney.edu.au

The faculty look forward to the commencement of your affiliation.

Yours sincerely,

On behalf of Professor Bruce Robinson
Dean, Sydney Medical School

Vitae

Name: Miss Arinda Ma-a-lee

Student ID: 5420330008

Educational Attainment:

Degree	**Name of institution**	**Year of Graduation**
B.Sc. (Mathematics)	Thaksin University	2002
M.Sc. (Research Methodology)	Songkla University	2006

Scholarship Awards during Enrolment

- This research was supported by grant funds from the program Strategic Scholarships Fellowships Frontier Research Networks (Specific for Southern region) for the Ph.D. Program Thai Doctoral degree from Office of the Higher Education Commission, Thailand. Research scholarship from the Centre of Excellence in Mathematics, the Commission on Higher Education, Thailand

- Scholarship for proceeding from the Graduate school and Faculty of Science and Technology, Prince of Songkla University, Songkhla, Thailand and Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Thailand

Work-Position and Address:

Lecturer, Prince of Songkla University

List of Publication and Proceeding:

Publication:

Ma-a-lee, A., Tongkumchum, P. and Ueranantasan, A. A comparison of cause-specific deaths in Thailand in 2005 with Japan in 2006, submitted to the *Pertanika Journal of Social Science & Humanities.*

Ma-a-lee, A., Tongkumchum, P. and Pipatjaturon, N. A statistical methods for comparing stroke death rates in Suphan Buri with Bangkok in 2005, submitted to the *Journal of Research in Health Sciences.*

Ma-a-lee, A., Pipatjaturon, N. and Tongkumchum, P. A statistical methods for correcting misreported multinomial outcome with application to stroke mortality in Thailand, submitted to the *Walailak Journal of Science and Technology*.

Proceeding:

Arinda, M. A comparison of deaths in Bangkok in 2005 with Japan in 2006. *The 3rd international conference on social sciences research 2015 (ICSSR 2015)* on 8 & 9 June at Melia Hotel Kuala Lumpur, Malaysia.