Appendix I Proceeding

Proceedings topic: Estimating HIV Deaths in Thailand in 2005 Using Verbal Autopsy Modeling.

Place: The 5th International Conference on Public Health among Greater Mekong Sub-Regional Countries. Yangon, Myanmar.

Time: 28-29 September, 2013

Estimating HIV Deaths in Thailand in 2005 Using Verbal Autopsy Modeling

Amornrat Chutinantakul [1], Phattrawan Tongkumchum [2] and
Virasakdi Chongsuvivatwong [3]

[1]Department of Mathematics and Computer Science; Faculty of Science and Technology, Prince of Songkla University, Muang, Pattani, 94000 Thailand
Tel: 66878828589 E-mail: yaiamorn@yahoo.com
[2]Department of Mathematics and Computer Science; Faculty of Science and Technology, Prince of Songkla University, Muang, Pattani, 94000 Thailand
E-mail: tphattra@bunga.pn.psu.ac.th
[3] Epidemiology Unit ; Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkla, 90000 Thailand
E-mail: cvirasak@medicine.psu.ac.th

Abstract

Background: Mortality data are important in disease measurement and health status indication. They reflect quality of life, life expectancy and burden of disease at local and national levels. Vital registrations (VR) in Thailand are of poor quality, not only lacking completeness but also inaccurate for cause of death. The precise numbers of HIV deaths are not known, but HIV deaths from VR data are under-reported. Thai Working Groups on HIV/AIDS projections 2005 estimate HIV mortality using Asian Epidemic Model (AEM) method but gave results exceeding VR data by a factor of 2.38.

Objective: To accurately estimate HIV reporting deaths in Thailand 2005 based on the 2005 VA study.

Methods: Verbal autopsy (VA) is used to determine individual cause of death. The latest VA study in Thailand was carried out in 2005. We investigated HIV deaths in Thailand reported 2005, using 9,644 cases of deaths from the 2005 VA study to accurately assess primary cause, fitting logistic regression models with province, gender-age group and reported cause in and outside hospital as factors. Receiver

Operating Characteristic (ROC) curve is used to assess the model’s ability. A triangulation method is used to interpolate province coefficients outside the VA study in Thailand.

Results: Based on 2005 VA data, the logistic regression model shows all factors are statistically significant. The ROC curve illustrated a good model prediction. The model gave sensitivity 68.9% and specificity 98.3% whereas VR reports gave sensitivity 31.1%. We applied the model to VR data in 2005 and found substantial under-representation in HIV. The model gave results exceeding VR data by a factor of 2.78 whereas Porapakkham’s study gave results a factor of 4.

Conclusions: These methods can substantially improve the quality of VR data based on VA surveys**.** Results from this study provide valuable information for Ministry of Public Health in Thailand, and also can improve the quality of VR data based on VA surveys in other developing countries. Further analysis could usefully apply this model to other years and focus on forecasting cause of death.

Keywords: HIV mortality, verbal autopsy, vital registration


5th
ICPH-GMS
2013
5th
INTERNATIONAL CONFERENCE ON PUBLIC HEALTH
AMONG GMS COUNTRIES
Evidence-based Health Policy: Current Status and Future Prospects
Venue : National Theater
University of Public Health
University of Medicine (1)
28 - 29 September, 2013

5th International Conference on Public Health among Greater Mekong Sub-Regional Countries

28-29 SEPTEMBER, 2013

UNIVERSITY OF PUBLIC HEALTH, YANGON, MYANMAR

LETTER OF INVITATION

Amornrat Chutinantakul
PhD Student,
Department of Mathematics and Computer Science,
Faculty of Science and Technology
Prince of Songkla University,
Thailand

Date: 10 August, 2013

On behalf of the Organizing Committee, it is a great pleasure for me to inform you that your abstract entitled '**Estimating HIV Deaths in Thailand in 2005 Using Verbal Autopsy Modeling'** has been reviewed and accepted for ORAL presentation at 5th International Conference on Public Health among Greater Mekong Sub-regional Countries at University of Public Health, Yangon, Myanmar on 28 – 29 September, 2013. Therefore you are cordially invited to participate in the above mentioned conference. Exact date and time of oral presentation will be announced later.

Yours sincerely,

Professor Dr Nay Soe Maung
Rector of University of Public Health
Chairman of Organizing Committee, 5th ICPHGMS
Postal address: Corner of Bogyoke Aung San Road and Myoma Kyaung Road, Latha Township, Yangon, Myanmar 11131
website: http://www.uph-myanmar.org
e-mail naysoemg26@gmail.com
Off Ph +951 395207, Fax +951 395212

Certificate of Presentation

This certificate is awarded to

Amornrat Chutinantakul

in recognition of presentation in

"5TH INTERNATIONAL CONFERENCE ON PUBLIC HEALTH AMONG GMS COUNTRIES"

28th - 29th September, 2013

Yangon, Myanmar

Prof. Nay Soe Maung

Rector

University of Public Health

Yangon, Myanmar

Appendix II Article

“Estimation of mortality with missing data using logistic regression”

Songklanakarin J. Sci. Technol.
36 (2), 249-254, Mar. - Apr. 2014

SJST SONGKLANAKARIN JOURNAL OF SCIENCE AND TECHNOLOGY
http://www.sjst.psu.ac.th

Original Article

Estimation of mortality with missing data using logistic regression

Amornrat Chutinantakul[1,2], Marzukee Mayeng[2], and Phattrawan Tongkumchum[2]*

[1] *Office of Disease Prevention and Control, Region 11 Nakhon Si Thammarat, Mueang, Nakhon Si Thammarat, 80000 Thailand.*

[2] *Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Mueang, Pattani, 94000 Thailand.*

Received 16 June 2013; Accepted 22 January 2014

Abstract

The present study aims to improve estimation of mortality data with unknown demographic factors using logistic regression, based on inflation factors for distributing such cases. The method is illustrated on death proportions based on numbers of deaths reported in 1996-2009 classified by gender, age and province in the death registration (DR) system. The results indicate that cases with unknown province mostly occurred in ages 0-4 years and 15-44 years and cases with unknown age mostly occurred in the central and southern regions. The method is straightforward, provides confidence intervals, and can be generally used for eliminating biases due to cases with unknown values of demographic factors in DR data. The resulting estimated numbers of deaths were used to examine age-specific mortality using cubic spline function. Cubic spline interpolation is effective to smoothly interpolate non-negative mortality data. These methods provide valuable information to the Ministry of Public Health.

Keywords: estimation, mortality, unknown demographic factors, logistic regression, cubic spline

1. Introduction

In vital registration surveys it is quite common for reports to have missing or unknown values for important variables. Yearly death certificates in the death registration (DR) database for the years 1996-2009 from the Ministry of Interior in Thailand had between 0.01% and 6.82% of records with unknown province and between 0.05% and 4.78% with unknown age. Although these percentages are relatively small and have decreased substantially since 2005 to 2009, it is important to have appropriate methods to adjust for biases that arise in such cases. Note that if cases with missing values are omitted from analysis under the possibly valid assumption that they are representative of the population, consequent estimates of incidence rates will be biased downwards unless appropriate inflation factors are applied.

Gender and age-specific mortality rates are obtained by dividing numbers of deaths for each sex and age group by corresponding populations (usually 1000s or 100,000s). Death rates in two regions with different age patterns are then compared by applying these rates to the same population. Based on such understanding of how Thai population distributions vary geographically throughout Thailand, we consider variation in numbers of deaths (the numerator in the mortality rate formula). Before examining these patterns, it is important to focus on data quality. As pointed out by Carmichael (2011) "significant under-registration and age-selective under-enumeration" occur in Thailand.

Although there is a sizable literature on statistical methods for handling missing data (see Little and Rubin, 2002), these methods are not directly appropriate for correcting incidence rates as described above. For subjects with unknown age in cancer registries, Fallah and Kharazmi (2008) compared several methods, finding them to be superior to

* Corresponding author.
Email address: tphattra@bunga.pn.psu.ac.th

conventional methods of simply inflating by a constant or allocating the unknown ages to the oldest age group. Although similar methods could be applied to categorical factors other than age group, including nominal grouped data such as province or district (using the size of the group to rank the data), these methods are essentially *ad hoc*.

Logistic regression (McNeil, 1996; Woodward 1999) provides a more promising general method for handling predictors with unknown values when estimating incidence rates, because the problem can be formulated in terms of a binary outcome (unknown or known value of the predictor) with other known factors as predictors. This method has been used to address missing data in other studies, including case-control studies (see, for example, Li *et al.,* 2004). An extensive study by Williams *et al.* (2005) approached the problem using a Bayesian formulation of logistic regression to calculate the posterior distribution by integrating out missing data. In this paper we describe a simple method for correcting mortality data in the presence of cases with unknown factors, based on the proportions of missing data with known factors for other variables. It involves fitting a logistic regression model and thus provides estimates and confidence intervals for proportions. The method is illustrated using all-cause mortality data from Thailand in 1996-2009.

2. Data and Methods

The DR database in year 1996-2009 was provided by the Bureau of Registration Administration, Ministry of Interior and coded as cause-of death by the Bureau of Policy and Strategy (2010), Ministry of Public Health. The reported deaths were classified in 18 five-year age groups up to 85-89 and 90 or more by gender in 76 provinces in Thailand. The numbers of deaths reported were tabulated into cells based on all combinations of the two sexes, the 77 groups comprising the provinces and an additional "unknown province" group, and 20 groups comprising the age groups and a further "unknown age" group.

The basic method involves fitting the generalized linear model from the binomial family with logit link by maximum likelihood as described in Venables and Ripley (2002). Logistic regression formulates the logit of the probability that a person died with unknown province or unknown age group as a linear function of the determinant factors. For the data considered, the binary outcome was defined as either (a) unknown or known province, or (b) unknown or known age. In each case the model fitted was an additive combination of two factors. The models are formulated as

$$logit\left(P_{ij}\right) = ln\left(\frac{P_{ij}}{1-P_{ij}}\right) = \mu + \alpha_i + \beta_j, \qquad \text{(a)}$$

where P_{ij} is the probability of death with unknown province, μ is a constant, α_i and β_j are individual parameters specifying gender i and age group j (j=1,2,3,…,19), respectively.

$$logit\left(P_{ik}\right) = ln\left(\frac{P_{ik}}{1-P_{ik}}\right) = \mu + \alpha_i + \gamma_k, \qquad \text{(b)}$$

where P_{ik} is the probability of death from unknown age group, α_i and γ_k are individual parameters specifying gender i and province k (k=1,2,3,…,76), respectively. The data for each model comprised all cases with known age or province, and thus only those cases with both province and age unknown were omitted. The models were fitted sequentially with the total numbers of known deaths in provinces for model (b) inflated and rounded to integers using results from model (a). The adequacy of fit of each model was assessed by using a plot of deviance residuals versus theoretical quantiles.

The adjusted proportions from model (a) were used to correct deaths in each age group with unknown province by multiplying the reported numbers in gender i and age group j by $1/(1-P_{ij})$, where P_{ij} is the estimated proportion of deaths with unknown province in gender i (i =1 for males) and age group j. Thus, the estimated numbers of deaths were obtained.

In the same way, we corrected deaths in each province with unknown age group by multiplying the estimated numbers of deaths from model (b) in gender i and province k by $1/(1-P_{ik})$, where P_{ik} is the estimated proportion of deaths with unknown age group in gender i and province k. These procedures thus redistribute reported deaths with unknown province for each gender and age group to known province and redistribute reported deaths with unknown age group for each gender and province to known age groups.

The adjusted proportions of deaths with unknown province or unknown age were presented using graphs of confidence intervals. Since it is more appropriate to compare province effects with their overall mean, rather than with an arbitrary province, the standard errors for the estimated parameters in the model are based not on the conventional treatment contrasts but on weighted sum contrasts. A method for doing this is described by Tongkumchum and McNeil (2009), with thematic mapping of regions based on this method described by Odton *et al.* (2010a,b).

After allocation of deaths with unknown province or unknown age, natural cubic splines were used to interpolate age-specific demographic data to ensure that relevant boundary conditions on second derivatives are satisfied as described by McNeil *et al.* (2011). Age-specific demographic data functions are necessarily non-negative. This method can be used to smoothly interpolate non-negative mortality data. "Cubic spline interpolation is a useful technique to interpolate between known data points due to its stable and smooth characteristics" (Kruger, 2003).

All data analysis was undertaken and graphical displays created using basic R software (R Development Core Team, 2012).

Table 1. Number of deaths and percentage unknown of province and age, 1996-2009

year	number of deaths	percentage unknown			
		province	age	both	total
1996	342,643	0.44	4.78	0.00	5.22
1997	300,321	1.06	2.72	0.14	3.92
1998	310,535	1.60	0.20	0.00	1.80
1999	362,607	1.15	0.89	0.00	2.05
2000	365,741	2.03	0.42	0.00	2.45
2001	369,494	0.34	0.30	0.00	0.65
2002	380,364	3.53	0.25	0.07	3.86
2003	384,131	4.27	0.24	0.01	4.52
2004	393,592	6.82	0.19	0.01	7.02
2005	395,374	0.51	0.35	0.00	0.86
2006	391,126	0.39	0.32	0.00	0.71
2007	393,254	0.04	0.20	0.00	0.23
2008	397,327	0.02	0.06	0.00	0.08
2009	393,916	0.01	0.05	0.00	0.06

3. Results

Annual numbers of deaths from 1996 to 2009 varied from 300,321 to 397,327 cases. Percentages of DR data for numbers of deaths with an unknown province and age are shown in Table 1. The highest unknown province is in the year 2004 whereas unknown age is in the year 1996. The percentages of deaths with unknown province or age are negligible after 2006, but need correction in earlier years.

Logistic regression models provide an acceptable fit for both proportions of unknown province and age. Figure 1 shows logistic regression modeling results from model (a). The graphs show adjusted percentages of deaths with unknown provinces in 1997 (upper panel) and 2004 (lower panel). The 95% confidence intervals are plotted for comparing estimated percentages of deaths with unknown provinces from the overall mean by gender and age group (red line). Overall percentage of unknown province deaths increased from 1.1% to 6.8% from 1997 to 2004.

The pattern of unknown provinces in 1997 and 2004 are similar. Percentage of unknown provinces above average mostly occurred in the worker age groups, 15-44 years, including the highest in age group 0-4 years. Unknown provinces below average occurred in ages over 60 years. The graphs also indicate that the percentage of deaths with unknown province for males were slightly lower than that for females in 1997 and higher in 2004.

Figure 2 graphs similar results from the logistic regression model (b). The model estimates the percentages of deaths with unknown age by gender and province. The provinces with codes of 10-27 and 70-77 are in the Central region, 30-49 in the North-East, 50-67 in the North and 80-96 in the South. In 1997, over 15% of female unknown age deaths occurred in Rayong (code 21) and Chanthaburi (code 22) provinces (upper panel). Unknown-age deaths decreased from 2.7% to 0.2% from 1997 to 2004 (red line). In 2004, most provinces had unknown age around the mean and below 1%

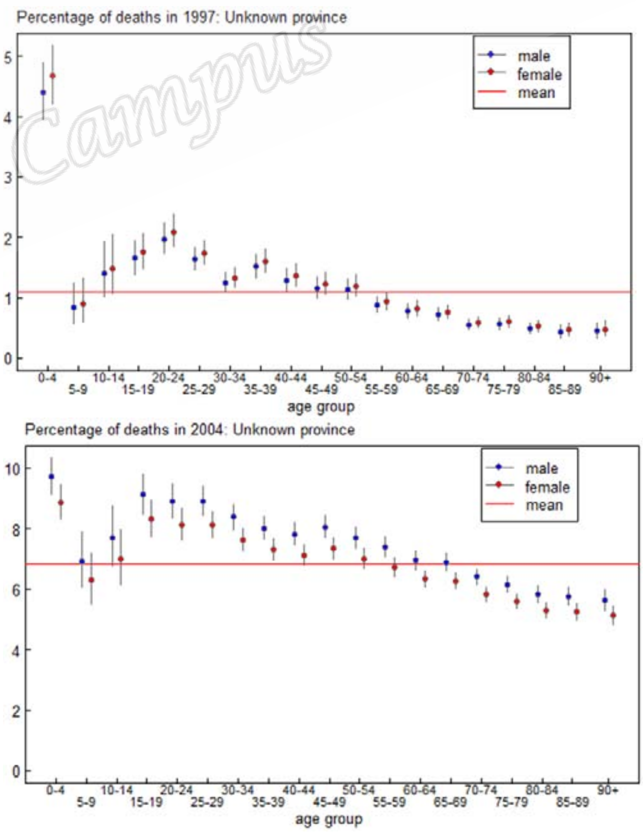

Figure 1. Estimates with 95% confidence intervals for percentages of deaths with unknown province in Thailand in 1997 (upper panel) and 2004 (lower panel)

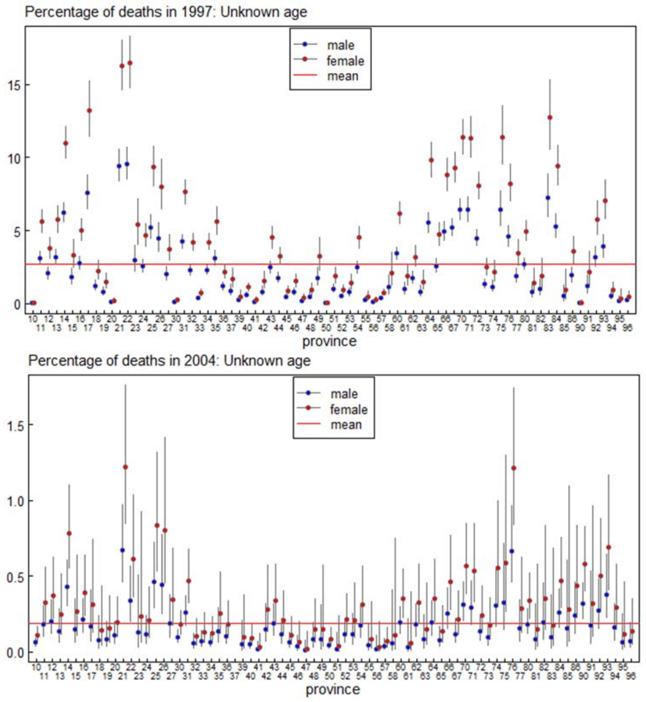

Figure 2. Estimates with 95% confidence intervals for percentages of deaths with unknown age in Thailand in 1997 (upper panel) and 2004 (lower panel)

with the exception of females in Rayong (code 21) and Phetchaburi (code 76) provinces (lower panel). Unknown-age deaths were reported to a greater extent for females.

The thematic maps clearly illustrate unknown age changed for both genders in each province from 1997 to 2004 as shown in Figure 3. They separate three levels of these percentages into groups where the confidence intervals are entirely above the mean (coloured in red), crossing the mean (orange) and entirely below the mean (cyan). In 1997 (left panel), nineteen provinces had above-average unknown age percentage occurring mainly in the provinces of the Central and the Southern regions. Lower than average percentages of reported deaths with unknown age mostly occurred in the North and North-East provinces. In 2004 (right panel), nine provinces had above-average unknown age percentages. Seven provinces in the central and Phatthalung remained unchanged from 1997 to 2004, whereas the percentage for Songkhla changed from below-average to above-average. Provinces below-average remained unchanged in Chiang Mai, Chiang Rai, Udon Thani, Sakon Nakhon, Khon Kaen, Bangkok and Amnat Charoen (no unknown age deaths in 2004).

Missing data cause biases and can lead to misunderstanding of information. The methods in this study applies to data with unknown province or age and thus provide estimated numbers of deaths for further investigation of mortality patterns and trends. This method allocates numbers of deaths and provides estimation with 95% confidence intervals.

We graphed the changes in smoothed age distributions of deaths for 76 provinces by fitting cubic spline to cumulative death counts after allocation of deaths with unknown province or age. It shows male peaks occurred around age 25 in several provinces in 1996 including Chiang Mai, Lumphun, Lampang, Payao, Chiang Rai, Phetchaburi and Ranong, but these peaks had disappeared by 2009.

Figure 4 shows three patterns of trends in spline-smoothed deaths from 1996-2009 in six provinces. First, Narathiwat and Phuket show increasingly high levels of infant mortality. Second, Bangkok and Khon Kaen have quite similar patterns, with minor peaks between ages 20 and 30. Third, Chiang Rai and Chiang Mai have very much higher peaks in this age range. Some anomalies are apparent. Infant mortality levels for Phuket in 1998 and 1999 were about three times higher than in other years, whereas no infant deaths were reported in Bangkok in 1997, probably due to incomplete data reporting.

We compared male and female patterns in Narathiwat, Bangkok and Chiang Rai provinces as shown in Figure 5. Numbers of female deaths are generally lower, but the comparison is distorted by the greater numbers of male deaths in ages 20-40 in Chiang Rai and to a lesser extent in Bangkok.

The total number of deaths in any year for a province is simply the area under its curve. For example, the excess number of male deaths in age group 20-40 years for Chiang Rai in 1996 is approximated by the triangle with area 240× 30/2=3600. The triangle is superimposed on the curve for male deaths in 2009 and shaded in yellow (top right panel). A triangle with similar area is also superimposed on the curve for female deaths in 2009 (bottom right panel). It shows that the number of male deaths in the age group 20-40 in 1996 is similar to the number of female deaths in all age groups in 2009.

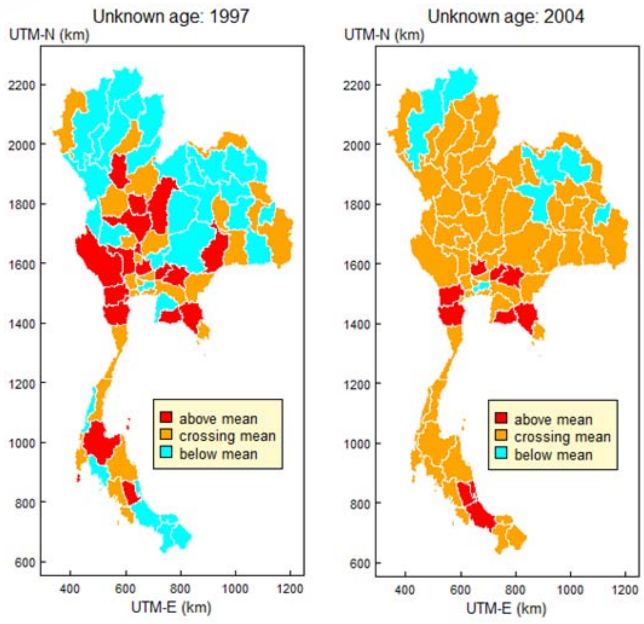

Figure 3. The thematic map of unknown age in 1997 and 2004

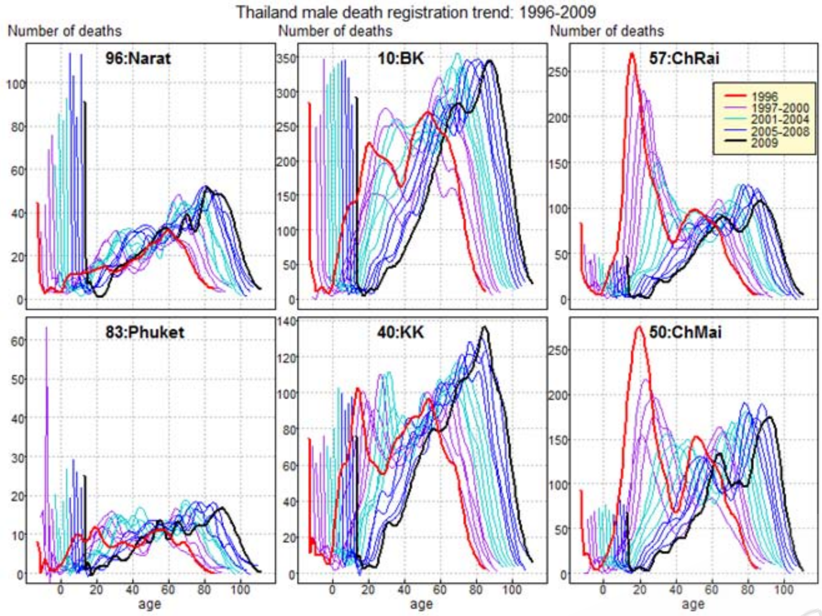

Figure 4. Trends in spline-smoothed deaths in six provinces from 1996-2009

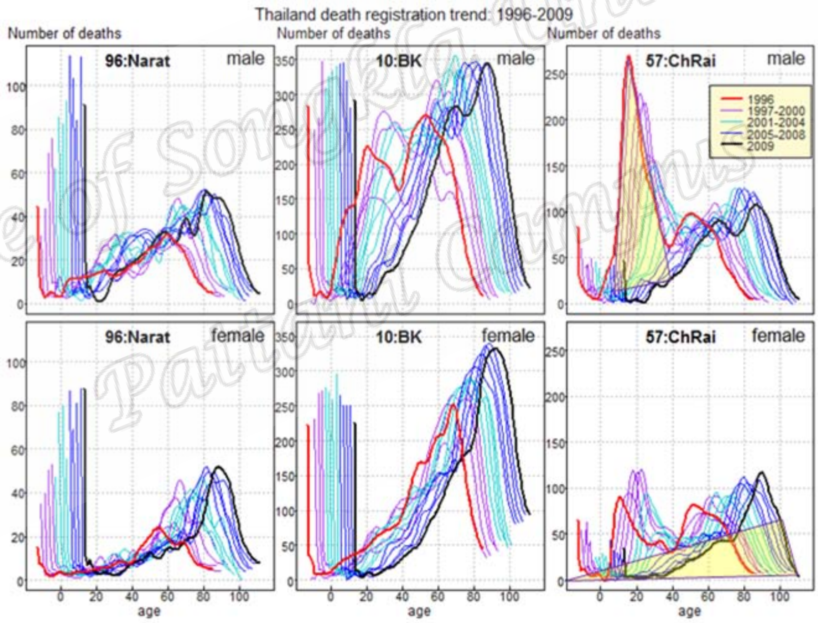

Figure 5. Trends in spline-smoothed deaths by gender from 1996-2009

4. Discussion

Reasons for the absence of demographic details on death certificates could be due to persons dying away from relatives and without ID cards. The age distribution of deaths with an unknown province of residence mostly occurred in worker age groups, suggesting that many such cases could be due to injuries occurring outside the province of residence or migration. The number of deaths with unknown province increased from 1996 to 2004, possibly partially due to migration (Vapattanawong and Saplon, 2011). Thus it would be informative to investigate such cases further. Our finding that deaths with unknown age were higher among females also needs further investigation.

Since death certificates in Thailand also provide information of the deceased person's age, gender and district (available from the Ministry of Interior), it is possible that further accuracy could be gained by using the 926 districts into which the nation is divided, rather than the 76 provinces. Faramnuayphol *et al.* (2008) used district as the location factor in their regional comparison of cause-specific mortality, and Odton *et al.* (2010b) used 235 aggregated districts with similar populations in their study of regional variation of age-group and gender-adjusted all-cause mortality. However, neither of these studies made adjustments for deaths reported with unknown region or age. Although in a subsequent study Odton *et al.* (2010a) investigated regional variations in deaths reported with unknown cause, correcting for unknown cause in mortality data is essentially a different problem, because in this case there is no cause-specific population denominator.

Logistic regression is a method to correct a misclassification binary outcome. It is appropriate for dealing with missing data. Many researchers have used this method to correct for missing data in their studies (Duffy *et al.*, 2004; Li *et al.,* 2004; Lyles *et al.,* 2011). For simplicity and conciseness in smoothed age distributions of deaths, cubic spline interpolation is an efficient strategy for smoothing cumulative death counts (Wand, 2000; McNeil *et al.,* 2011). The methods in this study minimize the effects of less than perfect data and provide more accurate estimation of mortality with 95% confidence intervals. They also provide valuable information to the Ministry of Public Health for efficient health policy planning.

Acknowledgements

We thank Prof. Dr. Don McNeil, Professor Emeritus of Statistics at Macquarie University in Australia, for his helpful guidance. We also thank the Bureau of Registration Administration in the Ministry of Interior and the Bureau of Policy and Strategy in the Ministry of Public Health for providing the data.

References

Bureau of Policy and Strategy. 2010. Vital Registration Database, Ministry of Public Health, Thailand.

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

Duffy, S.W., Warwick, J., Williams, A.R.W., Keshavarz, H., Kaffashian, F., Rohan, T.E., Nili, F. and Sadeghi-Hassanabadi, A. 2004. A simple model for potential use with a misclassified binary outcome in epidemiology. Journal Epidemiology Community Health. 58, 712-717.

Fallah, M. and Kharazmi, E. 2008. New methods of handling cases of unknown age in cancer registry data. Asian Pacific Journal of Cancer Prevention. 9(2), 259-262.

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

Kruger, C.J.C. 2003. Constrained cubic spline interpolation for chemical engineering Applications. Available from: http://www.korf.co.uk/spline.pdf. [November 28, 2012]

Li, X., Song, X. and Gray, R.H. 2004. Comparison of the missing-indicator method and conditional logistic regression in 1:m matched case-control studies with missing exposure values. American Journal of Epidemiology. 159(6), 603-610.

Little, R.J.A. and Rubin, D.B. 2002. Statistical Analysis with Missing Data, Wiley-Interscience, New York, U.S.A.

Lyles, R.H., Tang, L., Superak, H.M., King, C.C., Celentano, D.D., Lo, Y. and Sobel, J.D. 2011. Validation data-based adjustments for outcome misclassification in logistic regression: An illustration. Journal of Epidemiology. 22(4), 589-597.

McNeil, D. 1996. Epidemiological Research Methods, Wiley, Chichester, England, pp. 125-194.

McNeil, N., Odton, P. and Ueranantasun, A. 2011. Spline interpolation of demographic data revisited. Songklanakarin Journal of Science and Technology. 33(1), 117-120.

Odton, P., Bundhamcharoen, K. and Ueranantasun, A. 2010a. District-level variations in quality of mortality data in Thailand. Asia-Pacific Population Journal. 25(1), 79-91.

Odton, P., Choonpradub, C. and Bundhamcharoen, K. 2010b. Geographical variations in all-cause mortality in Thailand. Southeast Asian Journal of Tropical Medicine and Public Health. 41(5), 1209-1219.

R Development Core Team. 2012. The R project for statistical computing. Applications. Available from: http://www.r-project.org [September 9, 2012]

Tongkumchum, P. and McNeil, D. 2009. Confidence intervals using contrasts for regression model. Songklanakarin Journal of Science and Technology. 31(2), 151-156.

Vapattanawong, P. and Saplon, O. 2011. Deaths outside residential area of Thais: study from death registration, 1996-2009. Thai Population Journal. 3(1), 73-89.

Venables, W.N. and Ripley, BD. 2002. Modern Applied Statistics with S, Springer, New York, U.S.A., pp. 183-199.

Wand, M. P. 2000. A comparison of regression spline smoothing procedures. Computational Statistics. 15(4), 443-462.

Williams, D., Liao X., Ya, X. and Carin, L. 2005. Incomplete-data classification using logistic regression. Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, August 7-11, 2005, 972-979.

Woodward, M. 1999. Epidemiology: Study Design and Data Analysis, Chapman & Hall, U.S.A., pp. 448-481.

Appendix III Article

"Correcting and estimating HIV mortality in Thailand based on 2005 verbal autopsy data focusing on demographic factors, 1996-2009"

Chutinantakul et al. Population Health Metrics 2014, 12:25
http://www.pophealthmetrics.com/content/12/1/25

POPULATION HEALTH METRICS

RESEARCH Open Access

Correcting and estimating HIV mortality in Thailand based on 2005 verbal autopsy data focusing on demographic factors, 1996-2009

Amornrat Chutinantakul[1,2], Phattrawan Tongkumchum[2], Kanitta Bundhamcharoen[3] and Virasakdi Chongsuvivatwong[4*]

Abstract

Background: It is known that death registry (DR) underestimates HIV deaths. The objectives of this study were to examine under-reporting/misclassification and to estimate HIV mortality in Thailand during 1996-2009 from a model based on 2005 verbal autopsy (VA) data.

Methods: Logistic regression was used to predict HIV deaths from the VA dataset with and without demographic covariates. This full model was then used to predict individual HIV deaths from the DR dataset of provinces in which VA was conducted. The proportions in the remaining provinces were predicted from spatial interpolation based on coefficients of the VA provinces.

Results: Area under Receiver Operating Characteristic curve of the full model was 0.969 compared to 0.879 of the simple cross-referencing model when demographic covariates were not included. DR-reported HIV deaths accounted for only one-third of all VA-estimated HIV deaths. The most misclassified HIV deaths were those registered as tuberculosis and mental and nervous system. Under-reporting was most common among females and people aged 20-39 years, and effect of province was highest in the upper north and upper south regions.

Conclusions: For approximately two-thirds of all HIV deaths estimated by the full model, the causes were reported under other categories, not HIV. Demographic variables are essential for accurately correcting causes of death from death registries.

Keywords: HIV death, Verbal autopsy, Death registry, Under-reporting, Misclassification, Estimation

Background

Inaccurate and unreliable attribution of causes of death is especially high where causes are ill-defined. Under-reporting of HIV deaths is common in developing countries, and has been documented in Botswana [1], Brazil [2,3], South Africa [4,5], and Thailand [6-8], due to under-registered deaths and misclassification of cause of death. These problems severely limit the value of routine mortality data for public utility and affect resource allocations by policymakers.

Death registry (DR) data in Thailand provided by the Ministry of Interior through the Bureau of Policy and Strategy, Ministry of Public Health, are of poor quality, not only lacking completeness but also providing inaccurate cause of death [6,7,9-13]. Verbal autopsy (VA) surveys have been widely used in several countries including Uganda, China, Brazil, Tanzania, Bangladesh, South Africa, Zimbabwe, and Thailand to give more accurate information about causes of death [12,14,15]. The latest VA study in Thailand was carried out in 2005 in nine provinces by the SPICE (Setting Priorities using Information on Cost- Effectiveness analysis) project. The 2005 VA study [8,16-18] was used to estimate various causes of death including HIV. However, the simple cross-referencing method used in these studies ignored the effect of sex-age groups and locality of the deceased, which could give incorrect estimates due to confounding.

* Correspondence: cvirasak@medicine.psu.ac.th
[4]Epidemiology Unit, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla, Thailand
Full list of author information is available at the end of the article

We hypothesized that the utility of the 2005 VA data can be substantially improved if demographic variables were included to predict the cause of death. The aims of our study were 1) to examine under-reporting and misclassification of HIV deaths, based on modeling of the 2005 verbal autopsy data and 2) to estimate HIV deaths in all provinces of Thailand during 1996-2009.

Methods

Data sources and management

This study was confined to deaths of people aged 5 years and older, for which HIV death is common and often misclassified. DR data from 1996-2009 were obtained from the Bureau of Policy and Strategy database, Ministry of Public Health. The 2005 VA study was conducted by the SPICE project, and included a sample of 9,644 deaths (3,316 in-hospital and 6,328 outside-hospital) from 28 selected districts in nine provinces of four regions, of which 9,495 were deaths of persons aged 5 years and older.

Table 1 summarizes the cause groups based on VA counts. Accordingly, the chapter-block classifications of ICD-10 codes [19], consisting of blocks categorized mainly by human organs, were used to create 21 major cause groups for deaths at ages 5 years and older based on the distribution of VA-assessed deaths. For statistical accuracy, groups with small counts (mainly less than 200) were combined into larger groups using medical considerations (apart from septicemia, which received special attention due to over-reporting). The proportion of all deaths represented by these categories varied from 0.8% for septicemia to 11.3% for stroke and 5.4% for HIV deaths (ICD 10 code B20-24), as shown in Table 1.

Table 1 Cause groups based on VA counts

Cause groups	VA count	Percentages
1: TB (A15-19)	195	2.1
2: Septicemia (A40-41)	77	0.8
3: HIV (B20-24)	512	5.4
4: Other Infectious (A, B)$^{-}$	219	2.3
5: Liver Cancer (C22)	500	5.3
6: Lung Cancer^{+}(C30-39)	320	3.4
7: Other Digestive Cancer (C15-26^{-})	290	3.1
8: Other Cancer (C^{-}, D0-48)	697	7.3
9: Endocrine (E)	647	6.8
10: Mental, Nervous (F, G)	223	2.3
11: Ischemic (I20-25)	617	6.5
12: Stroke (I60-69)	1,076	11.3
13: Other CVD (I^{-})	540	5.7
14: Respiratory (J)	801	8.4
15: Digestive (K)	489	5.2
16: Genitourinary (N)	412	4.3
17: Ill-defined (R)	501	5.3
18: Transport Accident (V)	536	5.6
19: Other Injury (W, X0-59)	327	3.4
20: Suicide (X60-84)	158	1.7
21: All other	358	3.8
Total	9,495	100.0

$^{+}$*Respiratory/thoracic,* $^{-}$*exclude above.*

Misclassification was not at random. The effects of sex, age, and spatial variables were used to correct misclassification, using logistic regression. For efficiency, the predictors were optimally grouped to obtain sufficient sample size for relatively homogeneous risk groups. Nine provinces were included in the VA study (Bangkok, Nakhon Nayok, Suphan Buri, Ubon Ratchathani, Loei, Phayao, Chiang Rai, Chumphon, and Songkhla). The effects of age for males and females were considered separately (see Results). Sex and age were grouped together into 14 levels (with seven levels of age in years: 5-19, 20-29, 30-39, 40-49, 50-59, 60-69, and 70+).

Similarly, misclassification of cause of death was considered differently for deaths in and outside hospitals. Reported causes of death and location were grouped into 18 levels, which resulted from the combination of two levels of location (in and outside hospital) and nine major causes of death (HIV, respiratory, septicemia, tuberculosis (TB), other infectious, mental and nervous system, digestive, ill-defined, and the remainder, which were aggregated into a single group).

Through logistic regression [20-22] we estimated the logit of the probability P that a person died from HIV as a linear function of the determinant factor. The simple logistic regression model with simple cross-referencing is formulated as

$$logit(P_i) = ln\left(\frac{P_i}{1\text{-}P_i}\right)=\mu+\alpha_i, \qquad \text{(A)}$$

where P_i is the probability of death due to HIV, μ is a constant, and α_i is the only parameter of DR cause-location i. The simple cross-referencing model (A) was compared with the full model (B), which includes an additive linear function of the determinant factors, which could be expressed as

$$logit\left(P_{ijk}\right) = ln\left(\frac{P_{ijk}}{1-P_{ijk}}\right) = \mu + \alpha_i + \beta_j + \gamma_k, \qquad \text{(B)}$$

where P_{ijk} is the probability of death due to HIV and α_i, β_j and γ_k are individual parameters specifying DR cause-location group i, sex-age group j and province k, respectively.

We used "sum contrasts" developed by Tongkumchum and McNeil [23] and Kongchouy and Sampantarak [24] instead of conventional "treatment contrasts" where the first level is left out from the model to be the reference. This method allows us to compute the estimate and the 95% confidence interval of deaths for each of the covariate levels in the VA and the DR datasets.

To assess the accuracy of model prediction, the Receiver Operating Characteristic (ROC) curve from logistic regression was drawn based on a concept described by Chongsuvivatwong [22] and Fan et al. [25]. Area under the ROC curve (AUC) measures the performance of a model and represents model accuracy [26,27]. A cut-off point in the curve, where the predicted number of HIV deaths equals the observed value in the VA dataset (512 cases), was used to report sensitivity and specificity of the model. These were compared with results from the simple cross-referencing method.

Estimation of HIV mortality

For the nine study provinces, fitting the complete logistic regression model to the 2005 VA dataset resulted in nine province coefficients, 14 sex-age group coefficients, and 18 DR cause-location coefficients and the estimate of HIV deaths and 95% confidence intervals.

For the remaining 67 provinces, we used a simple and easily implemented spatial "triangulation method" [28,29] to interpolate province coefficients. This was preferred to the "kriging" method because it uses fewer points than kriging, and there were insufficient sample provinces (only nine) to provide the basis for kriging [30].

Triangles were drawn linking nine provinces in the 2005 VA study. The values of province coefficients in each triangle were assigned as an average of coefficients from nearby provinces in the model. For each triangle, values *a, b* and *c* were obtained by solving three equations using linear algebra based on latitude and longitude as follows.

$$a + longP_1 \times b + latP_1 \times c = \beta_{P1} \tag{C}$$

$$a + longP_2 \times b + latP_2 \times c = \beta_{P2} \tag{D}$$

$$a + longP_3 \times b + latP_3 \times c = \beta_{P3} \tag{E}$$

(Note: P = Province, β = coefficient)

The coefficient for any province j *within* a triangle could then be given by

$$\beta_{Pj} = a + longP_j \times b + latP_j \times c \tag{F}$$

Coefficients for provinces outside triangles were obtained similarly by extrapolation from nearby provinces. Province coefficients for all provinces were thus obtained and the magnitude of HIV deaths estimated.

R program version 2.15.2 [31] was used for all statistical analysis and graphical displays.

Results

Ubon Ratchathani, Suphan Buri, and Chiang Rai had the largest numbers of total deaths (2373, 1600, and 1437 deaths, respectively), while Chumphon had the lowest (310 deaths). The VA-assessment gave 512 HIV deaths, whereas only 164 HIV deaths (32%) were correctly DR-reported.

From the likelihood ratio test in Table 2, the logistic regression model gives the deviance reduction between the full and null models as shown. All p-values are statistically significant.

Figure 1 shows crude percentages of HIV deaths (among all deaths) by province, sex-age group, and DR cause-location group and the adjusted values with 95% confidence intervals. The values derived from the direct VA assessment and from the full model are similar, indicating variation among groups but with no substantial confounding. The plotted values above the average line reflect the groups that were more likely to die from HIV.

The 95% confidence interval for both Phayao and Chumphon is marginally higher than the mean, whereas for Loei it is marginally lower. Therefore, effect of province on misclassification of cause of death was marginal. The percentages of HIV deaths in age groups 20-49 are all substantially above the mean, with females higher than males when those aged 20-39 years were compared. Thus, age groups 20-49 were significantly more likely to have high levels of under-reporting. Finally, substantial numbers of HIV deaths were reported as TB, mental and nervous system, other infectious diseases, and respiratory for deaths in hospitals, whereas HIV deaths outside hospitals were reported as TB, other infectious diseases, and septicemia. These are the groups in which HIV deaths were often misclassified.

The full model was assessed using the ROC curve and compared with a simple cross-referencing model. Figure 2 shows the ROC curve of the simple cross-referencing model (model A) with only DR cause-location factor and the ROC curve of the full model (model B) with three factors of DR cause-location, sex-age groups, and province. The cut-off point marked by the star gives a total predicted number agreement of the number of VA-assessed HIV deaths in the model. The simple cross-referencing model represents an AUC of 0.879, 53.1% sensitivity, and 97.5% specificity, whereas the full model represents an AUC of 0.969, 69.3% sensitivity,

Table 2 Results from logistic regression, full model

Factors	Deviance reduction	df	p-value
DR cause-location	954.67	17	< 0.0000001
Sex-age group	783.58	13	< 0.0000001
Province	19.43	8	0.01271
Error	711.90	1154	

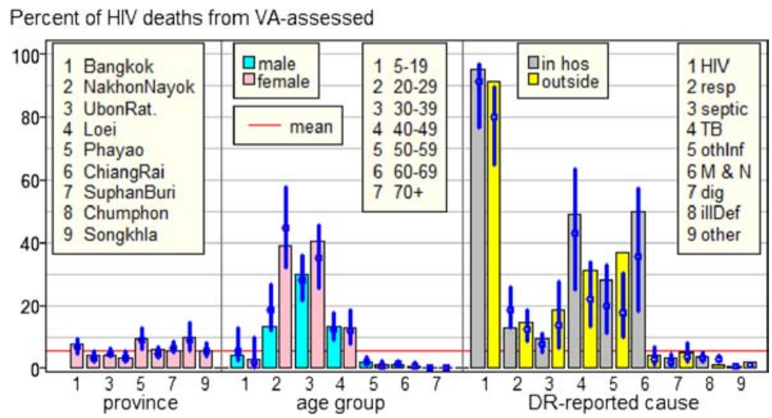

Figure 1 Estimated percentage of HIV-related deaths out of all deaths, based on the VA assessment (bars) and based on the full model (blue diamonds with error bars).

and 98.3% specificity. When we only compared an area above the diagonal line, the simple cross-referencing had AUC of 76% and the full model had AUC of 94%. In other words, the simple cross-referencing had an error of 24%, whereas the full model had an error of 6%. This means our model reduced the error by a factor of four. It is clear that the full model has the ability to predict the correct cause of HIV deaths better than the simple cross-referencing model. Just using the contingency table without statistical modeling, DR-reported cause had 32% sensitivity and 99.9% specificity.

The full model (B) was then extended to all provinces. The left panel of Figure 3 shows the nine study province coefficients from the logistic regression model plotted in black. Values plotted in blue are averages of coefficients from nearby provinces in each triangle using the triangulation method. The right panel classifies province from the equations (C, D, E, F in methods section), according to three levels of coefficients. The highest were found in the upper north (Phayao and Phrae) and the upper south (Prachuap Khirikhan, Chumphon, Ranong, Surat Thani, and Phang Nga). This implies that HIV deaths were proportionally highest among all deaths in the upper north and the upper south.

Finally, the simple cross-referencing and the full model were then applied to the DR data for male and female deaths in 1996-2009 and plotted as area graphs in Figure 4. The area of each color strip denotes the number of HIV deaths in each age group. The total number of DR-reported HIV deaths were much lower than those estimated by cross-referencing (model A) and logistic regression (model B) by factors of 2.8 and 3.1, respectively.

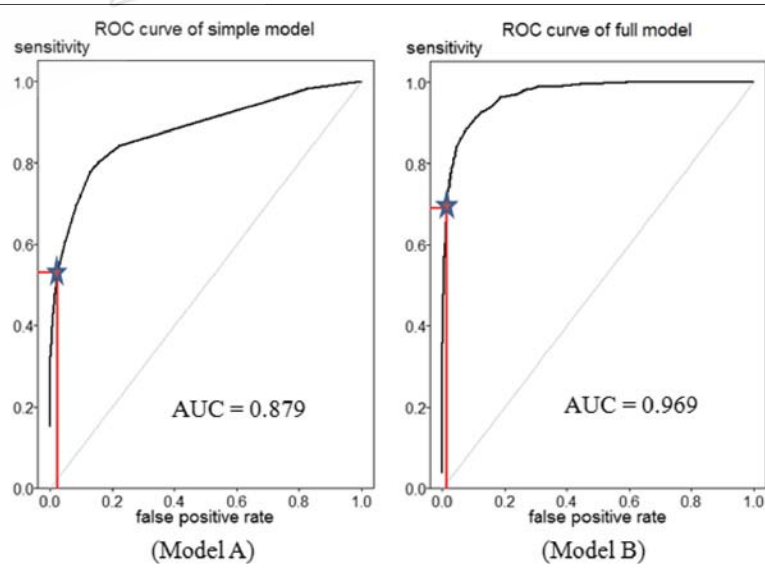

Figure 2 The ROC curve of logistic regression model with demographic covariates and without demographic covariates. Model **A** includes no demographic covariates whereas model **B** includes demographic covariates.

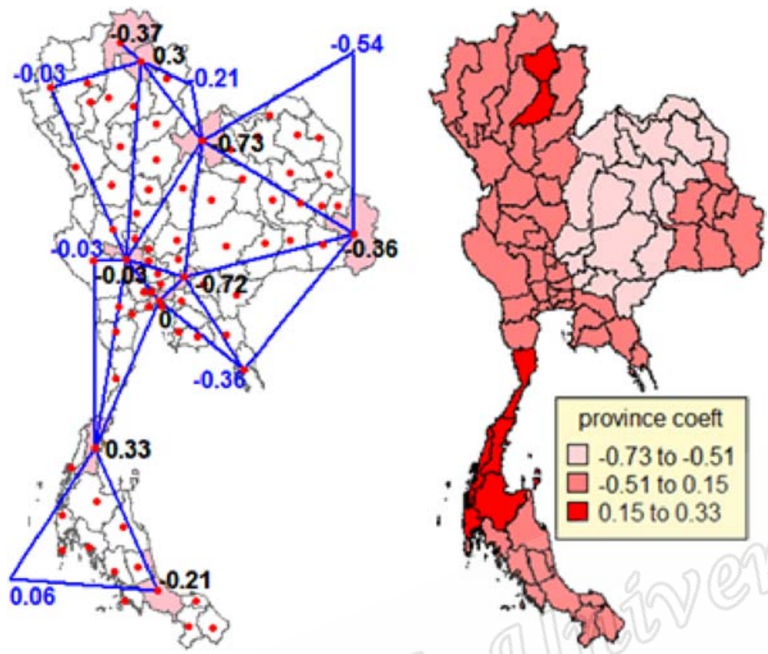

Figure 3 Triangulation for interpolating province coefficients in 2005.

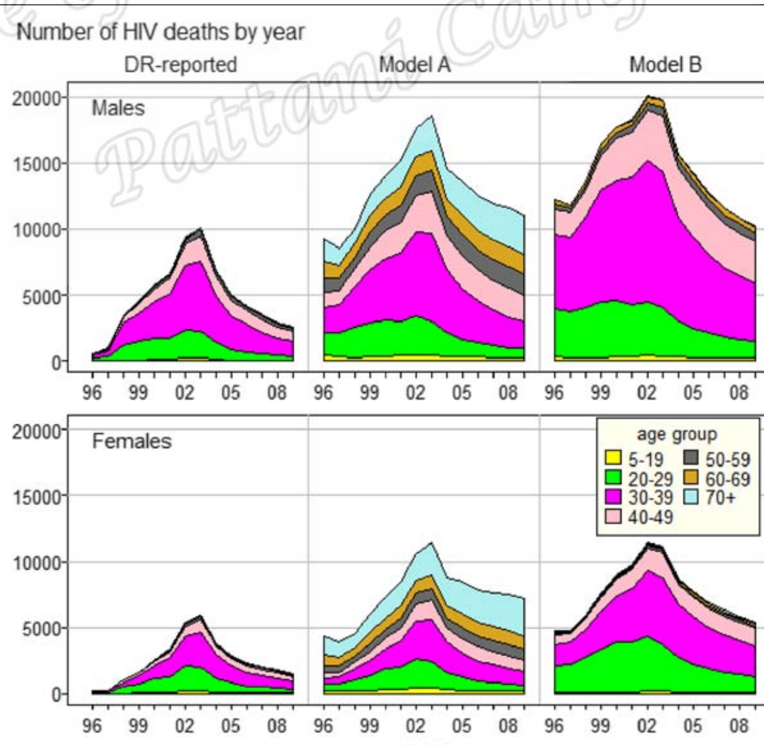

Figure 4 DR-reported and VA-estimated HIV deaths from the model with demographic covariates for both genders, 1996-2009. Model **A** includes no demographic covariates whereas model **B** includes demographic covariates.

While model A gave large proportions of HIV deaths at ages over 50 years (light blue, golden yellow, and grey), these were substantially reduced when the full logistic regression allowing for age/sex and province (model B) was used. On the other hand, for the young adult group, HIV cause of death was already substantially improved in accuracy by the simple logistic regression model.

Discussion and conclusions

Our logistic regression analysis showed that VA-assessed HIV deaths were more likely in female young adults compared to death registration, but many of those deaths were DR-registered as deaths from TB or from mental and nervous system disorders. A logistic regression-based method allowing for age/sex and geographical effects predicted HIV with higher sensitivity and specificity when compared with those HIV-estimated deaths derived from the cross-referencing from simple tabulation. Under-reporting was most common in the upper north and the upper south of the country. DR under-reported HIV deaths by a factor of three, whereas the simple cross-referencing method distorted the age distribution and could lead to a misunderstanding that HIV death was also common among the elderly.

HIV deaths were found to be relatively common among deaths in the age group 20-39 years, in agreement with other research [4,32]. AIDS is estimated to be the largest cause of death in Asian adults 15-44 years [33]. Before 1990, new HIV infections were highest among those injecting drugs and clients of sex workers. During 1995-2005, they were highest among the women with the category of housewife [34]. In other words, the most under-reporting of HIV deaths was found in females rather than males.

Most misclassifications of HIV deaths were classified as TB or mental and nervous system disorders. It is commonly known that TB and cryptococcal meningitis are the leading causes of opportunistic infections among HIV patients [35-38]. These infections were possibly recorded as the primary cause of deaths in death certificates either to avoid stigma to the family of the deceased, because the symptoms of TB and HIV are very similar, or because the people reporting the death might not have access to the results of a HIV test for the deceased. Another general condition often recorded was "immunodeficiency (D849: immunodeficiency, unspecified)." This might in fact be the more specific "HIV/AIDS (B20-B24: Human immunodeficiency virus disease)" in ICD10 coding [7].

Misclassification was associated with region (province). This could be due to difference in the levels of intensity of the HIV epidemic, stigmatization, and availability of qualified personnel for DR recording and their attitude toward HIV-related death across the regions. HIV mortality peaked in the upper north, especially in Phayao, because in the past two decades the HIV epidemic has been most severe in the upper north [32,34,39-42]. One-third of HIV deaths were predicted in the northern region since 1987-2014 [39]. Those HIV deaths were higher in the upper south than in the central region in spite of the less severe HIV epidemic [40]. HIV deaths in the south were more likely to be misclassified to other causes, as the area was perceived to have low levels of HIV [43]. In addition, mortality varies by geographic location, and the south has the lowest overall mortality [43,44].

Our full logistic regression model based on the 2005 VA data was shown to predict and estimate HIV deaths with high sensitivity, specificity, and AUC, better than the simple cross-referencing model. The specificity level from our model was higher than a verbal autopsy tool from Uganda, where sensitivity was not reported [45]. The cross-referencing method has been used in many previous studies [1,15-18,45]. Inadequate models can give misleading or incorrect inferences [26]. Our study showed that the use of this simple method should be discouraged because it distorts the HIV death estimate in various demographic groups. This distortion can mislead priority setting and resource allocation.

There were limitations in our analysis. First, the sample survey design did not stratify by strong predictors of the outcome such as reported cause and location of report. The study sample thus did not adequately cover the population at risk for HIV and the sample size did not allow precise estimation among certain minority groups, such as the Muslim group in the far south. Second, only nine of Thailand's 76 provinces were included in the VA study.

Third, we have assumed that the 2005 VA data can inform corrections in all years between 1996 and 2009, while it is clear that the coverage of antiretroviral treatment was near zero in 1996, 12% in 2003, 41% in 2005, and 76% in 2009 [46]. There would therefore be differences in misclassification of HIV-related deaths across the years, which are not captured by our methods. Finally, VA itself has limitations, in terms of inaccuracy of informants and recall bias. The results must therefore be carefully interpreted.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
All authors participated in the design of the study and the interpretation of the results. AC and PT performed the study analysis and drafted the manuscript and prepared all tables and graphs. VC critically reviewed the manuscript. All authors read and contributed to the final manuscript.

Acknowledgments
The authors gratefully acknowledge Prof. Dr. Don McNeil, Professor Emeritus of Statistics at Macquarie University, Australia for his valuable and helpful guidance and Greig Rundle for his assistance and suggestions. We also thank the SPICE project team for collecting the 2005 VA data.

Author details

[1]The Office of Disease Prevention and Control, Region 11, Nakhon Si Thammarat, Muang, Nakhon Si Thammarat, Thailand. [2]Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Muang, Pattani, Thailand. [3]International Health Policy Program (IHPP), Bureau of Policy and Strategy, Ministry of Public Health, Nonthaburi, Thailand. [4]Epidemiology Unit, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla, Thailand.

Received: 2 December 2013 Accepted: 24 September 2014
Published online: 03 October 2014

References

1. Taffa N, Will JC, Bodika S, Packel L, Motlapele D, Stein E, Roels TH, Kennedy G, Shenaaz EH: **Validation of AIDS-related mortality in Botswana.** *JIAS* 2009, **12**:24.
2. Fazito E, Cuchi P, Fat DM, Ghys PD, Pereira MG, Vasconcelos AMN, Pascom ARP: **Identifying and quantifying misclassified and under-reported AIDS deaths in Brazil: a retrospective analysis from 1985 to 2009.** *Sex Transm Infect* 2012, **88**(Suppl 2):i86–i94.
3. Pacheco AG, Saraceni V, Tuboi SH, Lauria LM, Moulton LH, Faulhaber JC, King B, Golub JE, Durovni B, Cavalcante S, Harrison LH, Chaisson RE, Schechter M: **Estimating the extent of underreporting of mortality among HIV-infected individuals in Rio de Janeiro, Brazil.** *AIDS Res Hum Retroviruses* 2011, **27**(1):25–28.
4. Yudkin PL, Burger EH, Bradshaw D, Groenewald P, Ward AM, Volmink J: **Deaths caused by HIV disease under-reported in South Africa.** *AIDS* 2009, **23**(12):1600–1602.
5. Birnbaum JK, Murray CJL, Lozano R: **Exposing misclassification HIV/AIDS deaths in South Africa.** *Bull World Health Organ* 2011, **89**:278–285.
6. Tangcharoensathien V, Faramnuayphol P, Teokul W, Bundhamcharoen K, Wibulpholprasert S: **A critical assessment of mortality statistics in Thailand: potential for improvements.** *Bull World Health Organ* 2006, **84**(3):233–238.
7. Khonhan K: **Quality of mortality data of HIV/AIDS surveillance reporting system in Mukdahan Province, Thailand (in Thai).** *Thai Popul J* 2009, **1**(1):125–135.
8. Rao C, Porapakkham Y, Pattaraarchachai J, Polprasert W, Swampunyalert N, Lopez AD: **Verifying causes of death in Thailand: rationale and methods for empirical investigation.** *Popul Health Metr* 2010, **8**:11.
9. Mathers CD, Fat DM, Inoue M, Rao C, Lopez AD: **Counting the dead and what they died from: an assessment of the global status of cause of death data.** *Bull World Health Organ* 2005, **83**(3):171–177.
10. Prasartkul P, Vapattanawong P: **The completeness of death registration in Thailand: evidence from demographic surveillance system of the Kanchanaburi project.** *World Health Popul* 2006, **8**:43–51.
11. Hill K, Vapattanawong P, Prasartkul P, Porapakkham Y, Lim SS, Lopez AD: **Epidemiologic transition interrupted: a reassessment of mortality trends in Thailand, 1980–2000.** *Int J Epidemiol* 2007, **36**:374–384.
12. Prasartkul P, Porapakham Y, Vapattanawong P, Rittirong J: **Development of a verbal autopsy tool for investigating cause of death: the Kanchanaburi project.** *JPSS* 2007, **15**(2):1–22.
13. Vapattanawong P, 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.
14. Lopez AD, Lozano R, Murray CJL, Shibuya K: **Verbal autopsy: innovations, applications, opportunities improving cause of death measurement.** *In Popul Health Metr* 2011, **9**:128–254.
15. Choprapawon C, Porapakkham Y, Sablon O, Panjajaru R, Jhantharatat B: **Thailand's national death registration reform: verifying the causes of death between July 1997 and December 1999.** *Asia Pac J Public Health* 2005, **17**(2):110–116.
16. Pattaraarchachai J, Rao C, Polprasert W, Porapakkham Y, Poa-in W, Singwerathum N, Lopez AD: **Cause-specific mortality patterns among hospital deaths in Thailand: validating routine death certification.** *Popul Health Metr* 2010, **8**:12.
17. Polprasert W, Rao C, Adair T, Pattaraachachai J, Porapakkham Y, Lopez AD: **Cause-of-death ascertainment for deaths that occur outside hospitals in Thailand: application of verbal autopsy methods.** *Popul Health Metr* 2010, **8**:13.
18. Porapakkham Y, Rao C, Pattaraachachai J, Polprasert W, Vos T, Adair T, Lopez AD: **Estimated causes of death in Thailand, 2005: implications for health policy.** *Popul Health Metr* 2010, **8**:14.
19. World Health Organization: *ICD-10 International Statistical Classification of Diseases and Related Health Problems.* Geneva: WHO; 2004.
20. McNeil D: *Epidemiological Research Methods.* New York: John Wiley & Sons Ltd; 1996.
21. Venables WN, Ripley BD: *Modern Applied Statistics with S.* New York: Springer; 2002.
22. Chongsuvivatwong V: *Graphs, Tables and Equations for Health Research (in Thai).* Bangkok: Chulalongkorn University Press; 2007.
23. Tongkumchum P, McNeil D: **Confidence intervals using contrasts for regression model.** *Songklanakarin J Sci Technol* 2009, **31**(2):151–156.
24. Kongchouy N, Sampantarak U: **Confidence intervals for adjusted proportions using logistic regression.** *Mod Appl Sci* 2010, **4**(6):2–7.
25. Fan J, Upadhye S, Worster A: **Understanding receiver operating characteristic (ROC) curves.** *Can J Emerg Med* 2006, **8**(1):19–20.
26. Sakar S, Midi H: **Importance of assessing the model adequacy of binary logistic regression.** *J of Appl Sci* 2010, **10**(6):479–486.
27. Takahashi K, Uchiyama H, Yanagisawa S, Kamae I: **The logistic regression and ROC analysis of group-based screening for predicting diabetes incidence in four years.** *Kobe J Med Sci* 2006, **52**(6):171–180.
28. Li J, Heap AD: *A Review of Spatial Interpolation Methods for Environmental Scientists.* Canberra: Geoscience Australia, Record 2008/23; 2008.
29. Yang CS, Kao SP, Lee FB, Hung PS: **Twelve Different Interpolation Methods: A Case Study of SURFER 8.0.** [http://www.isprs.org/proceedings/XXXV/congress/comm2/comm2.aspx]
30. Murphy M: **The Advantages of Kriging vs Triangulation Contour Mapping Methods.** [http://www.ehow.com/info_12002607_advantages-kriging-vs-triangulation-contour-mapping-methods.html]
31. R Development Core Team: **R program: A Language and Environment for Statistical Computing and Graphics** [http://cran.r-project.org/bin/windows/base/old/2.15.2/]
32. Punyacharoensin N, Viwatwongkasem C: **Trends in three decades of HIV/AIDS epidemic in Thailand by nonparametric backcalculation method.** *AIDS* 2009, **23**(9):1143–1152.
33. Kerr S, Phanuphak P: **An Asian perspective on HIV/AIDS.** *Asian Biomed* 2009, **3**(1):9–14.
34. The Thai Working Group on HIV/AIDS Projections (2005): *The Asian Epidemic Model (AEM) Projections for HIV/AIDS in Thailand: 2005-2025.* Bangkok: Family Health International (FHI) and Bureau of AIDS,TB and STIs, Department of Disease Control, Ministry of Public Health; 2008.
35. Kantipong P, Murakami K, Moolphate S, Aung MN, Yamada N: **Causes of mortality among tuberculosis and HIV co-infected patients in Chiang Rai, Northern Thailand.** *HIV AIDS Res Palliat Care* 2012, **4**:159–168.
36. Cain KP, Anekthananon T, Burapat C, Akksilp S, Mankhatitham W, Srinak C, Nateniyom S, Sattayawuthipong W, Tasaneeyapan T, Varma JK: **Caused of death in HIV-infected persons who have Tuberculosis.** *Thailand Emerg Infect Dis* 2009, **15**(2):258–264.
37. Kitkungvan D, Apisarnthanasak A, Plengpart P, Mundy LM: **Fever of unknown origin in patients with HIV infection in Thailand: an observational study and review of the literature.** *Int J STD AIDS* 2008, **19**:232–235.
38. Likittanasombut P: **Opportunistic central nervous system infection in human immunodeficiency virus infected patients in Thammasat hospital, Thailand.** *Neurology Asia* 2004, **9**:29–32.
39. Surasiengsunk S, Kiranandana S, Wongboonsin K, Garnett GP, Anderson RM, Griensven GJP: **Demographic impact of the HIV epidemic in Thailand.** *AIDS* 1998, **12**(7):775–784.
40. Bureau of Epidemiology, Ministry of Public Health: *AIDS Situation.* 2010 [http://www.boe.moph.go.th/report.php?cat=19&year=2010]
41. Chariyalertsak S, Sirisanthana T, Saengwonloey O, Nelson KE: **Clinical presentation and risk behaviors of patients with acquired immunodeficiency syndrome in Thailand, 1994-1998: regional variation and temporal trends.** *Clin Infect Dis* 2001, **32**:955–962.
42. Jones H, Pardthaisong L: **Demographic interactions and developmental implications in the era of AIDS: findings from Northern Thailand.** *Appl Geography* 2000, **20**(3):255–275.
43. Faramnuayphol P, Chongsuvivatwong V, Pannarunothai S: **Geographical variation of mortality in Thailand.** *J Med Assoc Thai* 2008, **91**(9):1455–1460.
44. Odton P, Choonpradub C, Bundhamcharoen K: **Geographical variations in all-cause mortality in Thailand.** *Southeast Asian J Trop Med Public Health* 2010, **41**(5):1209–1219.

45. Mayanja BN, Baisley K, Nalweyiso N, Kibengo FM, Mugisha JO, Van der Paal L, Maher D, Kaleebu P: Using verbal autopsy to assess the prevalence of HIV infection among deaths in ART period in rural Uganda: a prospective cohort study, 2006-2008. *Popul Health Metr* 2011, **9**:36.
46. National AIDS committee: *2012 Thailand AIDS Response Progress Report.* Bangkok: National AIDS Management Center, Department of Disease Control, Ministry of Public Health; 2013.

doi:10.1186/s12963-014-0025-x
Cite this article as: Chutinantakul *et al.*: Correcting and estimating HIV mortality in Thailand based on 2005 verbal autopsy data focusing on demographic factors, 1996-2009. *Population Health Metrics* 2014 **12**:25.

Appendix IV

Table of HIV VA-estimated and DR-reported deaths from 1996-2009 by sex and age groups.

		year													
	sex-age	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
VA-estimated	m 5-19	329	287.4	235.7	294.2	324.1	368.4	428.3	390.6	337	**275**	263.5	234.8	246.2	215
	m 20-29	3602.9	3502.6	3836.7	4201	4232.3	3876.5	4065.2	3628.9	2631.9	**2162**	1842.7	1551.2	1372.3	1275.6
	m 30-39	5652.9	5611.6	6839.6	8428.9	9169.6	9705.9	10648	10370	7927.4	**6939**	6026.3	5258.2	4837.1	4389
	m 40-49	1945	1826	2059	2696	3145	3344	3825	4179	3685	**3685**	3481	3390	3267	3150
	m 50-59	356	306	319	428	476	511	600	695	634	**656**	638	672	665	655
	m 60-69	354	287	285	373	393	403	431	439	399	**445**	443	436	435	439
	m 70+	22	17	17	22	23	25	28	32	30	**35**	36	39	42	43
	f 5-19	64	64	77	100	123	148	213	187	144	**103**	93	88	84	88
	f 20-29	2046	2161	2668	3256	3789	3720	4135	3548	2636	**2148**	1765	1515	1361	1186
	f 30-39	1617	1680	2159	2889	3537	4096	4972	5069	3972	**3516**	3062	2828	2576	2353
	f 40-49	719	678	779	1037	1228	1420	1729	1896	1527	**1521**	1527	1425	1407	1310
	f 50-59	66	60	63	95	103	119	155	175	153	**156**	147	157	146	145
	f 60-69	136	117	117	152	149	159	170	180	166	**181**	169	173	176	173
	f 70+	65	51	54	66	69	77	85	96	89	**103**	105	114	120	120
	total	16974	16647	19508	24036	26762	27972	31485	30887	24332	**21924**	19600	17881	16734	15542
DR-reported	m 5-19	8	14	35	70	111	133	226	224	151	**83**	70	53	60	39
	m 20-29	194	379	1211	1410	1599	1645	2120	1989	1238	**802**	647	480	360	332
	m 30-39	208	404	1607	2200	2795	3391	4910	5348	3524	**2560**	2037	1595	1300	1104
	m 40-49	65	107	425	651	954	1090	1574	1949	1471	**1192**	1076	974	854	809
	m 50-59	18	36	90	142	186	218	350	455	352	**289**	265	299	251	227
	m 60-69	16	17	20	48	69	60	90	103	72	**64**	76	61	49	61
	m 70+	18	23	6	19	17	29	43	39	18	**20**	19	18	26	24
	f 5-19	10	10	43	69	105	139	246	211	145	**89**	77	66	56	66
	f 20-29	68	121	463	663	1037	1164	1824	1785	1024	**701**	490	416	318	223
	f 30-39	52	54	385	590	921	1402	2274	2637	1756	**1345**	1035	909	808	655
	f 40-49	15	14	129	208	341	470	784	963	631	**515**	542	447	407	339
	f 50-59	7	15	24	61	74	100	178	215	165	**142**	111	135	120	108
	f 60-69	10	13	9	27	20	33	37	54	32	**36**	24	26	33	33
	f 70+	16	16	4	7	13	28	30	34	10	**10**	13	11	15	12
	total	705	1223	4451	6165	8242	9902	14686	16006	10589	**7848**	6482	5490	4657	4032
IF=VA/DR		24.08	13.61	4.38	3.90	3.25	2.82	2.14	1.93	2.30	**2.79**	3.02	3.26	3.59	3.85

28-29 September, 2013 Scholarship for proceedings from the Faculty of Science and Technology, Prince of Songkla University, Songkhla, Thailand

Work-Position and Address

Position: Senior Technical officer

Address: The Office of Disease Prevention and Control, region 11 Nakhon Si Thammarat, Muang, Nakhon Si Thammarat, 80000 Thailand

List of Publications and Proceedings

Publications:

Chutinantakul, A., Mayeng, M. and Tongkumchum, P. 2014. Estimation of mortality with missing data using logistic regression. Songklanakarin Journal of Science and Technology. 36(2), 249-254.

Chutinantakul, A., Tongkumchum, P., Bundhamcharoen, K. 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 Matrics. 12(25).

Proceedings:

Chutinantakul, A., Tongkumchum, P. and Chongsuvivatwong, V. Estimating HIV deaths in Thailand in 2005 using verbal autopsy modeling. The 5th International Conference on Public Health among Greater Mekong Sub-Regional Countries. 28-29 September, 2013. Yangon, Myanmar.