Chapter 3

Results

This chapter presents results from preliminary analysis and statistical modeling of the three parts. The first part presents the results of estimating mortality data with unknown province or age, the second part presents the results of correcting under-reporting/misclassification of HIV deaths. Finally, the third part presents the results after impute missing demographic factors and reclassification of cause of HIV deaths.

3.1 Part I: Estimation of mortality with missing data using logistic regression

(Chutinantakul *et al*., 2014a)

3.1.1 Preliminary results

Annual numbers of deaths from 1996 to 2009 varied from 300,321 cases in 1997 to 397,327 cases in 2008. Percentages of deaths with unknown province and age are shown in Table 3.1. The highest unknown province is in 2004 whereas unknown age is in 1996. Overall percentage of unknown province deaths increased from 1.1% to 6.8% from 1997 to 2004. The percentages of deaths with unknown province or age are negligible after 2006, but need correction in earlier years.

Table 3.1 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

Separate logistic regression models were fitted to unknown province or age in DR data in each year from 1996-2009. Figure 3.1 shows the deviance residuals that are plotted against corresponding theoretical quantiles for model from 1996-2005. For the year 2006-2009, it has small sample size of missing cases and the models gave widely 95% confidence interval. Thus, these years were not presented. These plots display the models provided an acceptable fit to both proportions of deaths with unknown province and age.

Several models fitted quite well for unknown province and age by each year, except unknown province in 1998 and 2002 and unknown age in 1996, 1997 and 2001.

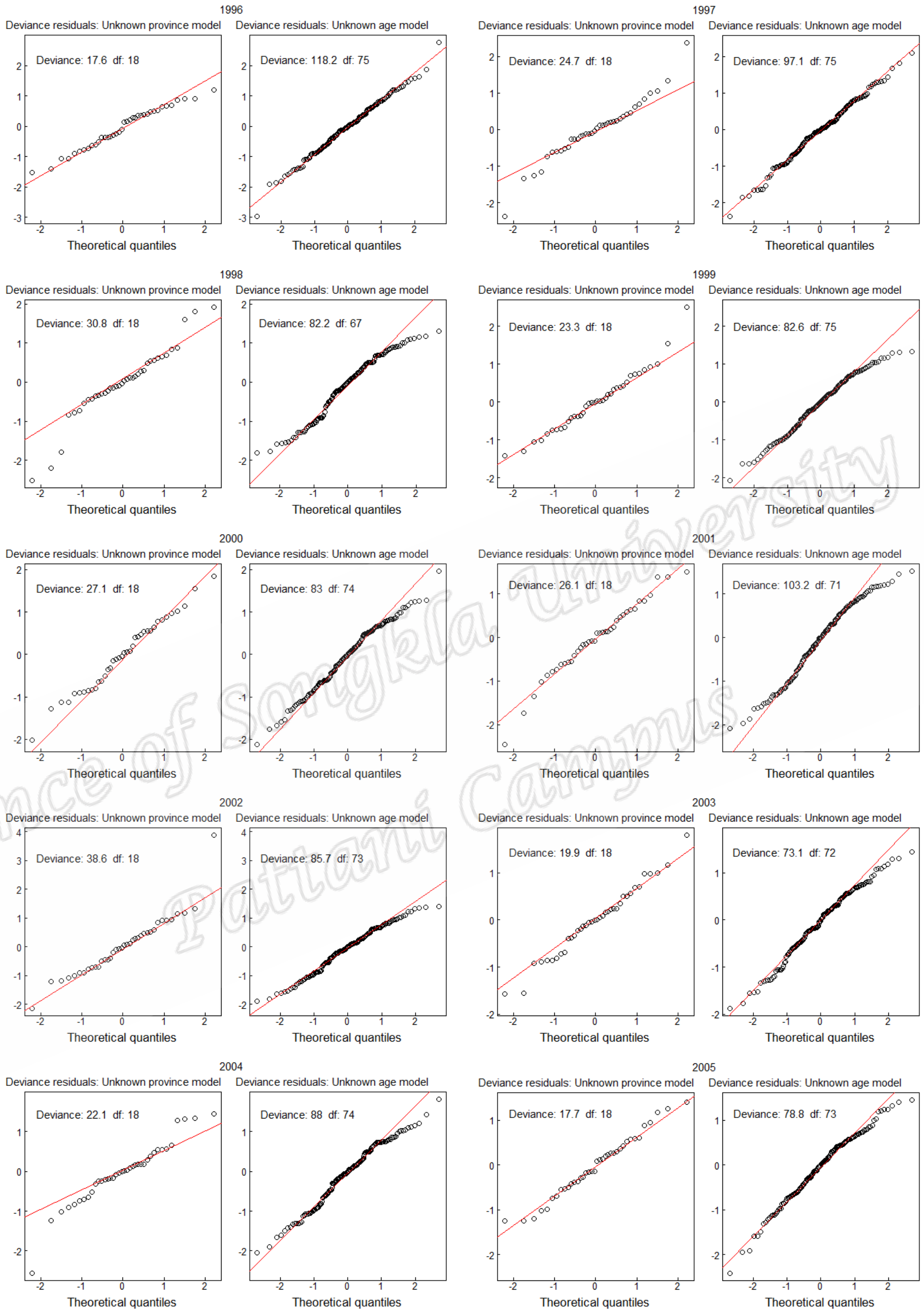

Figure 3.1 Plots of deviance residuals versus theoretical quantiles based on logistic regression models for predicting the percentage of deaths in Thailand from 1996-2005 with unknown province and age

We graphed percentages of unknown province from model (a) as described in method section 2.2.2 and shown in Figure 3.2. The results are classified into three groups. The first pattern shows unknown provinces all age groups around the mean as shown in 1996. The second pattern shows highly percentages of unknown provinces in only working age groups as shown in 1998, 2000, 2002 and 2005. The third pattern presents highly percentages of unknown provinces in working age groups including 0-4 years as presented in 1997, 1999, 2001, 2003 and 2004.

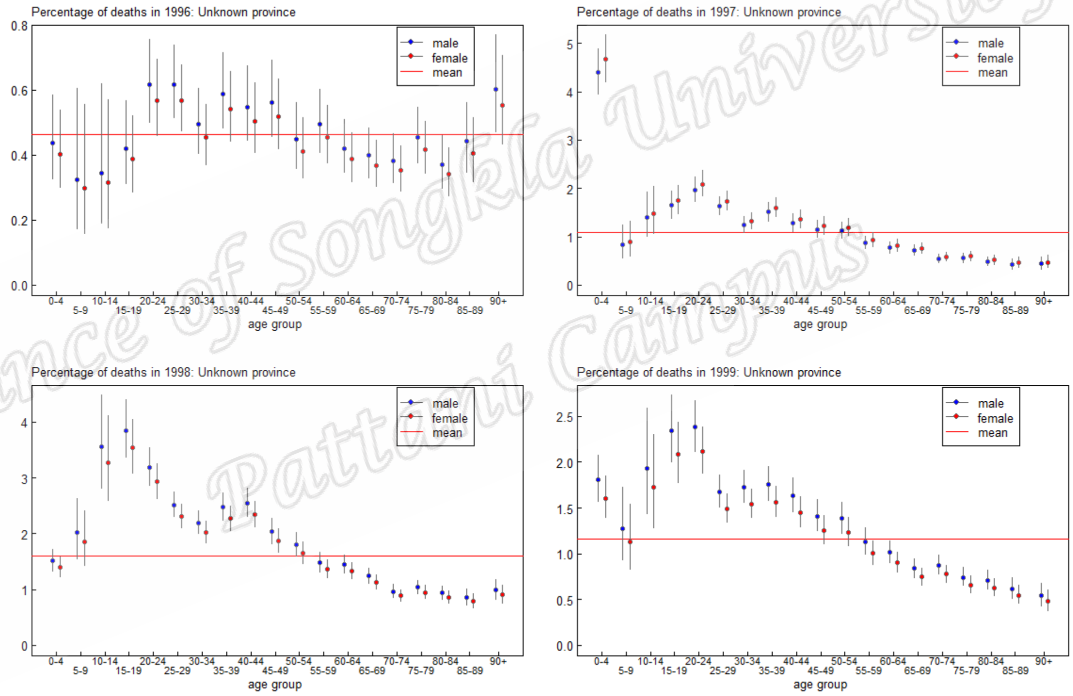

Figure 3.2 Estimates with 95% confidence intervals for percentages of deaths with unknown province in Thailand from 1996-2005

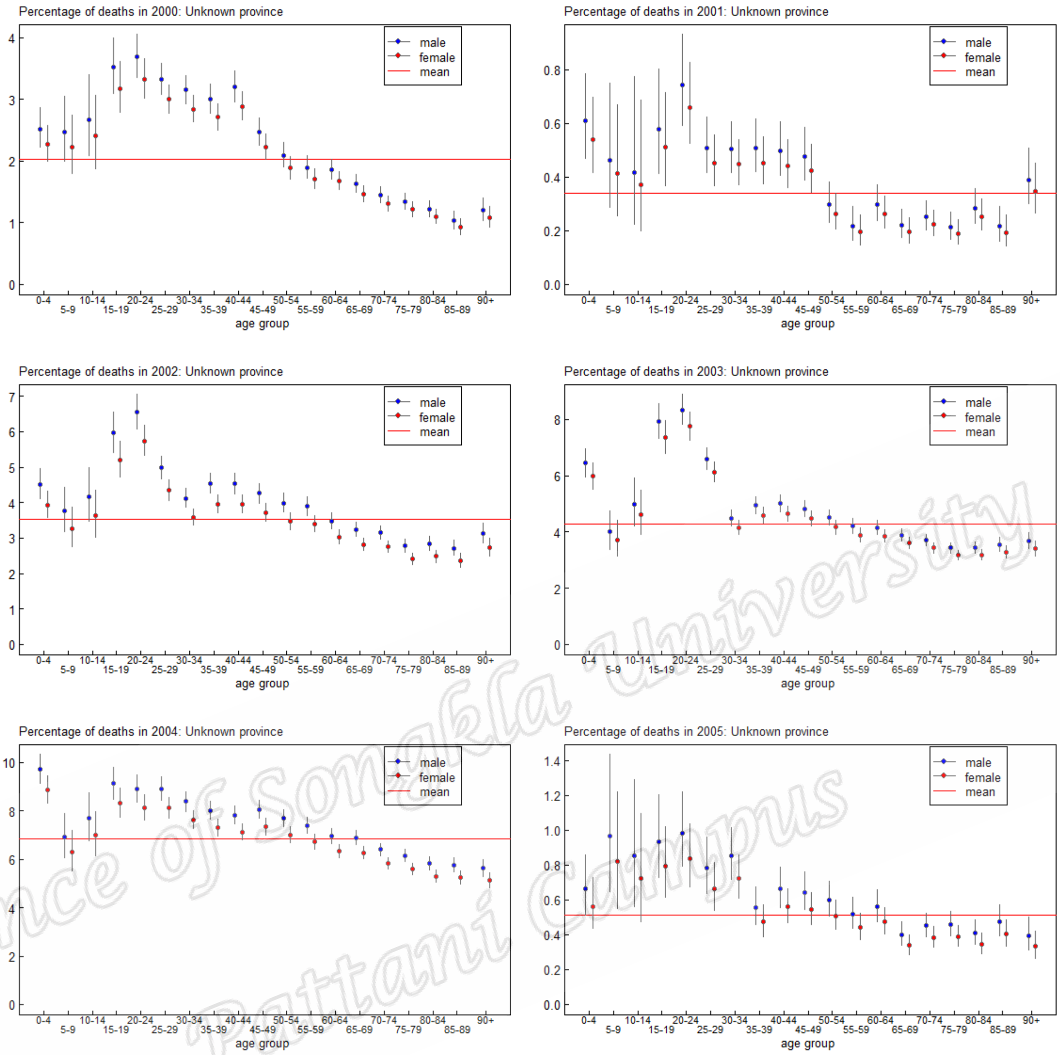

Figure 3.2 Estimates with 95% confidence intervals for percentages of deaths with unknown province in Thailand from 1996-2005 (cont.)

For unknown age in 1996-2005, Figure 3.3 graphs the results from the logistic regression model (b) as described in method section 2.2.2. The percentages of unknown age above average mostly occurred in the Central region in the first two years, then diffused all regions and moved to the South in 2005. Trends of percentages of unknown age decreased from 1996 to 2005. Twenty provinces of unknown age are above average in 1996 and decreased to eight provinces in 2005.

This implied improving quality data of unknown age. For 2005, half of them occurred in the Southern region.

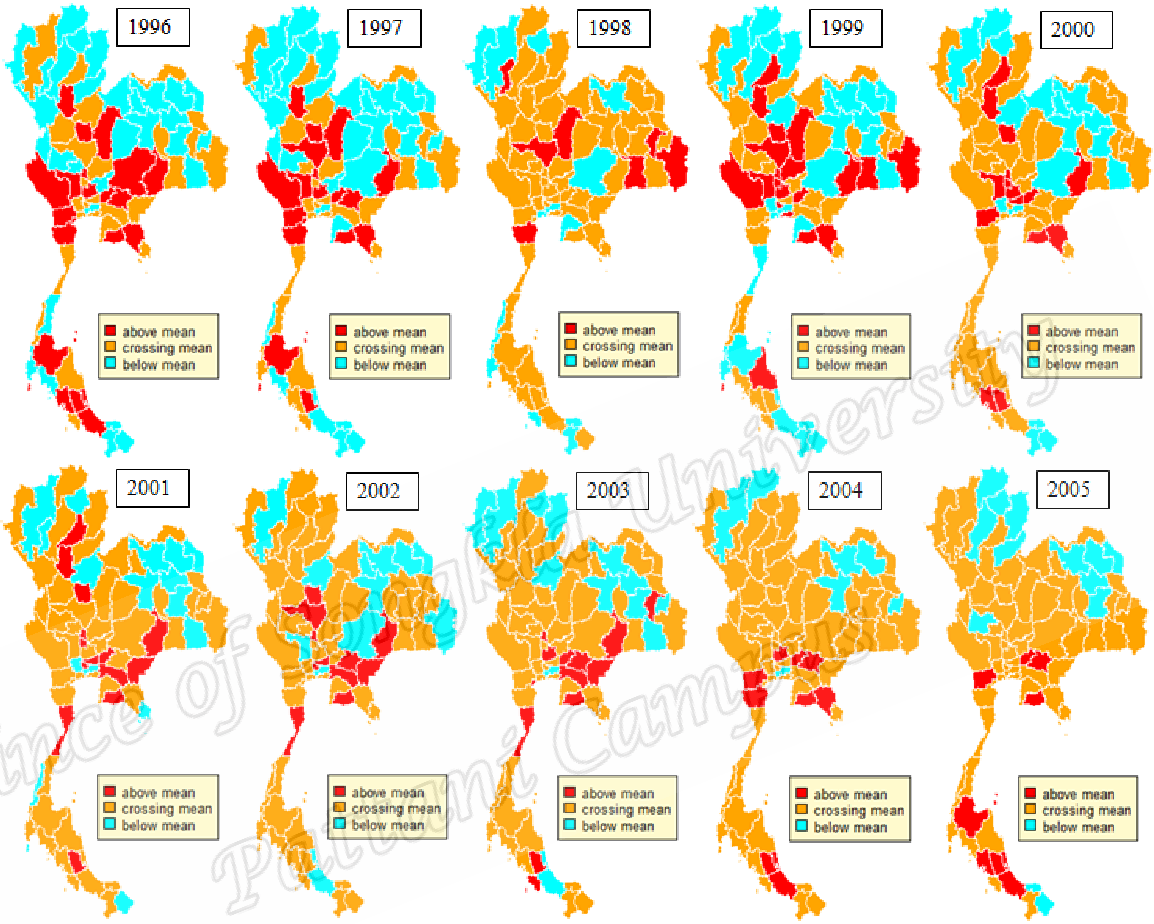

Figure 3.3 The thematic map of unknown age from 1996-2005

3.1.2 Results

From the preliminary results, unknown province or age occurred in the earlier years and varied with time. We selected 1997 and 2004 to represent unknown province and age. These years illustrated high percentage of unknown province or age.

Figure 3.4 shows 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 by sex and age group with unknown provinces from the overall mean (red line).

The pattern of unknown provinces in 1997 and 2004 are similar. Percentage of unknown provinces above average mostly occurred in the working 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.

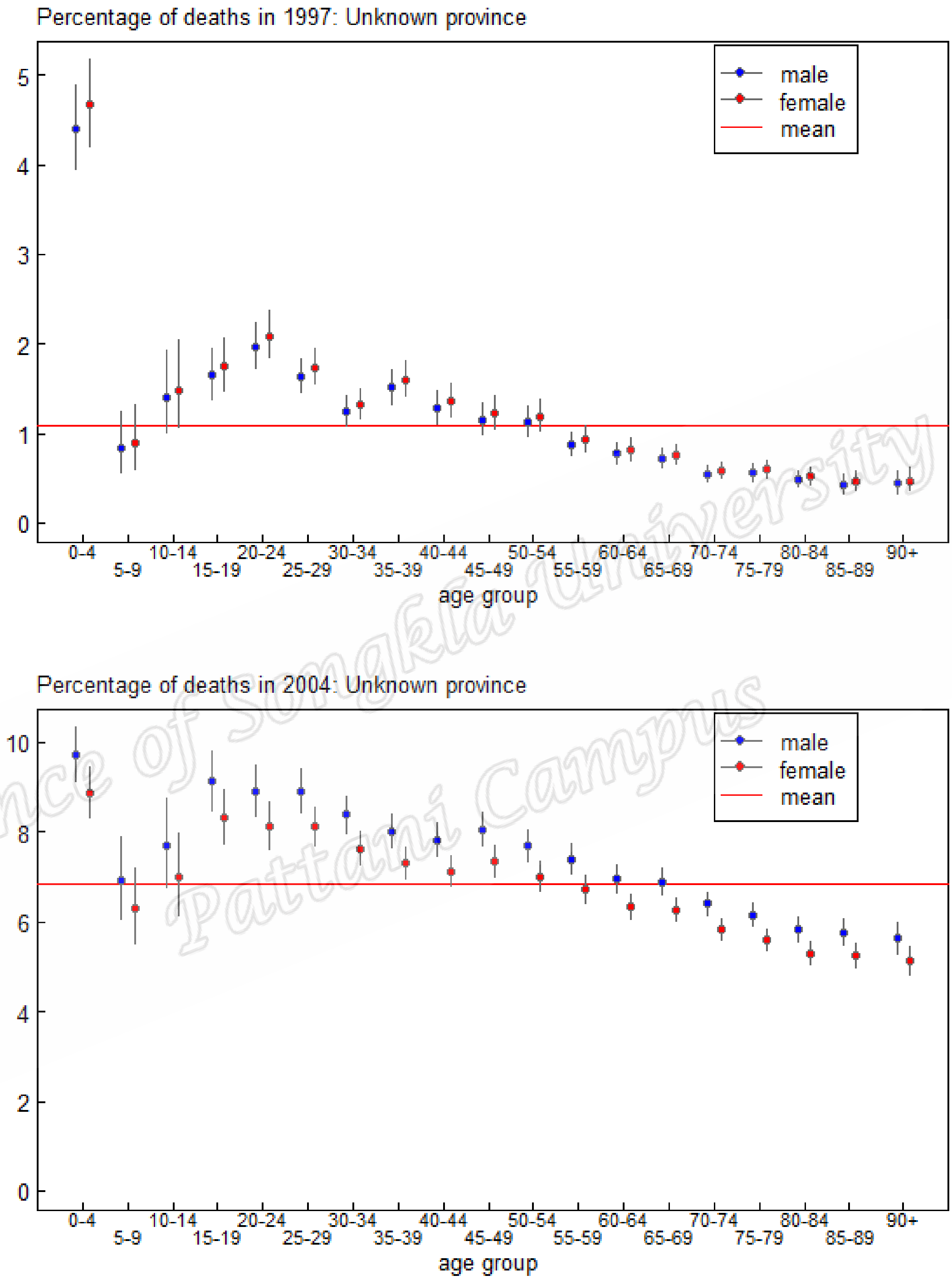

Figure 3.4 Estimates with 95% confidence intervals for percentages of deaths with unknown province in Thailand in 1997 (upper panel) and 2004 (lower panel)

Figure 3.5 graphs similar results from the logistic regression model. The model estimates the percentages of deaths with unknown age by sex 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.

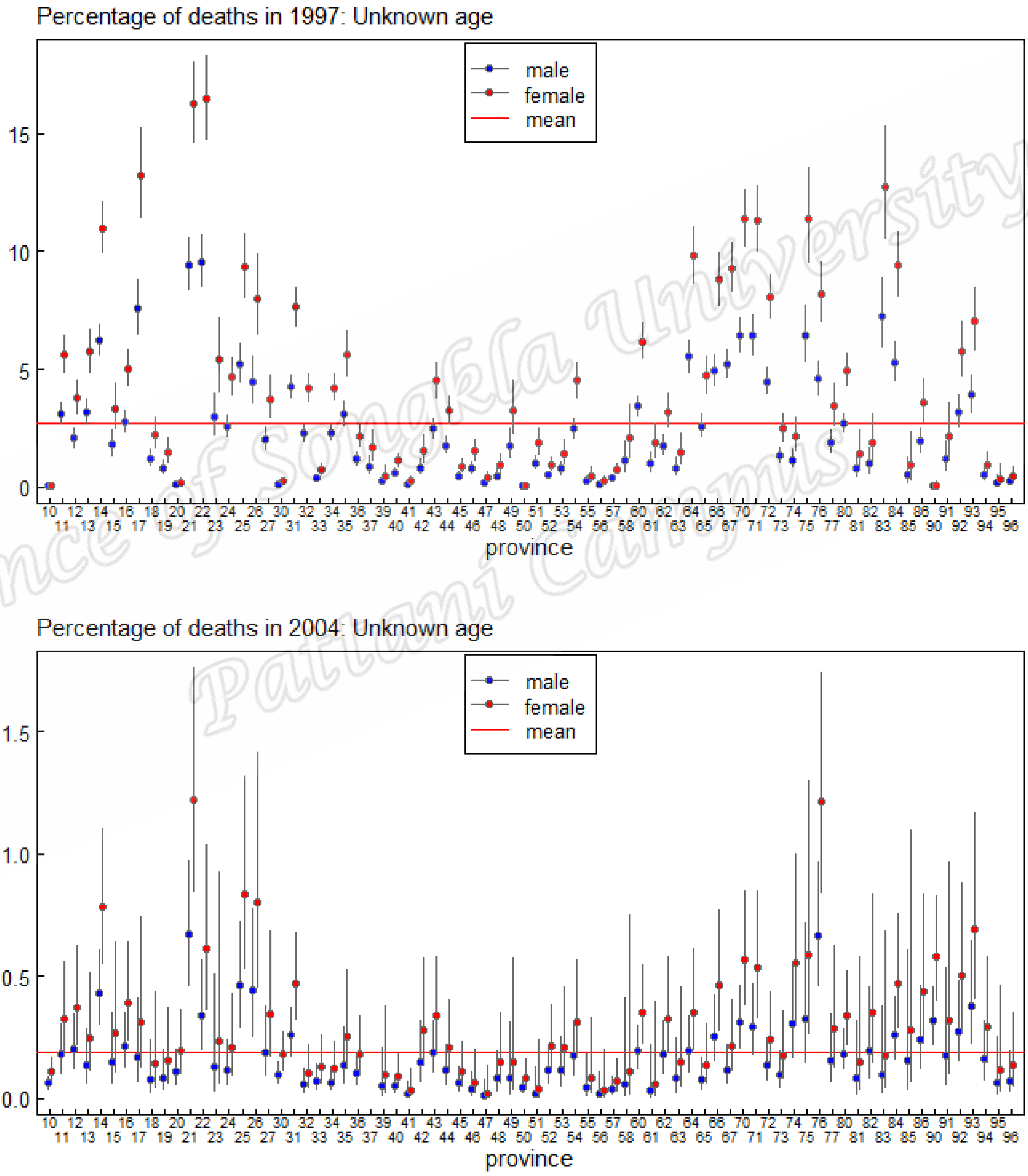

Figure 3.5 Estimates with 95% confidence intervals for percentages of deaths with unknown age in Thailand in 1997 (upper panel) and 2004 (lower panel)

In 1997, over 15% of female unknown age deaths occurred in Rayong (code 21) and Chanthaburi (code 22) provinces (upper panel of Figure 3.5). 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% 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 sexes in each province from 1997 to 2004 as shown in Figure 3.6. They separate three levels of these percentages into groups where the confidence intervals are entirely above the mean (colored 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).

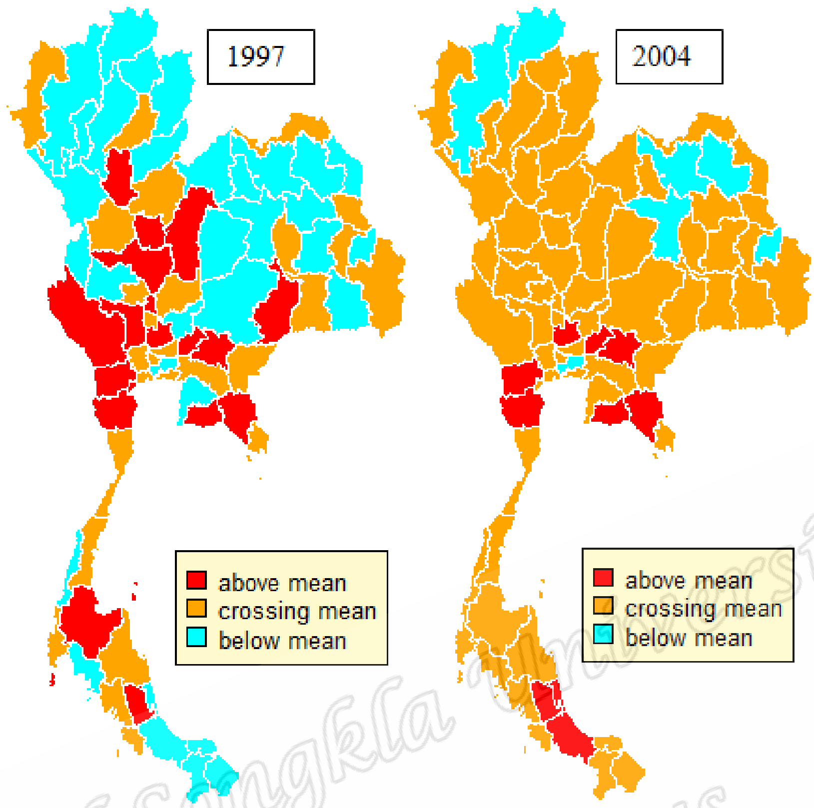

Figure 3.6 The thematic map of unknown age in 1997 and 2004

3.2 Part II: Correcting and estimating HIV mortality in Thailand based on 2005 verbal autopsy data focusing on demographic factors, 1996-2009 (Chutinantakul *et al.*, 2014b)

3.2.1 Preliminary results

Part II examines HIV deaths in DR in Thailand during 1996-2009 based on the 2005 VA data. The simple cross-referencing between VA cause groups and DR cause groups of the 2005 VA data are shown in Figure 3.7. It presented over-reporting substantially found in ill-defined and septicemia whereas under-reporting found in stroke, HIV and liver cancer. Numbers of HIV deaths in DR cause groups were 176 cases whereas 512 cases by VA-assessed. Only 164 HIV deaths (32%) were correctly DR-reported.

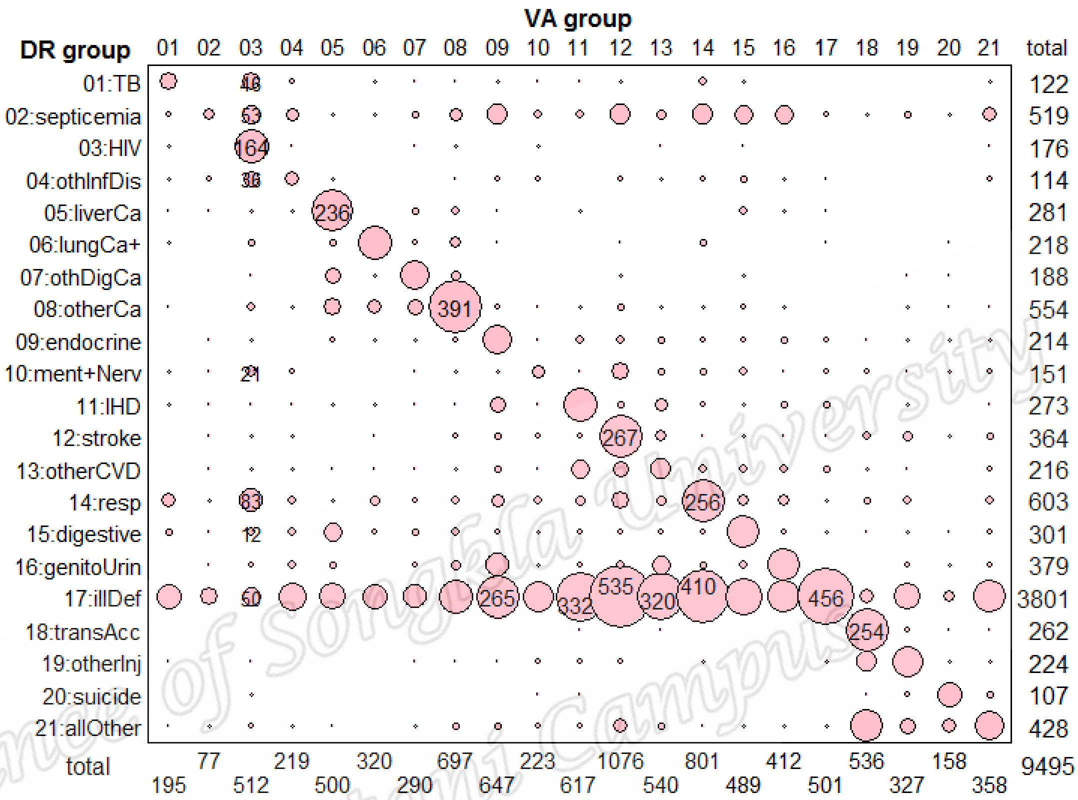

Figure 3.7 The cross-referencing between DR cause groups and VA cause groups of the 2005 VA data

All of 176 cases of HIV DR-reported deaths were shown in Figure 3.8. Phayao, Chiang Rai and Chumphon show higher percentages of HIV deaths than the average (1.85%) including males and females in age groups 20-49 years. HIV deaths were misclassified reporting to TB, other infectious diseases, mental and nervous system, and digestive diseases.

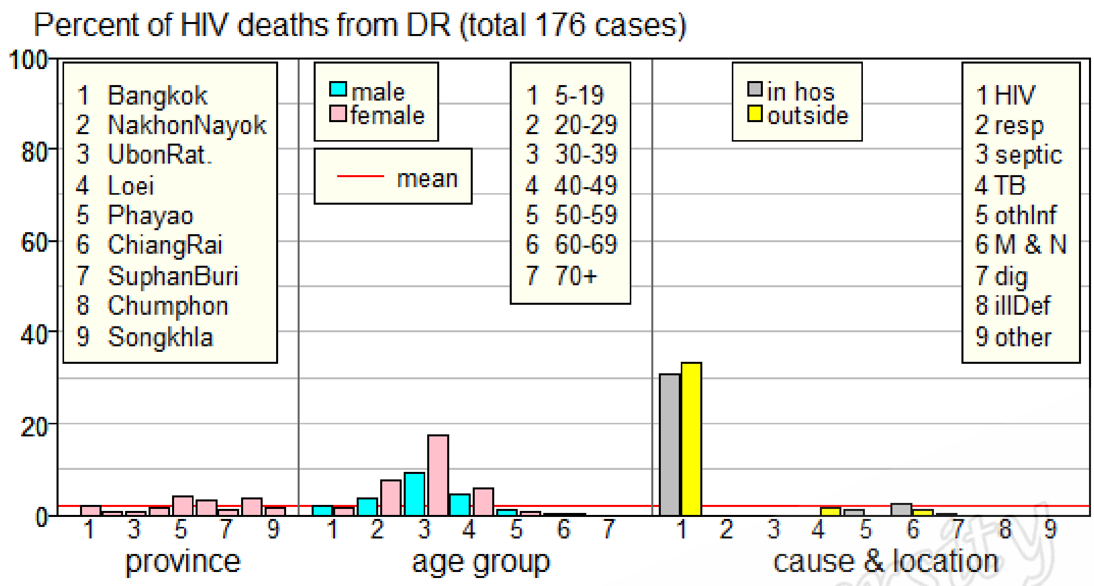

Figure 3.8 Percentage of HIV DR-reported deaths out of all deaths

3.2.2 Results

From the likelihood ratio test, Table 3.2 and 3.3 show deviance reduction from the simple cross-referencing model and the full model, respectively. The simple cross-referencing model with single DR cause-location factor reduced large deviance reduction with highly statistically significant of p-value. This indicated an important factor of DR cause-location. However, the model gave large error.

Table 3.2 Results from the simple cross-referencing model

Factors	Deviance reduction	df	p-value
DR cause-location	1583.5	17	< 0.0000001
Error	1522.2	1175	

Table 3.3 shows the results of the full model after added sex-age group and province factors. The p-values show that all three factors in the model are highly statistically significant. The error from the simple cross-referencing model (Table 3.2) is larger than the full model (Table 3.3) suggested that the full model is well predicted.

Table 3.3 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 3.9 shows crude percentages of HIV deaths from VA-assessed (among all deaths) by province, sex-age group, and DR cause-location group in bar charts 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 (5.39%) reflect the groups that were more likely to die from HIV.

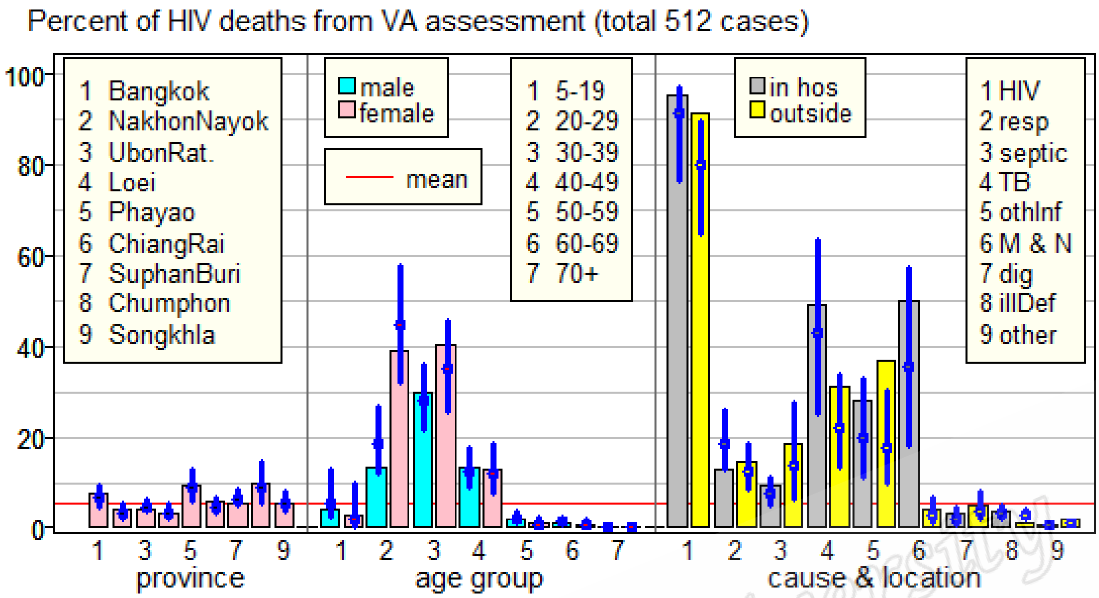

Figure 3.9 Estimated percentage of HIV-related deaths out of all deaths, based on the VA assessment (bars) and based on the full model (blue diamonds with error bars)

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 causes 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 hospital were reported as TB, other infectious diseases, septicemia and respiratory. 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 3.10 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 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.

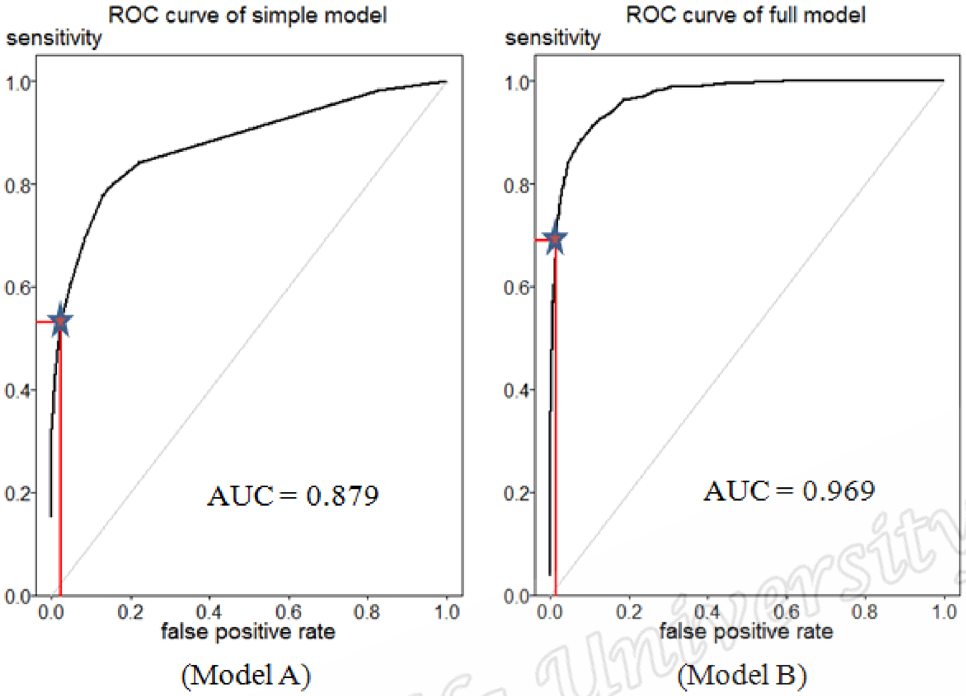

Figure 3.10 The ROC curve of logistic regression model with demographic covariates and without demographic covariates (Model A includes no demographic covariates whereas model B includes demographic covariates)

The full model (B) was then extended to all provinces in 2005. The left panel of Figure 3.11 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 equation (1-4, in method section 2.3.2), according to three levels of coefficients in 2005. 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. We will use this full model to generalize HIV mortality in all provinces from 1996-2009 in Part III.

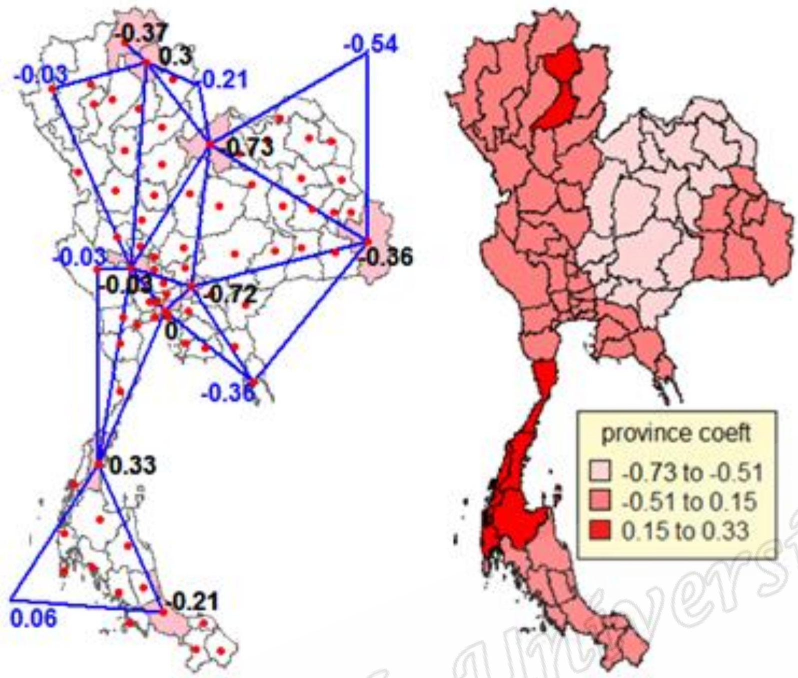

Figure 3.11 Triangulation for interpolating province coefficients in 2005

3.3 Part III: Exploring mortality patterns of all-cause and HIV from 1996-2009

3.3.1 Results *(Part III A: All-cause mortality patterns from 1996-2009)*

We graphed the changes in smoothed age distributions of deaths for 76 provinces by fitting cubic spline to cumulative death counts. Figure 3.12 and Figure 3.13 show three mortality patterns in spline-smoothed deaths for males and females from 1996-2009, respectively. The first pattern shows increasing high levels of infant mortality. The second pattern has quite similar patterns, with minor peaks between ages 20 and 30. The third pattern has very much higher peaks in this age range. Figure 3.12 shows male peaks occurred around age 25 in 1996 including Chiang Mai, Lamphun, Lampang, Phayao, Chiang Rai, Phetchaburi and Ranong, but these peaks had disappeared by 2009. Female peaks occurred in Chiang Mai, Phayao and Chiang Rai as shown in Figure 3.13.

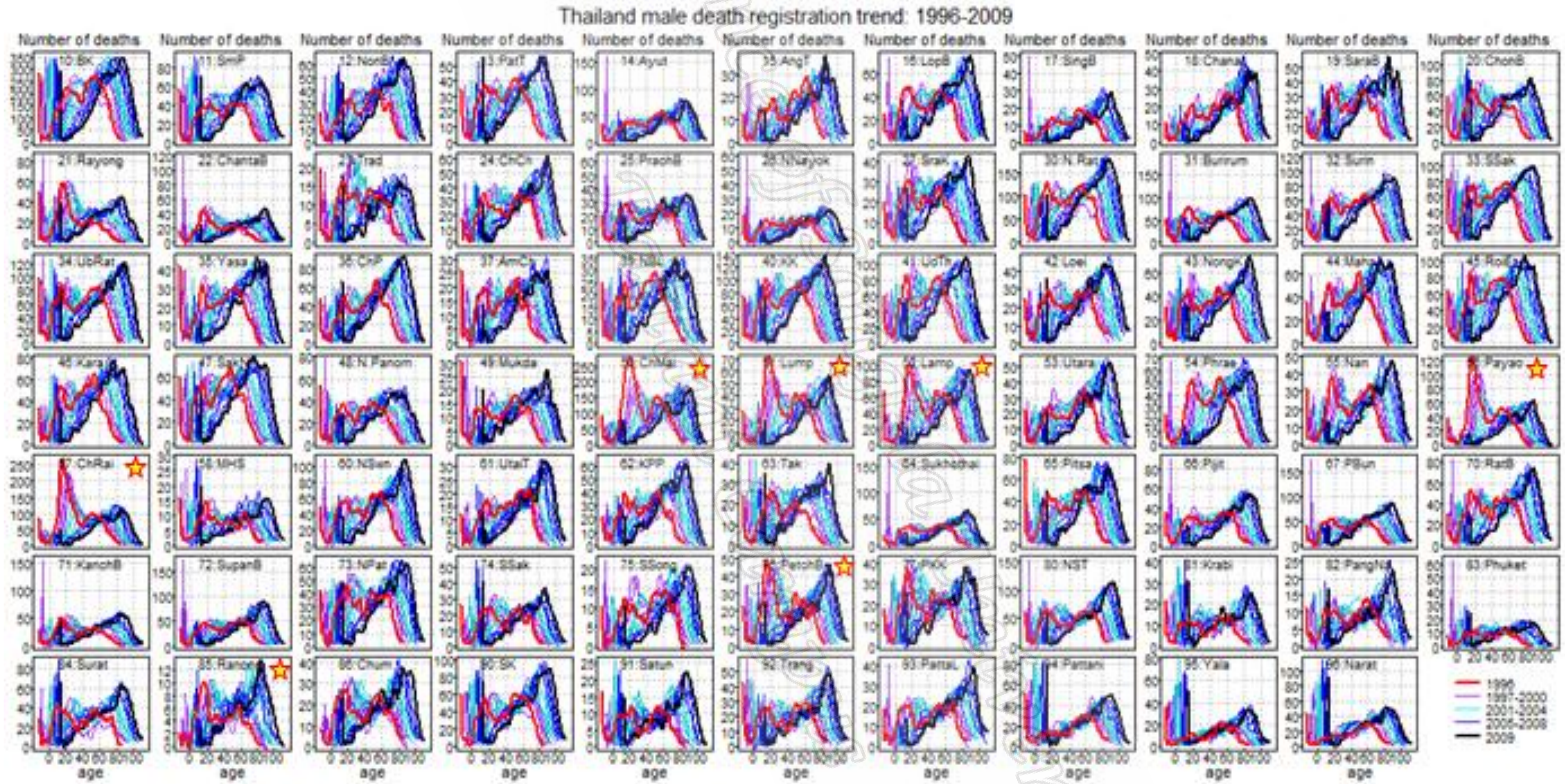

Figure 3.12 Thailand male death registration trend after corrected for unknown province and age, 1996- 2009

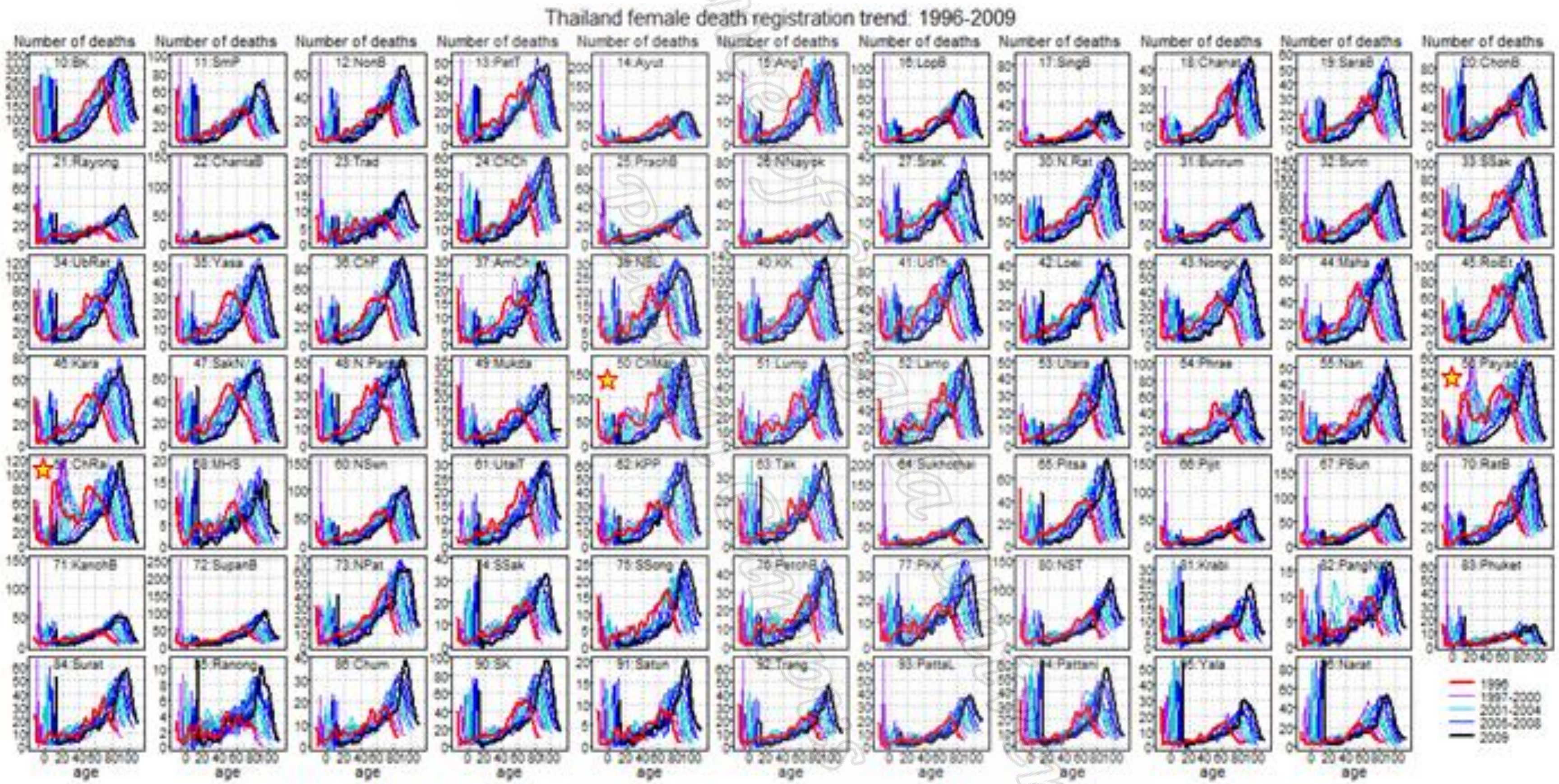

Figure 3.13 Thailand female death registration trend after corrected for unknown province and age, 1996- 2009

Note: province codes

10 Bangkok	31 Buri Rum	51 Lamphun	73 Nakhon Pathom
11 Samut Prakan	32 Surin	52 Lampang	74 Samut Sakhon
12 Nonthaburi	33 Sri Sa Ket	53 Uttaradit	75 Samut Songkhram
13 Pathum Thani	34 Ubon Ratchathani	54 Phrae	76 Phetchaburi
14 Ayutthaya	35 Yasothon	55 Nan	77 Prachuap Khiri Khan
15 Ang Thong	36 Chaiyaphum	56 Phayao	80 Nakhon Si Thammarat
16 Lopburi	37 Amnat Charoen	57 Chiang Rai	81 Krabi
17 Sing Buri	39 Nong Bua Lamphu	58 Mae Hong Son	82 Phang Nga
18 Chainat	40 Khon Kaen	60 Nakhon Sawan	83 Phuket
19 Saraburi	41 UdonThani	61 Uthai Thani	84 Surat Thani
20 Chon Buri	42 Loei	62 Kamphaeng Phet	85 Ranong
21 Rayong	43 Nong Khai	63 Tak	86 Chumphon
22 Chanthaburi	44 Maha Sarakham	64 Sukhothai	90 Song Khla
23 Trat	45 Roi Et	65 Phitsanulok	91 Satun
24 Cha Choeng Sao	46 Kalasin	66 Phichit	92 Trang
25 Prachin Buri	47 Sakon Nakhon	67 Phetchabun	93 Phatthalung
26 Naknon Nayok	48 Nakhon Panom	70 Ratchaburi	94 Pattani
27 Sa Kaew	49 Mukdahan	71 Kanchanaburi	95 Yala
30 Nakhon Ratchasima	50 Chiang Mai	72 Suphan Buri	96 Narathiwat

Now, we focused on three patterns in six provinces as shown in Figure 3.14. Narathiwat and Phuket show increasingly high levels of infant mortality. Bangkok and Khon Kaen have quite similar patterns, with minor peaks between ages 20 and 30. Chiang Rai and Chiang Mai have very much higher peaks between ages 20 and 30. 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.

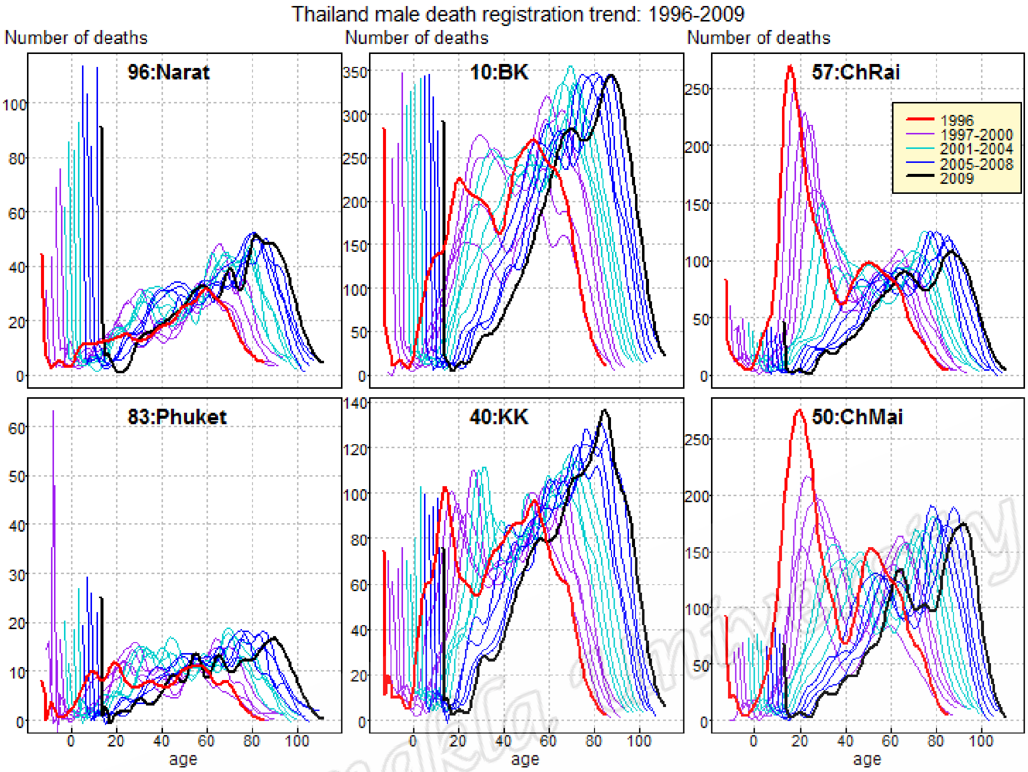

Figure 3.14 Trends in spline-smoothed deaths in six provinces from 1996-2009

We compared male and female patterns in Narathiwat, Bangkok and Chiang Rai provinces as shown in Figure 3.15. 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 were similar to the number of female deaths in all age groups in 2009. This implied the magnitude of young male deaths in Chiang Rai in 1996. It indicated that these deceased might be due to HIV or transport accident (need further analysis).

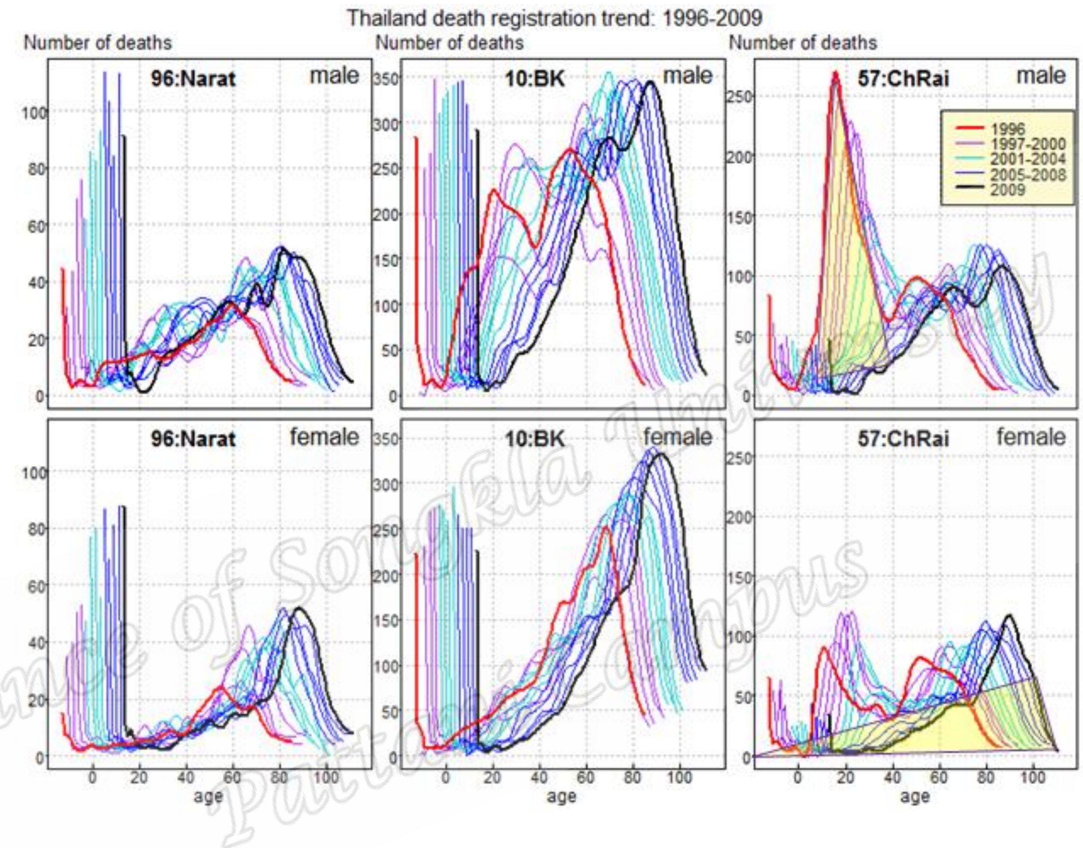

Figure 3.15 Trends in spline-smoothed deaths by sex from 1996-2009

3.3.2 Results *(Part III B: HIV mortality patterns from 1996-2009)*

The simple cross-referencing and the full model were then applied to the DR data for male and female deaths from 1996-2009 and plotted as area graphs in Figure 3.16. 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 3.09, respectively. 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.

The numbers of VA-estimated HIV deaths peaked in 2002 whereas DR-reported HIV deaths peaked in 2003. The VA-estimated HIV deaths predominantly affected age group 20-49 years (91.3% of males and 93.6% of females). Totally, the model B gave VA-estimated HIV deaths at 5 years and older for both sexes from 1996-2009 as 310,283 cases (205,071 males and 105,212 females) whereas DR-reported HIV deaths were 100,478 cases (66,160 males and 34,318 females). It shows substantial under-reporting for both sexes in all age groups by a factor of 3.09 (3.10 for male and 3.07 for female). This suggests that DR-reporting of HIV deaths account for only one third of all VA-estimated HIV deaths.

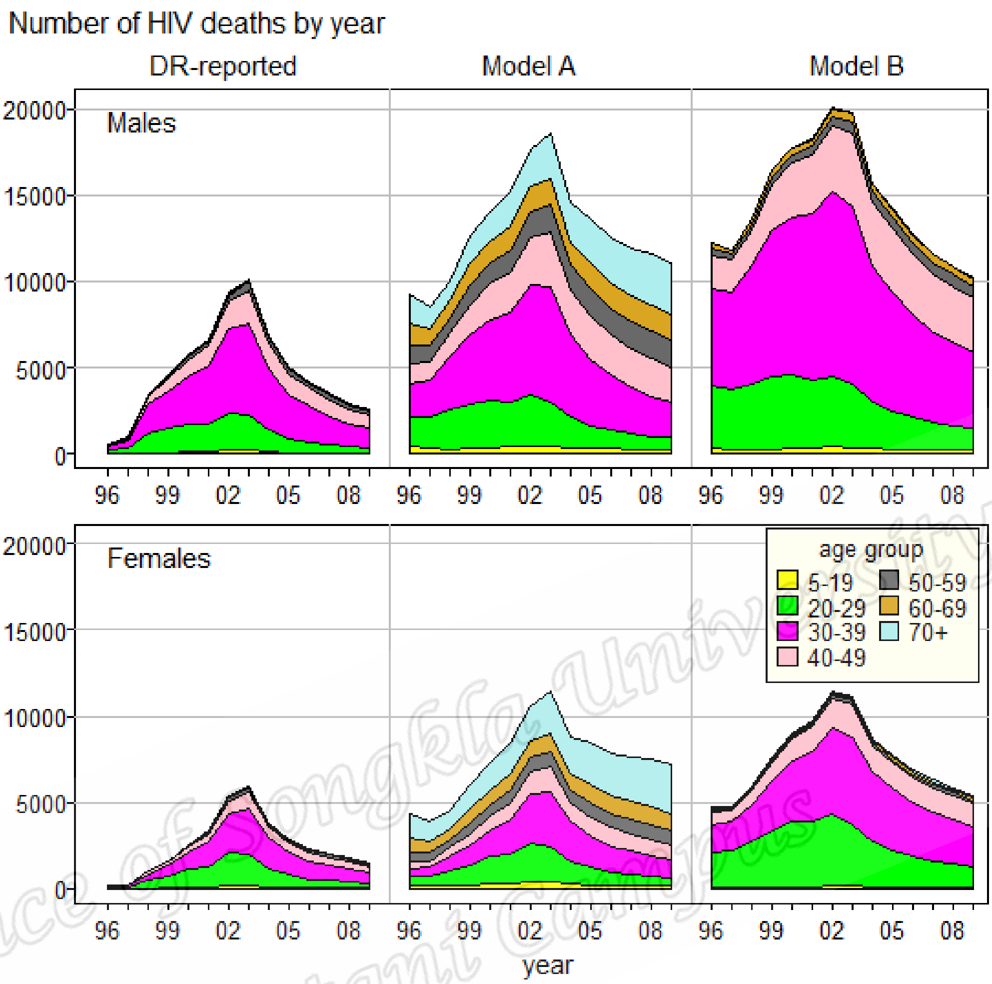

Figure 3.16 DR-reported and VA-estimated HIV deaths from the model with demographic covariates for both sexes, 1996-2009 (Model A includes no demographic covariates whereas model B includes demographic covariates)

Figure 3.17 shows HIV-estimated deaths variation in regions with time trend in three levels based on percentile rank. The dark shade, medium shade and light shade illustrated percentages of HIV deaths as high-level, medium-level and low-level, respectively. In 1996, HIV-estimated deaths mostly occurred in the upper North, slightly change until 2000, and clearly change in 2004-2009. Their trends show high-level in the North and moved to the South. However, Phayao, Rayong, Prachuap Khiri Khan, Chumphon and Ranong are still high-level from 1996 to 2009.

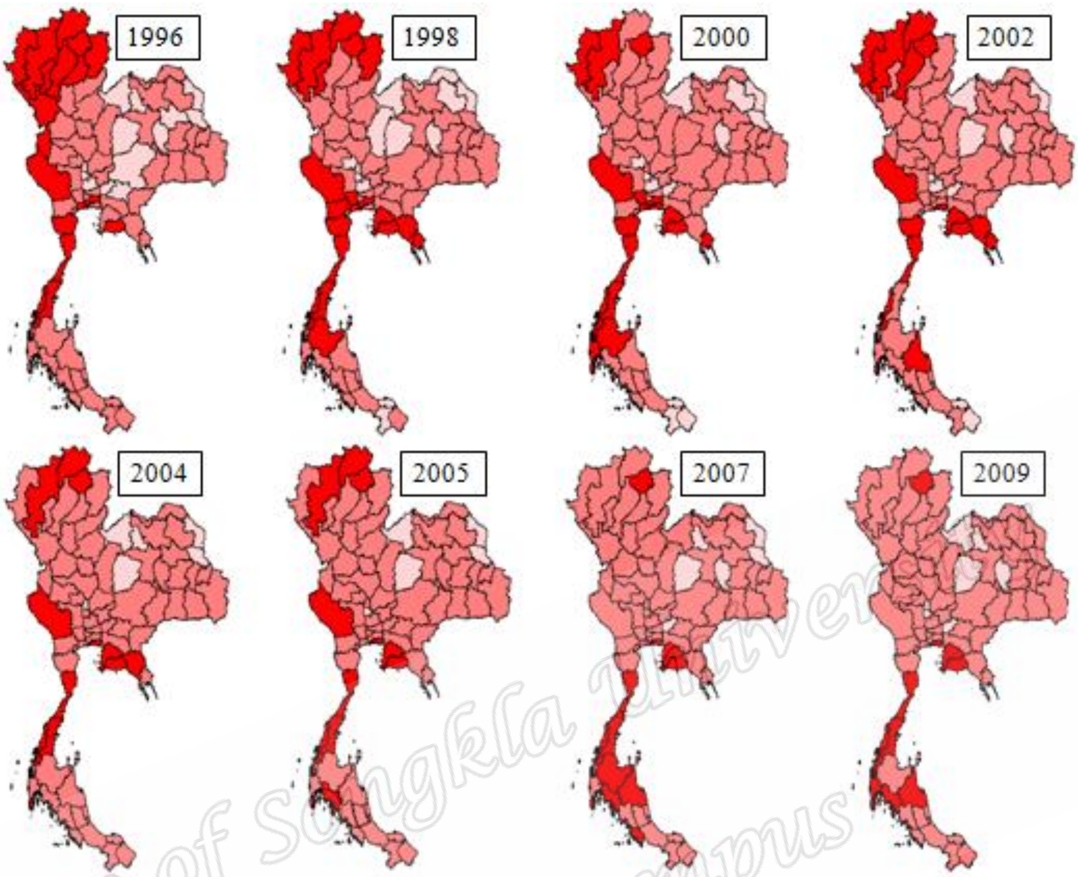

Figure 3.17 Trends of HIV-estimated deaths from 1996-2009