
Patama Vapattanawong & Pramote Prasartkul

Abstract

Objective To estimate the magnitude of under-registration of deaths, by age and sex, in Thailand.

Methods The data in this study were derived from two sources: the Thai Survey of Population Changes (SPC) 2005–2006, a consecutive multi-round household survey conducted over a 12-month period, and Thailand’s vital registration records. SPC death entries for people of all ages were matched to 2005–2006 death records from vital registration. The principles of a dual records system were applied to estimate the magnitude of under-registration of deaths, classified by age and sex, using the Chandrasekaran-Deming formula.

Findings Overall under-registration of deaths during 2005–2006 was 8.69% (95% confidence Interval, CI: 8.95–9.05) for males and 8.36% (95% CI: 8.31–8.41) for females. For both males and females, under-registration decreased as age increased. Under-registration was greatest among people of either sex aged 1–4 years, whereas it was < 10% among people 60 years of age and older, both males and females.

Conclusion These findings provided correction factors that can be used for adjusting mortality data from the registration system.

Introduction

Mortality is a demographic indicator not only of health outcomes in various population groups, but also of a country’s state of development. Death data are fundamentally crucial for identifying health problems and for monitoring health programmes. They are also used to measure and compare mortality rates in subpopulations within the same country or across different countries. Accordingly, the quality of mortality data is the key to reliable indicators and should therefore be carefully assessed before the data are used.
In Thailand, the most important source of mortality data is the vital registration 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, births must be registered within 15 days of delivery, while deaths must be registered within 24 hours. For deaths, the registration process is not straightforward because two steps are involved: (i) notifying an authorized person about the death, and (ii) validating and registering the death. After both registration steps are completed, the name of the deceased is deleted from the household roster and the details concerning the death 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.

In addition to the vital registration system, Thailand has a second important source of population data: the population census, which is conducted every 10 years. Theoretically, these two sources of demographic data should provide very similar population figures, but in reality the data sometimes conflict. During the first Population Seminar held in Bangkok in March 1963, disagreements arose surrounding the rate of population growth in Thailand because vital registration and population census data were in conflict. To learn the real rate of population growth, the National Statistical Office, which is in charge of the population census, was requested by the cabinet to conduct a national survey known as the Survey of Population Change (SPC). The first SPC was undertaken in 1964–1966. It later became an intercensal population survey conducted every 10 years at about the mid-period between the two censuses. The main purposes of the SPC are to obtain reliable estimates of vital event rates and of the rate of population growth, as well as to measure the under-registration of vital events. Thailand’s most recent SPC was conducted in 2005–2006.

Although estimating under-registration is a main objective of each SPC, the methods used in the surveys have not been the same. The differences in such methods are described in detail elsewhere. In brief, the estimation methods used in the first three SPCs (1964–1966, 1974–1976, 1984–1986) consisted of a dual records procedure, while those used in the 1995–1996 and 2005–2006 SPCs were based directly on survey questions, whether vital events were registered or not. According to the SPC results, during 1960s and 1970s approximately 60% of the deaths in the Thai population were registered. In the mid-1980s the fraction rose to 76% and in the two most recent SPCs (1995–1996 and 2005–2006) 95% completeness was reported. The very large increase in vital event registration observed in the 1995–1996
SPC relative to previous SPCs raised doubts, and the completeness of death registration in Thailand remains a controversial issue requiring further scrutiny.

This study aimed to investigate the quality of mortality data in Thailand’s civil registration system. Its objectives were to estimate the magnitude of under-registration of deaths and its associated age and sex patterns in 2005–2006 by cross-matching data from two sources.

**Methods**

**Data sources**

The two sources of data used in this study were the SPC 2005–2006 and vital registration records. The SPC 2005–2006 was a national, longitudinal, multi-round household survey with a large sample: 2050 sample rural villages/urban blocks and 82,000 households. This survey was designed to determine changes in household composition. To overcome the errors generated by the respondents’ limited ability to correctly recall events that occurred over an extended period, each sample household was visited five times within a year, or about once every three months. The survey period began in July 2005 and ended in July 2006.

All households in the sample rural villages and urban blocks were visited by the enumerators to obtain information on the usual household residents. In the initial interview, all persons residing in each household were listed and their basic demographic characteristics were recorded. A total of 327,735 individuals were counted in this first interview. As an additional check, to obtain as complete a coverage of births and deaths as possible, the enumerator also asked in every household: (i) if any women were known to be pregnant, and, if so, the number of months of pregnancy; and (ii) if any individuals were known to be seriously ill. For the second and subsequent interviews, the enumerator determined and recorded any changes in household composition that had taken place since the previous visit on account of births, deaths or movements in and out. If such changes had taken place, the date when they occurred was recorded. An additional question was asked about any babies born or deceased since the enumerator’s last visit.

To obtain vital registration data, individual death records for 2005–2006 from the interior ministry were used. Each record was in electronic format and had a 13-digit unique personal identifier (PID). Death records from before and after the SPC survey period were excluded from the study.
The individual records of the SPC had been linked to death records from vital registration using a 13-digit unique PID never recorded in previous SPCs except the last one. Of a total of 327,735 records in the SPC, 37,110 were excluded because their unique PIDs were missing or incomplete. Thus, only 290,625 records (88.7%) could be used to link to vital registration data. The method allowed for four possible outcomes from the cross-matching: (i) deaths reported in the SPC matched with deaths in the vital event registry; 2) deaths reported in the SPC but not matched to deaths in the registry; 3) people reported as survivors in the SPC but matched to deaths in the registry; and (iv) people reported as survivors in the SPC but not matched to deaths in the registry. All of these outcomes except the fourth were used to analyse the under-registration of deaths by age and sex.

**Data analysis**

This study applied the Chandrasekaran-Deming formula to carry out a dual record system estimation. The dual records system is a method used to estimate the completeness of a data source by cross-matching its records case by case against those of another source. We used the Chandrasekaran-Deming formula primarily to estimate the data that were missing from both systems, i.e. the number of people known to have died but not reported in either data source. We used this approach based on four crucial assumptions: (i) that the population of interest was closed; (ii) that events recorded in both systems could be cross-matched; (iii) that the probability of an event occurring in one system was independent of the probability of its occurring in the other, and (iv) that each case had an equal probability of ascertainment in either source. The under-registration of deaths was calculated by constructing a $2 \times 2$ contingency table (Table 1) containing the results of cross-matching the deaths from the SPC with those from vital registration records.

To estimate total events ($N$) according to the Chandrasekaran-Deming:

$$N = \frac{(M + NS) \times (M + NR)}{M}$$

where $M$ (matched) stands for the events found in both systems; $NS$ (not surveyed) stands for the events missed by the survey but found in vital registration records; and $NR$ (not registered) stands for the events missed by registration but found in the survey. An estimate of the variance of $N$ was calculated using the following formula:

$$Var(N) = \frac{Nq(1-q)}{pp}$$
where
\[ p_1 = \frac{M + NS}{N}, \]
\[ p_2 = \frac{M + NR}{N}, \]
and
\[ p_1 + q_1 = p_2 + p_1 = 1 \]

The 95% confidence intervals (CIs) of \( N \) were calculated using the formula:
\[ 95\% \text{ CI} = \hat{N} \pm 1.96\sqrt{\text{Var}(N)} \]

The formula of \( p_1 \) gave the estimated completeness of death registration. Thus, the under-registration (\( D_u \)) could be estimated from
\[ 1 - p_1 \text{ (or } q_1 \text{)} \]
and the variance of \( q_1 \) was calculated as:
\[ \text{Var}(q_1) = \frac{(1 - p_1)p_1}{N(1 - p_1)} \]

Results

Cross-matching of deaths
The diagram of the cross-matching process between deaths listed in the SPC and vital registration data is shown in Fig. 1. Of 290 625 individual records from the SPC, 1882 were deaths that could be matched to vital registration data; 179 were deaths that could not be matched; 556 were people listed as survivors in the SPC who could be matched to vital registration death records; and 288 008 were people listed as survivors in the SPC who could not be matched to vital registration death records (Table 2).

As shown in Table 2, the SPC listed a total of 2061 deaths as having occurred during 2005–2006. Over that period, 2438 deaths were included in the vital event registry. The figures in this table also reflect counting errors, since 556 people listed in the death registry were reported as being alive during the survey. The table also shows the magnitude of the under-registration: 179 people reported to be dead during the survey were not registered in the vital event registry.

Estimation of under-registration of deaths
A second $2 \times 2$ contingency table (Table 3) was created with the death data in Table 2 to determine how many people had been missed by both systems (the cell with ‘???’). The estimated total number of deaths occurring during the 2005–2006 survey and its 95% CI were:

$$N = \frac{(1882 + 556) \times (1882 + 179)}{1882} = 2670$$

$$95\% \text{ CI of } N = 2670 \pm 1.96 \times \sqrt{\frac{2670 \times \left(1 - \frac{1882 + 556}{2670}\right) \times \left(1 - \frac{1882 + 179}{2670}\right)}{\left(\frac{1882 + 556}{2670}\right) \times \left(\frac{1882 + 179}{2670}\right)}}$$

$$= 2653 - 2687$$

where the result was rounded to the nearest integer. In addition, the estimated under-registration of death, and its 95% CI could be calculated as:

$$D_r = \left[1 - \left(\frac{1882 + 556}{2670}\right)\right] \times 100 = 8.69\%$$

$$95\% \text{ CI of } D_r = 8.69 \pm 1.96 \times \sqrt{\frac{\left(1 - \frac{1882}{1882 + 179}\right) \times \frac{1882}{1882 + 179}}{2670 \times \left(1 - \frac{1882}{1882 + 179}\right)}} = 8.65\% - 8.72\%.$$

To estimate the under-registration of deaths by age and sex, we repeated the above calculation using data classified by age and sex instead. Table 4 presents the age- and sex-specific numbers of deaths in 2005–2006 as recorded by the SPC and vital registration records, and Table 5 shows the estimated percentage of under-registration of deaths. In 2005–2006 under-registration was 8.69% (95% CI: 8.65–8.72%) for both sexes combined and, for all ages overall, it was 9.00% (95% CI: 8.95–9.05%) for males and 8.36% (95% CI: 8.31–8.41%) for females.

When we stratified by broad age groups, the highest under-registration was found among children 1–4 years old: 54.55% (95% CI: 54.25–54.84%) in males and 71.43% (95% CI: 71.11–71.75%) in females. Under-registration of male and female deaths decreased as age increased, with the sole exception of deaths among females less than 1 year of age, which showed much lower under-registration than deaths among male infants (8.33%; 95% CI: 7.87–8.80% versus 34.78%, 95% CI: 34.48–35.08, respectively). Furthermore, deaths among older males and females showed less than 10% under-registration.
Discussion
The aim of this study was to estimate the magnitude of under-registration of deaths by making use of two valuable sources of data that already existed in Thailand. Although the method used herein has been known for more than half a century, it is still useful, particularly in countries having several sources of mortality data, such as Egypt and the Philippines. This methodology is frequently used in studies and is often referred in epidemiology and ecology as capture–recapture.

In Thailand, the dual records procedure or a similar method based on cross-matching of data from two different sources has been practised since the mid-1960s, and in that early period it was applied in a manner very similar to the one we have applied in this study in the sense that data from independent sources – in our case the survey and the vital event registration system – were cross-matched. However, some practical differences do exist. Whereas Decades ago the deaths reported in the survey were manually cross-matched with the ones included in the registration system on the basis of name and other characteristics common to both systems, such as sex, age of deceased, place of death and place of residence, in this study manual cross-matching was replaced by computerized cross-matching using a 13-digit unique PID. In addition, to ensure correct cross-matching, we double checked the age and sex of each pair of cross-matching cases. Thus, this study was stronger than earlier studies. Moreover, the dual records procedure made it possible to generate estimates for population subgroups. Thus it is possible to estimate under-registration of deaths at a subnational level or among subgroup of the population if the data are available.

The estimated under-registration of deaths for both sexes combined and for males and females separately was 8.69%, 9.00% and 8.36%, respectively. These figures are approximately 1.8, 1.7 and 1.9 times higher than the percentages estimated directly by Thailand’s National Statistical Office from the questions in the SPC. This is because our study captured more deaths, particularly 556 deaths that were in the vital event registration system but that had not been counted in the survey. Two factors could account for the substantial under-reporting of deaths in the SPC (~20% of all deaths in both systems): (i) the assumption that the population of interest was closed and (ii) the quality of the counting of deaths in the survey. The life/death status of the 556 people who were in the death registry but not in the SPC had been carefully reassessed in the fifth (last) visit paid to the households during the survey period (data not shown). Of these people, 78.8% were found to be regular residents, 17.4% had moved away and 3.8% were temporarily absent or transient residents.
Thus, the quality of the counting of deaths may have been the main reason that deaths were under-reported in the SPC, followed by the mistaken assumption that the population surveyed was a closed one. Similar findings have not been reported in other countries.\textsuperscript{11,19}

Our figures may be an underestimation of the actual rate of under-reporting. One reason is that not all SPC 2005–2006 sample data could be analysed because some information (11.3\%) was incomplete; either the PID was missing (10.3\%) or it had fewer than 13 digits (0.2\%) or was duplicated (0.8\%). Another reason is that only one unique identifier was used as a linkage key. If some identifiers were incorrectly recorded during the interview or when entered into the system, the results of the cross-matching could have been affected. Thus, this study arguably provides a minimum estimate of the rate of under-registration of deaths in Thailand during 2005–2006. Its findings lend support to the notion that the use of different estimation methods accounts for the large improvement in the completeness of death registration observed after the mid-1990s.

Conclusions
In Thailand, overall under-registration of deaths in 2005–2006 was around 8.69\%, but the percentage varied by age and sex. The age- and sex-specific estimates of under-registration from this study can be used as correction factors to improve the accuracy of the death statistics in the country’s registration system. Several studies have calculated the magnitude of under-registration overall, but few have made age- and sex-specific estimates. By using mortality data from two important sources – the SPC 2005–2006 and the vital registration system – this study provides some degree of assurance that the results are reliable. Age- and sex-specific numbers of deaths, after adjustment by the correction factors generated in this study, can be used as numerators to make various mortality estimates. The newly-adjusted age- and sex-specific death rates can even be used, for example, to construct life tables of Thailand’s general population and subpopulation groups.

Acknowledgements
We thank the Thailand’s National Statistical Office and the Ministry of Interior for allowing us to use the Survey of Population Change and the mortality data in the vital registration system for this study.

Competing interests:
None declared.

References


Table 1. **A 2 × 2 contingency table showing case-by-case cross-matching of deaths from the Survey of Population Change (SPC) and from the vital registration system, Thailand, 2005–2006**

<table>
<thead>
<tr>
<th>Death in SPC</th>
<th>Vital registration</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>NR</td>
</tr>
<tr>
<td>No</td>
<td>NS</td>
<td>MB</td>
</tr>
<tr>
<td>Total</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

*M*: matched (i.e. in both systems); *MB*: missing from both systems; *N*: number of events (estimated total); *NR*: not in registry (i.e. in survey only); *NS*: not surveyed (i.e. in vital registry only).

Table 2. **Life/death status of individuals obtained by cross-matching data from the Survey of Population Change (SPC) and the vital registration system, Thailand, 2005–2006**

<table>
<thead>
<tr>
<th>Status in SPC</th>
<th>Vital registration</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alive</td>
<td>No</td>
<td>556</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>556</td>
</tr>
<tr>
<td>Dead</td>
<td>No</td>
<td>179</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>1882</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2438</td>
</tr>
</tbody>
</table>

Table 3. **Number of deaths in 2005–2006, obtained by cross-matching data from the Survey of Population Change (SPC) and the vital registration system, Thailand, 2005–2006**

<table>
<thead>
<tr>
<th>Death in SPC</th>
<th>Vital registration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>

*a Missed by both systems.*
Table 4. **Number of deaths in 2005–2006, by age and sex, obtained by cross-matching data from the Survey of Population Change (SPC) and the vital registration system**

<table>
<thead>
<tr>
<th>Sex/death data details</th>
<th>No. of deaths</th>
<th>Age group (years)</th>
<th>&lt; 1</th>
<th>1–4</th>
<th>5–14</th>
<th>15–59</th>
<th>60–74</th>
<th>75+</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male deaths</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In both systems (M)</td>
<td>15</td>
<td>5</td>
<td>11</td>
<td>289</td>
<td>222</td>
<td>408</td>
<td>950</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In SPC only (NR)</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>29</td>
<td>18</td>
<td>24</td>
<td>94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In vital registry only (NS)</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>112</td>
<td>74</td>
<td>89</td>
<td>286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated total no. (N)</td>
<td>28</td>
<td>20</td>
<td>27</td>
<td>441</td>
<td>320</td>
<td>526</td>
<td>1358</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Female deaths</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In both systems (M)</td>
<td>11</td>
<td>2</td>
<td>4</td>
<td>138</td>
<td>205</td>
<td>572</td>
<td>932</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In SPC only (NR)</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>24</td>
<td>17</td>
<td>37</td>
<td>85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In vital registry only (NS)</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>75</td>
<td>73</td>
<td>117</td>
<td>270</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated total no. (N)</td>
<td>16</td>
<td>11</td>
<td>5</td>
<td>250</td>
<td>301</td>
<td>734</td>
<td>1312</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CI, confidence interval.

*These figures are not the sum of the values in each age group.

Table 5. **Estimated percentage under-registration of deaths, by age and sex, Thailand, 2005–2006**

| Age group (years) | Males | | Under-registration of deaths | | | Both sexes | | |
|------------------|-------|---|-----------------------------|---|---|-----------------|---|
|                  | %     | 95% CI | %     | 95% CI | %     | 95% CI |
| < 1              | 34.78 | 34.48–35.08 | 8.33 | 7.87–8.80 | 25.71 | 25.46–25.97 |
| 1–4              | 54.55 | 54.25–54.84 | 71.43 | 71.11–71.75 | 61.11 | 60.89–61.33 |
| 5–14             | 45.00 | 44.72–45.28 | 20.00 | 19.22–20.78 | 40.00 | 39.74–40.27 |
| 60–74            | 7.50  | 7.39–7.61  | 7.66  | 7.55–7.77  | 7.58  | 7.50–7.65  |
| 75+              | 5.56  | 5.47–5.64  | 6.08  | 6.01–6.15  | 5.86  | 5.81–5.91  |
| Overall          | 9.00  | 8.95–9.05  | 8.36  | 8.31–8.41  | 8.69  | 8.65–8.72  |

CI, confidence interval.
Fig. 1. Diagram showing cross-matching of deaths in the Survey of Population Change (SPC) and in the vital registration system, Thailand, 2005–2006

SPC 2005-2006 (n = 327,735)

Exclude

Incomplete PID (37,110 cases) Complete PID (290,625 cases)

Death registration (DR)

In 2005: 394,315 cases In 2006: 391,126 cases

Merged by 13-digit PID

Exclude deaths occurred before & after survey period

1) Deaths in SPC were matched to DR.
2) Deaths in SPC were not matched to DR.
3) Survivors in SPC were matched to DR.
4) Survivors in SPC were not matched to DR.

PID, personal identifier.