Capture-recapture to Estimate Completeness of Pulmonary Tuberculosis with Two Sources

RATCHADAPORN UNGCHAROEN,
LILY INGSRISAWANG
DEPARTMENT OF STATISTICS, FACULTY OF SCIENCE
KASETSART UNIVERSITY, THAILAND
Overview

Introduction
Objective
Methodology
Results
Discussion
Conclusion
References
Introduction

Tuberculosis (TB) is caused by a bacterium called “Mycobacterium tuberculosis”. The bacteria usually attack the lungs ➔ Pulmonary tuberculosis (PTB)

But TB bacteria can attack any part of the body such as the kidney, spine, and brain ➔ Extra-Pulmonary tuberculosis

TB is spread through the air from one person to another. The TB bacteria are put into the air when a person with TB disease of the lungs or throat coughs, sneezes, speaks, or sings. People nearby may breathe in these bacteria and become infected.
Introduction

In 2013,

- 6.10 million cases of TB were reported to WHO
- 5.7 million were newly diagnosed cases, only 0.1 million cases being already in treatment.
- Most of estimated number of cases in 2013 occurred in Asia (56%).
- Thailand is 1 of 22 countries the high-burden countries (HBCs).
- WHO estimated the new Tuberculosis cases about in Thailand about 80,000 cases/year.
- Ministry of Public Health showed the reported the patients registration about 60,000 cases/year, an approximate 50% was pulmonary tuberculosis with positive smear.
Introduction

Nakhon Pathom Province

Suphanburi province

Kanchanaburi province

Kampangsan

Banglen

Dontoom

Nontaburi province

Ratchaburi province

Mueng

Nakhonchaisri

Pudtamontol

Sampran

Samutsakorn province

Bangkok

Introduction

In 2014, the report data migrant workers was 74,944 in Nakhon Pathom province.

Have Visa Non-immigrant-B for work was 56,688
- Myanmar was 52,071
- Cambodia was 2,772
- Laos was 1,845

No registration for migrant workers was 25,148
- Myanmar was 14,991
- Cambodia was 6,861
- Laos was 14,991
Objective

To estimate the completeness of pulmonary tuberculosis cases confirmed by two sources.
Methodology

Data sources

Source 1: The Hospital registration with pulmonary tuberculosis
Source 2: The Surveillance system by Nakhon Pathom Public Health Office with pulmonary tuberculosis

Case Identification

PTB cases are diagnosed by National Tuberculosis Control Programme of Thailand Guidelines.
Methodology

Record-linkage

Source 1 is the Hospital registration
Source 2 is the Surveillance system by Nakhon Pathom Public Health Office
Methodology

Capture-recapture methods

Basic Concept

Capture-recapture is an indirect method to estimate the number of cases based on the degree of overlap between two or more samples.

Assumptions:

(i) Independent between sources
(ii) Closed population: no change in the population during the period under investigation
(iii) Independent between individuals
Capture-recapture

Close population

- Two sources or Dual systems (Lincoln-Petersen)
- More two sources or Multiple systems

Combination closed and open Popn models

Open population

- Survival-Cormack-Jolly-Seber model
- Popn Size and Recruitment-Full Jolly-Seber Model
- Multi-state Models
Methodology

Classical model

Lincoln-Petersen model

\[ \hat{N}_{LP} = \frac{n_1 \times n_2}{n_{11}} \]

Chapman’s modification to reduce bias

\[ \hat{N}_{CPM} = \frac{(n_1 + 1)(n_2 + 1)}{n_{11} + 1} - 1 \]

<table>
<thead>
<tr>
<th>Source</th>
<th>Yes (1)</th>
<th>No (0)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes (1)</td>
<td>( n_{11} )</td>
<td>( n_{01} )</td>
<td>( n_2 )</td>
</tr>
<tr>
<td>No (0)</td>
<td>( n_{10} )</td>
<td></td>
<td>( x )</td>
</tr>
<tr>
<td>Total</td>
<td>( n_1 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chao’s lower bound

Brittain & Bohning, (2009); proposed the mixed binomial with 2 parameter;

\[ \hat{N}_C = n + n_1^2 / 4n_{11} \]

where \( n = n_{11} + n_{10} + n_{01} \)

Seber (1970) & Wittes (1972) suggested the variance that

\[ \text{Var}(\hat{N}) = \frac{(n_1 + 1)(n_2 + 1)n_{10}n_{01}}{(n_{11} + 1)^2(n_{11} + 2)} \]
**Methodology**

Dependence source?

\[ ORs = \frac{n_{11} x}{n_{01} n_{10}} \]

<table>
<thead>
<tr>
<th>Source 1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes (1)</td>
<td>No (0)</td>
</tr>
<tr>
<td>n_{11}</td>
<td>n_{01}</td>
</tr>
<tr>
<td>n_{10}</td>
<td>x</td>
</tr>
<tr>
<td>n_{1}</td>
<td></td>
</tr>
</tbody>
</table>

From the study of Ballivet S. et. al., 2000, CR method : application to a thyroid cancer registry.

Example: ORs = 0.02 (95%CI: 0.00-0.05) → Negative dependence → overestimate “. ” and underestimates the completeness.

ORs = 8.67 (95%CI: 3.63-20.71) → Positive dependence → underestimate “. ” and overestimates the completeness.

ORs = 4.83 (95%CI: 0.77-30.45) → None dependence
## Results

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Estimated PTB cases N (95%CI)</th>
<th>Completeness of source (%)</th>
<th>OR (95%CI)</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Registration Hospital (S1)</td>
<td>Surveillance system (S2)</td>
<td>Both sources (S1 &amp; S2)</td>
</tr>
<tr>
<td>Original Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chapman</td>
<td>360 (349-371)</td>
<td>66.9</td>
<td>84.7</td>
<td>57</td>
</tr>
<tr>
<td>Chao</td>
<td>365 (354-376)</td>
<td>66</td>
<td>83.6</td>
<td>55.9</td>
</tr>
<tr>
<td>Bootstrap</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chapman</td>
<td>347 (309-385)</td>
<td>69.5</td>
<td>87.9</td>
<td>58.8</td>
</tr>
<tr>
<td>Chao</td>
<td>353 (315-390)</td>
<td>68.3</td>
<td>86.4</td>
<td>57.8</td>
</tr>
</tbody>
</table>
Discussion

- The violation of the assumption of independence between two sources. The results showed positive dependent. The estimated total number cases will be underestimated.


- The assumption is more likely to be violated in human populations than animal populations.[Ernest B. Hook (1995), Brittain & Bohning (2009) and Van Hest (2008)].

- The Chao’s estimator is less biased than the Chapman estimator and is just as simple to calculate.

- Hook & Regal (2012) suggested that the formal analysis of dependencies could be used to define boundaries around the capture-recapture estimates.
Conclusion

➢ The dependence sources were found positive dependence.

➢ Both estimators of Chapman and Chao have not considered other variables in the model and may be consider the covariate in the model.

➢ For future research, we should considered to add the potential covariates in the estimation model.
References


References


References


We welcome your Question, Suggestion and comments.