Design of 2\textsuperscript{nd} survey in Cambodia

WHO workshop on Repeat prevalence surveys in Asia: design and analysis
8-11 Feb 2012
Cambodia

Norio Yamada
RIT/JATA
Primary Objectives

1. To determine the prevalence of pulmonary TB among the population aged 15 years or older at a defined point in time (2010) in Cambodia as measured by:

2. To assess the trend in TB prevalence
1\textsuperscript{st} survey: Area and Population

Target areas: \textit{4 remote provinces excluded}

Target population: \textit{aged} \geq 10\text{yrs} (tuberculin survey: aged \leq 14\text{ yrs})

42 sites (clusters): 7 in urban, 35 in rural (allocation of clusters to each strata by population size of strata): 720 (all age group) x 42 = 30240

This is based on expected prevalence(483.2), 25\% Relative precision, DEFF=1.25: not taking into account future comparison
Survey Areas and population

• Areas: The whole country
  The same areas as the 1\textsuperscript{st} survey (comparable with 1\textsuperscript{st} survey), consisting of two strata
  Urban
  Rural
  
+ The areas excluded in the 1\textsuperscript{st} survey

• Target population: Aged \geq 15\text{yrs}
## Survey Areas and Population

### 1st and 2nd survey

<table>
<thead>
<tr>
<th>Survey Areas</th>
<th>Population</th>
<th>1st survey</th>
<th>2nd survey</th>
<th>Use of data of 2nd survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>National prevalence comparison with 1st survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-14yrs</td>
<td>Not included</td>
<td>Not included</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15yrs-</td>
<td>Not included</td>
<td>Included</td>
<td>used</td>
<td></td>
</tr>
</tbody>
</table>

### Other areas

<table>
<thead>
<tr>
<th>4 Remote provinces (Stratum-3)</th>
<th>10-14yrs</th>
<th>Not included</th>
<th>Not included</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15yrs-</td>
<td>Not included</td>
<td>Included</td>
<td>used</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Urban (stratum-1)</th>
<th>10-14yrs</th>
<th>Included</th>
<th>Not included</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15yrs-</td>
<td>Included</td>
<td>Included</td>
<td>used</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rural (Stratum-2)</th>
<th>10-14yrs</th>
<th>Included</th>
<th>Not included</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15yrs-</td>
<td>Included</td>
<td>Included</td>
<td>used</td>
<td></td>
</tr>
</tbody>
</table>


Requirement of sample size
Urban/Rural Strata

1) to obtain 25% relative precision for expected range of prevalence in study population
   – Range: between no change and 42% reduction (corresponding to 50% reduction in 10 years) from the 1st survey. 441.7 – 256.2

2) to detect difference in prevalence in population aged >= 15yrs between two surveys if prevalence in the 2nd survey is 58% of prevalence in the first survey (42% reduction):
   • 441.7 vs 256.2 with type-I error of 5% (two side) and 80% Power
Procedures of sampling calculation

• Calculate SRS-sample size for the requirement-1 using formula-5.1 (Limebook): 23,932

• Calculate effective sample size of the 1st survey:
  – The number of participants / DEFF = 15,251

• Calculate SRS sample size for the objective-2 under effective sample size of 1st survey using Stata command corresponding to formula-9.2 (Limebook): 18596

• Compare two sample sizes and select the larger one (=sample size for requirement-1)

• See different combinations of cluster size and the number of clusters taking into account DEFF (depending on cluster size) to obtained the SRS sample size of the 2nd survey.
Assumption of DEFF in 2\textsuperscript{nd} survey

- DEFF = 1+(cluster size –1) x ICC  
  (ICC: intraclass correlation coefficient)
- DEFF of the 1\textsuperscript{st} survey: 1.15664
- Intraclass correlation coefficient (ICC) calculated from DEFF:0.000747685
- Adopt \{2 x ICC\} of 1\textsuperscript{st} survey as rather conservative ICC: 0.000747685 x 2
- Calculate DEFF for 2\textsuperscript{nd} survey by using the formula: DEFF = 1+(cluster size –1) x ICC
Cluster size and No. Clusters

• What was considered:

• What cluster size is acceptable?
  – Determined by acceptable workload in each cluster
  – Acceptable range of participants: say < 150-180 (max 200) per day x 4 days: 600-800
  – the average village population aged ≥15 years or more: 632.
  – A cluster size of 600-650 is appropriate if each cluster's operation is to be completed within a week.

• How many clusters are acceptable?
  – More incremental increase in budget by no. clusters than cluster size
  – Schedule of Global fund
  – Actual operation 8 months = 32 weeks -> about 60 clusters might be appropriate if two team operation
• Relation of Cluster size, No. clusters and DEFF

\[ \text{No. Clusters} = \frac{\text{SRS sample size} \times \text{Deff}}{\text{Cluster size}} \]

\[ \text{Deff} = 1 + (\text{cluster size} - 1) \times \text{ICC} \]

• Final cluster size = cluster size / participation rate(0.9)

• Total sample size = final cluster size x no. clusters

• We checked whether cluster size is acceptable if 60 clusters

  • 640 per cluster x 60 clusters: judged acceptable

  • sample size for urban/rural: 38400
Sampling Methods (1)

- Stratified cluster sampling with PPS
- 60 clusters are allocated for rural and urban strata into
  - Urban: 13 clusters
  - Rural: 47 clusters
- Additional 4 provinces excluded in 2002 survey: 2 clusters, which is proportional to population
- Total sample size = 62 x 640 = 39680
Sampling Methods (2)

• Sampling frame: Census 2008 data
  – Nested list of province, district, commune and village with population for each of strata.
  – Randomly listed within each level
  – Calculate cumulative number of population from the top of list to the end
Sampling Methods (3)

- Draw one number between 1 and pop size and select district(PSU)/commune(SSU)/village with cumulative number including this random number
- Select next district/commune/village systematically with the interval (Stratum population / No. cluster in strata) in the same way
Target areas: the whole of country

Target population for TB prevalence survey: aged $\geq 15$yrs

62 sites (clusters): 13 in urban areas, 47 in rural areas, 2 in remote province (allocation of clusters to each strata by population size of strata)
Sampling Methods (4)

• Last stage of sampling
  – If selected village has population larger than 800, households were grouped
    • one group was drawn randomly
    • Direction of next village is decided randomly to reach the sample size close to 640 (610-670)
  – If selected village has population smaller than 600, select villages within the same commune.
    • Select randomly one village neighboring the selected village and select village clockwise until reaching sample size (610-670)
During survey implementation

Braboeung village: 4 blocks: 262+500+400+378 peoples all
SdaeungChey commune
CheungPrey district
Kg Cham province

Cluster # 36

Block 250~262 from list
Theoretical issues

• In the last stage of sampling, sampling probability is not the same regardless of different situations while it may not be so.
  – Up to commune level, designed to be self-weighted

• Point estimates and variances are calculated as if sampling were made by PPS with replacement.
  – Might be acceptable because overall sampling fraction is not large and units at each level are randomly listed
Major Differences between two survey (1)

• Age group of survey population
  – 1st survey: >= 10yrs
  – 2nd survey: >= 15yrs

• Area
  – 1st survey: 4 provinces are excluded
  – 2nd survey: the whole country

• Sampling frame
  – 1st survey: all age group
  – 2nd survey: aged >= 15yrs

• Stratification: Definition of urban and rural is different
Major Differences between two survey (2)

- **Sampling methods**
  - 1\textsuperscript{st} survey: selecting district with PPS -> randomly selecting village
  - 2\textsuperscript{nd} survey: PPS up to village level

- **Screening**
  - Symptom criteria:
    - 1\textsuperscript{st} survey: duration of cough $\geq 3$wks
    - 2\textsuperscript{nd} survey: duration of cough $\geq 2$wks
  - Smear examination
    - 1\textsuperscript{st} survey: all sputum examined by ZN
    - 2\textsuperscript{nd} survey: all sputum examined by FL -> recheck (positive) by ZN
Thank You