TB incidence estimation

Philippe Glaziou
Glion, March 2015
Outline – main methods to estimate incidence

• Main methods in 6 groups of countries, based on available TB data
• HIV disaggregation
Group 1 – high income
Decline in TB burden in England and Wales

- TB mortality: -3%/year
- TB incidence: -9.3%/year

Graph showing the decline in TB burden from 1920 to 2000.
High-income settings

- Limited under-diagnosis assumed
- Undocumented under-reporting

\[ I = uf(N), u \in [1,1.3] \]

<table>
<thead>
<tr>
<th>( I )</th>
<th>incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>case notifications</td>
</tr>
<tr>
<td>( f )</td>
<td>cubic spline function in countries with large year-to-year fluctuations in ( N ), or else, the identity function</td>
</tr>
</tbody>
</table>
Group 2 – capture-recapture
Inventory studies

1. Evaluate under-reporting

2. Estimate incidence
   - Retrospective record-linkage
   - Prospective, sampling fraction > 0.5
Capture-recapture in 20 seconds

Assuming independence between events A and B,

\[ P(A \cap B) = P(A) \cdot P(B) \]

\[ \frac{N_{AB}}{N} = \frac{N_A}{N} \cdot \frac{N_B}{N} \]

\[ N = \frac{N_A}{N_{AB}} \cdot N_B \]

Total TB cases = \(N\)

\(N_A\) non-NTP

\(N_{AB}\)

\(N_B\) Reported
A simple model for 3 lists

• For an incomplete $2^3$ contingency table

$$\log E(Z_{ijk}) = u + u_1 I(i = 1) + u_2 I(j = 1) + u_3 I(k = 1) +$$

$$u_{12} I(i = j = 1) + ... + u_{123} I(i = j = k = 1)$$

Log of unobserved cases

• Model with 8 terms
  – Number expected in all list ($u$)
  – 3 main effects, log odds of appearing in list 1, 2, 3
  – 3 two-factor interactions $u_{12}$, $u_{13}$, $u_{123}$
  – 1 three-factor interaction, assumed zero
Capture-recapture in Iraq (2011)

(Int J Tuberc Lung Dis 2013; 17: 462–7.)

1980 cases detected, **under-reporting** = 16%
473 additional cases estimated (394–565)
One of those things may break CR

- Record linkage misclassifications
- Failure to link with records in non-sampled areas, or before/after the study period
- Unobservable events
- Non closed population
- Heterogeneity
- Between list dependences
- Ill-fit due to small numbers
Inventory studies

- Planned: 6 (including 1 repeat)
- Completed: 8
Nested inventory study in Indonesia: what did it tell us?
WHO estimates of incidence in Indonesia are implausibly low

- Incidence: $183 \ (159 \ - \ 204)$ per 100,000/year (2013)
- Case notifications: 130/100,000 (2013)
- Under-reporting (prevalence survey): 56%
- Treated in 2013:
  
  $130 / (1 - 0.56) = 296/100,000$
Group 3 – prevalence surveys
Two methods to derive incidence from prevalence

1. Based on the ratio untreated / treated, using a simple deterministic model

2. Based on standard assumptions* about disease duration in 4 different case categories

incidence ≈ prevalence / duration

Method 1

\[
\frac{dU}{dt} = I - (\mu_U + \theta_U + \delta)U
\]

\[
\frac{dT}{dt} = \delta U - (\mu_T + \theta_T)T
\]

At equilibrium,

\[
I = \frac{U}{\frac{dU}{dt}}
\]

\[
\delta U = \frac{T}{\frac{dT}{dt}}
\]

\[
d_U = (1 - \pi) \frac{U}{T} d_T
\]

\[
\pi = \text{proportion of incidence that dies or self-cures before being treated}
\]
## Method 1

\[ \pi \sim U(0, 0.1) \]

<table>
<thead>
<tr>
<th></th>
<th>( U )</th>
<th>( T )</th>
<th>Prevalence (per 1000)</th>
<th>Duration (year)</th>
<th>Incidence (per 1000/yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cambodia 2002</td>
<td>260</td>
<td>42</td>
<td>12 (10-15)</td>
<td>2.9 (1.9-4)</td>
<td>4 (2.5-5.8)</td>
</tr>
<tr>
<td>Cambodia 2011</td>
<td>205</td>
<td>80</td>
<td>8.3 (7.1-9.8)</td>
<td>1.2 (0.8-1.6)</td>
<td>6.7 (4.5-9.3)</td>
</tr>
<tr>
<td>Myanmar 2009</td>
<td>300</td>
<td>79</td>
<td>6.1 (5-7.5)</td>
<td>1.8 (1.1-1.6)</td>
<td>3.3 (2-4.8)</td>
</tr>
<tr>
<td>Thailand 2012</td>
<td>136</td>
<td>60</td>
<td>2.5 (1.9-3.5)</td>
<td>1.1 (0.5-1.6)</td>
<td>2.3 (1-3.5)</td>
</tr>
<tr>
<td>Indonesia 2013</td>
<td>407</td>
<td>122</td>
<td>6.6 (5.2 – 8.1)</td>
<td>1.6 (1 – 2.2)</td>
<td>4.1 (2.4 – 5.8)</td>
</tr>
</tbody>
</table>
Limitations of method 1

- Information on culture positivity for self-reported $T$ cases not available (smear microscopy only)
  - $T$ may be over-estimated
  - $d_u$ and Incidence estimates may be biased
- Surveys not powered to estimate $U/T$ with precision
Method 2

\[ I \approx \sum_{i,j} \frac{P_{i,j}}{d_{i,j}} \]

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Notified HIV-</td>
</tr>
<tr>
<td>Notified</td>
<td>HIV+</td>
<td>U (0.2 - 2) y</td>
</tr>
<tr>
<td>Not notified</td>
<td>HIV-</td>
<td>U (1 - 4) y</td>
</tr>
<tr>
<td>Not notified</td>
<td>HIV+</td>
<td>U (0.01 - 0.2) y</td>
</tr>
</tbody>
</table>
Limitations of method 2

- Invariance over space and time
- Biases of unknown direction and magnitude
Two methods compared: overlapping uncertainty ranges

<table>
<thead>
<tr>
<th></th>
<th>Prevalence (10^{-3})</th>
<th>Incidence – method 1 (10^{-3} y^{-1})</th>
<th>Incidence – method 2 (10^{-3} y^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cambodia 2002</td>
<td>12 (10-15)</td>
<td>4 (2.5-5.8)</td>
<td>2.2 (1.5 – 2.9)</td>
</tr>
<tr>
<td>Cambodia 2011</td>
<td>8.3 (7.1-9.8)</td>
<td>6.7 (4.5-9.3)</td>
<td>3.8 (2.2 – 5.8)</td>
</tr>
<tr>
<td>Myanmar 2009</td>
<td>6.1 (5-7.5)</td>
<td>3.3 (2-4.8)</td>
<td>3.4 (2 – 5.1)</td>
</tr>
<tr>
<td>Thailand 2012</td>
<td>2.5 (1.9-3.5)</td>
<td>2.3 (1-3.5)</td>
<td>1.1 (0.7 – 1.6)</td>
</tr>
<tr>
<td>Indonesia 2013</td>
<td>6.6 (5.2 – 8.1)</td>
<td>4.1 (2.4 – 5.8)</td>
<td>4.3 (2.2 – 7.2)</td>
</tr>
</tbody>
</table>
Estimated incidence, ensemble model

Ensemble 402 (276 - 552) per 100,000/year
Ensemble - principles

Incidence distributed Beta, \( I_i \sim B(a_i + 1, b_i + 1) \)

Two estimates \( I_1 \) and \( I_2 \), so that

\[
Prob(x = TB) = \int_0^1 x \ B(a_i + 1, b_i + 1) \ dx = \frac{a_i + 1}{a_i + b_i + 2}
\]

Let \( c = \sum a_i \) and \( d = \sum b_i \)

**Combined estimate**

\[
Prob(x = TB) = \frac{c + 1}{c + d + 2}
\]

\[
Var = \frac{(c + 1)(d + 1)}{(c + d + 2)^2(c + d + 3)}
\]
Incidence estimated from prevalence

- Uncertainty in estimated incidence due to:
  - Sampling uncertainty about prevalence of B+ in adults
  - Uncertainty about extra-pulmonary and childhood TB
  - Uncertainty about disease duration

- Combine two estimation methods (ensemble approach)
Group 4 – notifications + VR
Methods

• Use VR data to inform expert opinion about incidence and its time trends (former USSR)

• Internal consistency, where $I^- > 0$

\[
0 < \frac{M^-}{I^-} < 0.45
\]
Rise and fall in mortality in most former USSR countries.
Incidence and notifications (Russia)

Incidence (all)

Notifications (all)

Incidence (HIV-pos)

Rate per 100,000 population/year

Group 5 – notifications + ARI
Available data on trends in incidence in India

- Trends in ARI: \(-0.037 \, y^{-1} (-0.051, -0.024)\)
- Trends in Notification Rates in cohorts of DOTS implementation
  - 1999 districts (n=62): \(-2\% / yr\)
  - 2000 districts (n=95): \(-1.8\% / yr\)
  - 2001 districts (n=67): \(-3.6\% / yr\)
  - 2002 districts (n=75): \(-2.1\% / yr\)
  - 2003 districts (n=129): \(0.6\% / yr\)
Assumptions on trends, India

• Flat over 1954–2001
• Two sources: repeat ARI, average trend in DOTS districts
• Combined using a Bayesian model
  – Prior based on repeat ARI measures
  – Data from DOTS districts
  – Posterior mean rate of change
    \[ r = -1.5 \times 10^{-2} \text{ y}^{-1} \]
Group 6 – case notifications
Only case notifications (and HIV estimates)

• Epi reviews
  – Eliciting expert opinion on plausible range of CDR (regional workshops: 3 time points)

• In the absence of reliable data to estimate trends
  – Assume flat trend
Based on case notifications

- Expert opinion - non reproducible
- Too few experts
  - Conflict of interest?
  - Variance in opinion
  - Bias?
- Over-diagnosis among notified cases not accounted for (except for Kazakhstan)
Incidence disaggregated by HIV
Risk ratio \( \frac{\text{HIV}^+}{\text{HIV}^-} \)

\[
\rho = \frac{I^+/N^+}{I^-/N^-} > 1
\]

\[
\rho \frac{I^-}{I^+} = \frac{N^-}{N^+}
\]

\[
I^+ = \frac{\rho N^+/N}{1 + (\rho - 1)N^+/N}
\]

\[
I = \frac{I^+}{I} = \frac{\rho N^+/N}{1 + (\rho - 1)N^+/N}
\]

\[
t = \frac{I^+}{I} = \frac{h\rho}{1 + h(\rho - 1)}
\]

\[
\log(\rho) = \log \left( \frac{t}{1-t} \right) - \log \left( \frac{h}{1-h} \right), (t, h) \in ]0,1[
\]

| \( t = \frac{I^+}{I} \) | Prevalence of HIV among incident TB |
| \( \rho \) | Incidence Rate Ratio |
| \( h = \frac{N^+}{N} \) | Prevalence of HIV in the general population |
\[
\log(\rho) = \log\left(\frac{t}{1-t}\right) - \log\left(\frac{h}{1-h}\right), \quad \rho \in \mathbb{R}^+, \quad (t, h) \in [0,1[.
\]
HIV+ TB incidence

• Main assumption (\(n^+\) notified HIV+):
  \[
  \frac{I^+}{I} \approx \frac{n^+}{n}
  \]

• Infer HIV prevalence among incident TB \(\left(\frac{I^+}{N}\right)\)
  – National HIV surveys measuring \(n^+/n\)
  – Routine HIV testing with high coverage
  – HIV sentinel surveillance

• Since 2013, implementation in Spectrum (Avenir, US), with disaggregation by groups of CD4 cell counts