Methods for estimating the incidence of drug-resistant TB

Anna Dean, Pete Dodd, Philippe Glaziou

WHO Global Task Force on TB Impact Measurement:
Meeting of subgroup to review WHO methods for estimating TB disease burden

11-12 May, 2022
Mövenpick Hotel, Geneva, Switzerland

Confidential, not for sharing
Some slides with provisional results for absolute numbers have been removed
History of consultations and estimates for RR-TB

See Table 1 in background doc

Global MDR-TB stakeholder meeting
- 2013
WHO STAG-TB
- 2014
WHO TF
- 2016
WHO TF
- 2018
WHO TF
- 2022

Consultations

Publication of estimates
- 2008
- 2010
- 2014-2015
- 2016-2017
- 2018-2020

Global and county-level estimates
Global estimates only
Extended from MDR to RR-TB; Country-level estimates reintroduced
Methods refined
Source of data, 1995-2020
Data inclusion

Surveillance (criteria since 2015)

• test results available for ≥80% of bacteriologically confirmed new and/or previously treated patients with pulmonary TB;
• ratio of new patients to patients with unknown treatment history is at least 4:1;
• data ≤15 years;
• no obvious data irregularities, following clarifications with national TB programmes (NTPs) during WHO’s annual round of global TB data collection.

Surveys

• data ≤15 years
In 2020, 59% of notified pulmonary TB cases are bacteriologically confirmed
Year of data, 2005-2020

2015-2020: 147 countries (15 HBCs have no data)
2006-2014: 27 countries
Progress in coverage of routine surveillance

2015

50 countries (including 3 HBCs)

2020

93 countries (including 20 HBCs)
Fluctuations in proportions with RR-TB – new cases
Fluctuations in proportions with RR-TB – previously treated cases
Number of countries with excluded surveillance data

<table>
<thead>
<tr>
<th>Year</th>
<th>New cases i</th>
<th>Previously treated cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2017</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>2018</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>2019</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2020</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

i Reasons for exclusion were:
- Concerns raised by NTP: 2 countries
- Concerns identified in expert missions: 2 countries
- Sudden unexplained changes in % tested or % with rifampicin resistance: 3 countries
- Preference for recent surveys: 5 countries
- Molecular epidemiological reasons: 1 country
\[ I_{rr} = I[(1 - f)p_n((1 - r) + r\rho) + fp_r] \]
Limitations of current approach

- Data may be old (up to 15 years); 68 countries do not have any data for the period 2015–2020
- Proportion of RR-TB among bacteriologically confirmed pulmonary TB cases assumed to be the same as among
  - undiagnosed/non-notified TB cases
  - clinically diagnosed cases
  - extra-pulmonary cases
- Data informing parameter values for estimating the incidence of RR-TB have some limitations
- Estimates can change between consecutive Global TB Reports
  - estimates use only the most recent data point
  - estimates are for a single year only
Purpose and application of RR-TB estimates

• RR-TB incidence estimates can be used for
  - indication of total burden
  - advocacy
  - trends (if using new methods)

• RR-TB incidence estimates should not be used for
  - setting targets for treatment (because non-bacteriologically confirmed incident cases cannot be detected nor treated)

• Targets for case detection and treatment should rather be based on
  - proportion with RR-TB
  - any anticipated increases in coverage of bacteriological confirmation and drug susceptibility testing over time
Estimating RR-TB proportions

Pete Dodd
p.j.dodd@sheffield.ac.uk
Evaluation criteria

- Predictive quality
- Responsiveness to local data (and transparency of response)
- Stability to fluctuations
- Computational tractability
- Face validity
- Simplicity of exposition and explanation
- Use of all data
- Range of outputs

Main emphasis

More subjective

Models including isoniazid resistance
Data cases & likelihood

Surveillance data exx. *multinomial*

Data from ≥2000 used

Survey data exx. *lognormal*
Model types

For each country $i$ & patient group $g$ in \{new, ret\}

$$f(p^g_i) = C^g_i + S^g_i t + \gamma^g_{it} \quad \text{linear with respect to time (+ space-time random effects)}$$

\[ C = X\beta + \varphi \quad \text{regression for intercept + random effects} \]

\[ S = X\alpha + \delta \quad \text{regression for slope + random effects} \]

- \( X = \text{WHO regions + Former Soviet Republics; no other explanatory variables}\)
- Either include 4 resistance types or 2
- Joint or independent random effect priors on $\varphi$ & $\delta$ for the two patient groups
- Prior for random effects $\varphi$ & $\delta$ (hierarchical structure to share information)
Model types

Prior for random effects $\varphi \ & \ \delta$:

- Independent fixed level (variance) of Gaussian ‘noise’
- Hierarchical with level of noise learned globally
- Hierarchical with level of noise learned at WHO regional level
- Areal spatial models (Leroux conditional autoregressive) 
  Gaussian noise with locally-learned, correlated noise
  Based on a specified country adjacency structure
- Some other variants that were intractable or poorly inferred
Model types

Illustration of adding space-time interaction

General contrast in degree of smoothing
Evaluating models

Models fitted by MCMC ⇒ multiple fits × multiple models = slow

1. Use ‘PSIS-LOO’ method to approximate predictive performance
   *Method approximates leave-one-out experiment with only single fit*

2. Select smaller set of models for explicit leave-one-out experiments
   *Based on results of 1 & tractable run time*

3. Leave out one country at a time or leave out year=2020 for each country
   *With pragmatic limitations on number of countries omitted*

4. Convert RR proportion estimates → RR incidence estimates
   *Face validity & additional experiments*

5. ‘Rollback’ experiments on incidence stability with new data
Approximate leave-one-out

<table>
<thead>
<tr>
<th>model</th>
<th>elpd_diff</th>
<th>se_diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>together_LerouxInterceptLerouxSlope_2</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>separate_regionall_2</td>
<td>-22.37</td>
<td>12.66</td>
</tr>
<tr>
<td>together_CARanova_2</td>
<td>-28.87</td>
<td>21.07</td>
</tr>
<tr>
<td>together_LerouxInterceptLerouxSlope_2</td>
<td>-2360.47</td>
<td>290.01</td>
</tr>
<tr>
<td>separate_LerouxInterceptLerouxSlope_2</td>
<td>-2362.20</td>
<td>288.14</td>
</tr>
<tr>
<td>separate_regional_2</td>
<td>-2363.25</td>
<td>288.45</td>
</tr>
<tr>
<td>separate_globalH_2</td>
<td>-2412.16</td>
<td>289.41</td>
</tr>
<tr>
<td>together_LerouxIntercept_2</td>
<td>-2414.95</td>
<td>291.88</td>
</tr>
<tr>
<td>separate_global_2</td>
<td>-2419.81</td>
<td>291.19</td>
</tr>
<tr>
<td>separate_LerouxIntercept_2</td>
<td>-2424.59</td>
<td>290.96</td>
</tr>
<tr>
<td>together_CARanova_2</td>
<td>-3486.79</td>
<td>373.09</td>
</tr>
<tr>
<td>separate_regional_4</td>
<td>-29442.45</td>
<td>2834.46</td>
</tr>
<tr>
<td>together_LerouxInterceptLerouxSlope_4</td>
<td>-29491.99</td>
<td>2855.27</td>
</tr>
<tr>
<td>together_LerouxIntercept_4</td>
<td>-29504.11</td>
<td>2835.36</td>
</tr>
<tr>
<td>separate_LerouxInterceptLerouxSlope_4</td>
<td>-29540.63</td>
<td>2847.64</td>
</tr>
</tbody>
</table>
Explicit leave-one-out experiments

Experiment types

- Leave one country out (all data in country)
- Leave 2020 data out in a country (one country at a time)

Refit each time and compare predictions against withheld data

- Computationally expensive ⇒ restrict countries
- 39 countries have 50% of data points
- 30 of these countries have data for 2020
Explicit leave-one-out experiments

Metrics computed to compare data & predictions:

- **H** - cross-entropy; same as ELPD; prediction/data distribution ‘overlap’
- **MAPE** - mean absolute percentage error (→relative)
- **MAE** - mean absolute error (→not relative)
- **U** - uncertainty/precision as interquartile range (IQR)
- **C** - coverage: the % of time measurement within 95% prediction interval

\[ H = -E[\log(p)] \]
Explicit leave-one-out experiments: omit country

<table>
<thead>
<tr>
<th>model</th>
<th>H</th>
<th>H.sd</th>
<th>MAPE</th>
<th>MAE</th>
<th>U</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>together_LerouxInterceptLerouxSlope_2</td>
<td>33.309</td>
<td>9.977</td>
<td>65.349</td>
<td>2.357</td>
<td>3.451</td>
<td>50.947</td>
</tr>
<tr>
<td>together_LerouxInterceptLerouxSlope_4</td>
<td>52.337</td>
<td>16.574</td>
<td>69.149</td>
<td>5.011</td>
<td>6.730</td>
<td>44.867</td>
</tr>
<tr>
<td>together_LerouxInterceptLerouxSlope_2</td>
<td>119.147</td>
<td>22.865</td>
<td>79.066</td>
<td>12.706</td>
<td>15.897</td>
<td>61.937</td>
</tr>
<tr>
<td>together_LerouxInterceptLerouxSlope_4</td>
<td>121.704</td>
<td>30.799</td>
<td>72.731</td>
<td>14.537</td>
<td>22.645</td>
<td>55.871</td>
</tr>
<tr>
<td>together_LerouxInterceptLerouxSlope_2</td>
<td>130.882</td>
<td>30.724</td>
<td>73.234</td>
<td>16.427</td>
<td>26.438</td>
<td>53.855</td>
</tr>
</tbody>
</table>
Explicit leave-one-out experiments: omit country 2020 data

<table>
<thead>
<tr>
<th>model</th>
<th>H</th>
<th>H.sd</th>
<th>MAPE</th>
<th>MAE</th>
<th>U</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>together_LerouxInterceptLerouxSlope_4</td>
<td>39.976</td>
<td>0.412</td>
<td>55.498</td>
<td>0.935</td>
<td>0.893</td>
<td>93.914</td>
</tr>
<tr>
<td>together_LerouxInterceptLerouxSlope_2</td>
<td>40.104</td>
<td>0.451</td>
<td>56.139</td>
<td>0.930</td>
<td>0.872</td>
<td>93.951</td>
</tr>
<tr>
<td>together_LerouxInterceptLerouxSlopeI_2</td>
<td>40.575</td>
<td>0.463</td>
<td>56.578</td>
<td>0.927</td>
<td>1.081</td>
<td>96.594</td>
</tr>
<tr>
<td>together_LerouxInterceptLerouxSlope_4</td>
<td>163.019</td>
<td>0.833</td>
<td>78.864</td>
<td>7.034</td>
<td>5.001</td>
<td>95.607</td>
</tr>
<tr>
<td>together_LerouxInterceptLerouxSlope_2</td>
<td>163.931</td>
<td>0.921</td>
<td>76.772</td>
<td>6.977</td>
<td>5.186</td>
<td>95.508</td>
</tr>
<tr>
<td>together_LerouxInterceptLerouxSlopeI_2</td>
<td>164.615</td>
<td>0.869</td>
<td>75.927</td>
<td>7.050</td>
<td>5.716</td>
<td>97.813</td>
</tr>
</tbody>
</table>
Incidence outputs

Convert RR proportion estimates → RR incidence estimates:

1. Use 2020 all-TB incidence estimates (for all years, for all outputs)
2. Use the formula: $I_{nc_{RR}} = I_{nc_{all}} \times [(1-f)(1-r+\varphi).P_{new} + f.P_{ret}]$
3. All country-years, but with $f$, $r$, & $\varphi$ not varying with time
4. Aggregate & propagate uncertainty, including that in RR proportions
Incidence outputs: global proportions

together_LerouxInterceptLerouxSlope_2

together_LerouxInterceptLerouxSlopel_2
Incidence outputs: regional proportions
Incidence outputs: regional proportions
“Rollback” experiments

Motivating consideration was stability of estimates between rounds:

How do estimates change if removing data from last 1, 2, 3, or 4 years?

Note: only updating estimates of proportions (refitting to each ‘rolled-back’ data); all-TB incidence estimates held fixed (therefore underestimating changes)
Summary observations

1. Many models relatively close in metrics and predictions
2. Joint modelling of new/ret random effects (‘together’) typically better
3. Leroux-intercept/Leroux-slope typically slightly better, but inferentially & conceptually more complex; less local
4. Space-time interactions increase local responsiveness, often slightly better, but add appearance of noise and computational burden
5. 4-category models also give INH & MDR estimates, but more expensive, worse accuracy, and some ‘crazy’ results (Leroux slope)
Extras
TB incidence

PAST

NEW

PRESENT

REL

FUTURE

RNR