A mathematical modelling approach to estimating TB burden in children

Prepared by:
Pete Dodd, James Seddon

Questions for discussion

1. Are there appropriate (likely sub-national) settings for validation? I.e. areas where TB is relatively generalized, but where the paediatric notification data are considered reliable or there are relevant survey data.

2. This approach does not use paediatric notification data or vital registration data – how could it be rigorously combined with other approaches (e.g. Jenkins et al.) that do inform estimates from such data?

3. Given data on local TB epidemiology, what would an appropriate notion of calibration look like for a model such as this?
Modelling approach to the paediatric TB burden

Mathematical modelling approach to estimating TB burden in children

P.J. Dodd¹ and J.A. Seddon²

1. Health Economics & Decision Science, School of Health & Related Research, University of Sheffield, UK
2. Department of Paediatric Infectious Diseases, Imperial College London, London, UK

ABSTRACT

Tuberculosis (TB) in children can be challenging to diagnose and may go unreported. Here, we use a mathematical model starting from adult data for 2013 to estimate the incidence and prevalence of *Mycobacterium tuberculosis* infection, and the incidence of TB disease in children. We perform this for 180 countries, with a total population of over 7 billion, comprising the vast majority of the global population. We find an incidence of infection of 9.0 (IQR: 6.9-11.7) million, a prevalence of latent of infection of 62.9 (IQR: 48.9-80.6) million, and an incidence of TB disease of 827,000 (IQR: 549,000-1,245,000). A model variant where the efficacy of BCG is assumed not to vary by latitude estimates a lower TB disease incidence of 593,000 (IQR: 379,000-914,000).

Introduction ................................................................. 1
Methods ........................................................................... 3
Summary of methods ......................................................... 3
The modelling logic is show in Figure 1. ............................. 4
Figure 2 Countries used in analysis and the WHO estimates of TB incidence. ............. 4
Description of data used..................................................... 5
Country linking and exclusions ........................................... 5
Infection rates .................................................................. 7
Table 1: Literature available used to inform community transmission parameters .......... 7
Figure 3: Log-normal fit to literature estimates of the transmission parameter, beta .......... 9
Table 2: Choice of transmission parameters describing the rates of infection for a given exposure ........................................................................................................... 9
Disease progression ................................................................ 9
Table 3: Choice of parameters for TB disease progression risks .................................... 10
Modelling approach to the paediatric TB burden

Protection from BCG ................................................................. 10
Table 4: Choice of parameters describing the protective effect of BCG vaccination .... 11
The effect of HIV ................................................................. 11
Table 5: Choice of parameter describing the IRR for developing TB disease given HIV infection ................................................................. 11
Numerical methods ................................................................. 12
Limitations ................................................................. 12
Summary of differences from previous work ......................................... 14
Results .................................................................. 14
Figure 4: Model estimates for all TB incidence in those aged under 15 years in all 180 countries ........................................................................................................ 15
Table 6: Model estimates for all TB incidence in those aged under 15 years in all 180 countries ........................................................................................................ 15
Figure 5: A mosaic plot of the mean proportion of the estimated burden that is pulmonary vs extrapulmonary and in younger vs older age groups .................................................................................. 16
Figure 6: Model estimates for all TB incidence in those aged under 15 years in all 180 countries, stratified by age ........................................................................................................ 17
Table 7: Model estimates of total TB incidence in children 0-4 years and 5-14 years, in all 180 countries ........................................................................................................ 17
Table 8: Estimates of the prevalence and incidence of M. tb infection in all 180 countries, by age group ........................................................................................................ 18
Figure 7: Estimated paediatric TB incidence in the countries with the highest 50 burdens ........................................................................................................ 19
Figure 8: Proportion of TB incidence in children predicted against total TB incidence for each country ........................................................................................................ 20
Figure 9: Incidence predicted against reported cases in children for 180 countries ........ 21
Funding .................................................................. 21
Abbreviations used in text .................................................................. 22
References .................................................................. 23
Introduction

TB in children (those less than 15 years of age) can be difficult to diagnose, and may be underreported even where diagnosed. Given that the gap between notifications of TB in children and the true incidence is not directly measurable, our approach is to use a mathematical model to estimate TB infection rates in children from estimates of adult TB prevalence, and then model the numbers of children that then go on to develop disease.

This report recaps the methodology in Dodd et al. (1) and its online appendix, before extending its scope from the 22 highest TB burden countries considered there, to a current global estimate. More detail and discussion can be found in (1).

Methods

Summary of methods

![Diagram of modelling logic](image)

**Figure 1: Overview of modelling logic.** Numbers at risk are fed through models of exposure and infection, and risks of progression to disease (modified by BCG vaccination and HIV) to arrive at estimates of tuberculosis incidence in children. Diamonds represent data sources, squares represent numbers estimated at each stage, and stadiums represent modelling stages.
The modelling logic is shown in Figure 1. For each country considered, data on demography is used to determine numbers in different age groups who are at risk of infection; data on TB prevalence is used to model the risk of exposure and infection. Data from the literature are used to inform an age-dependent model of progression to extrapulmonary and pulmonary disease, and the rates of progression are modified by BCG vaccination, at the coverage relevant to a given country. Every input is treated as uncertain to a quantified extent, and this uncertainty is propagated through into all outputs.

**TB incidence per 100k**

Figure 2 Countries used in analysis and the WHO estimates of TB incidence. Countries in white were excluded as described in text.
Description of data used

We obtained data on country demography for 2013 from UN ESA, Population Division (2). Where necessary, 5-year age categories were disaggregated under the assumption of uniformly distributed ages. These data were used to generate the number of children at risk in each country by age.

WHO estimates of adult TB prevalence were obtained from for 2013, together with 95% uncertainty bounds (from (3)). Uncertainty in per-capita prevalence was represented by gamma distributions, parameterised by taking the quoted ranges defined by the upper and lower bounds as $1.96 \times$ the standard deviation, and the quoted point estimate as the mean. WHO notification data from 2010 were used to estimate the proportion of incident TB that is smear positive for the community ARI estimate. The same estimate was used for all countries to avoid bias resulting from different case detection infrastructures etc.

BCG vaccination coverage estimates were obtained for 2013 from WHO (4). The BCG vaccination coverages were used to determine the fraction of children whose risks of progression from infection to disease were moderated by BCG (in the manner described in the relevant section below).

HIV prevalence estimates in those aged under 15 were available for 82 countries from (5), together with 95% uncertainty bounds. Countries for which there were not estimates reported from this source were assumed to have negligible HIV prevalence in those under 15 years of age. Uncertainty in the prevalences was represented by gamma distributions, parameterised by taking the quoted ranges defined by the upper and lower bounds were taken as $1.96 \times$ the standard deviation, and the quoted point estimate as the mean. This HIV prevalence was assumed to be uniform by age in those under 15. Degree of immunosuppression or ART was not considered.

Country linking and exclusions

The WHO TB estimate and notification data were linked with the demographic, HIV, and BCG sets by 3 letter ISO code where possible, and by hand otherwise. Various countries were excluded where it was not possible to link them across the data. The WHO version of country names was used.
The following countries were dropped as not appearing in the WHO TB data:
American Samoa; Andorra; Anguilla; Bermuda; Bonaire, Saint Eustatius and Saba; British Virgin Islands; Brunei Darussalam; Cayman Islands; Cook Islands; Dominica; Greenland; Marshall Islands; Monaco; Montserrat; Nauru; Netherlands Antilles; Niue; Northern Mariana Islands; Palau; Saint Kitts and Nevis; San Marino; Serbia & Montenegro; Sint Maarten (Dutch part); Solomon Islands; Tokelau; Turks and Caicos Islands; Tuvalu; Wallis and Futuna Islands.

The following countries were dropped as not present in the BCG data:
Aruba; China, Hong Kong SAR; China, Macao SAR; Curacao; French Polynesia; Guam; New Caledonia; Puerto Rico; US Virgin Islands; West Bank and Gaza Strip.

The following countries were dropped from the BCG data as not present in the demographic data:
Channel Islands; French Guiana; Guadeloupe; Martinique; Mayotte; Other non-specifed areas; Réunion; Southern Asia; Western Sahara.

The following country was dropped as in the BCG data but not the WHO TB data:
Marshall Islands (the).

The following countries were dropped from the HIV dataset:
Moldova.

The following country was dropped as not having TB data for relevant year:
Tajikistan.

The following countries were included:
Afghanistan; Albania; Algeria; Angola; Antigua and Barbuda; Argentina; Armenia; Australia; Austria; Azerbaijan; Bahamas; Bahrain; Bangladesh; Barbados; Belarus; Belgium; Belize; Benin; Bhutan; Bolivia (Plurinational State of); Bosnia and Herzegovina; Botswana; Brazil; Bulgaria; Burkina Faso; Burundi; Cabo Verde; Cambodia; Cameroon; Canada; Central African Republic; Chad; Chile; China; Colombia; Comoros; Congo; Costa Rica; Croatia; Cuba; Cyprus; Czech Republic; Côte d'Ivoire; Democratic People's Republic of Korea; Democratic Republic of the Congo; Denmark; Djibouti; Dominican Republic; Ecuador; Egypt; El Salvador; Equatorial Guinea; Eritrea; Estonia; Ethiopia; Fiji; Finland; France; Gabon; Gambia; Georgia; Germany; Ghana; Greece; Grenada; Guatemala; Guinea; Guinea-Bissau; Guyana; Haiti; Honduras; Hungary; Iceland; India; Indonesia; Iran (Islamic Republic of); Iraq; Ireland; Israel; Italy; Jamaica; Japan; Jordan; Kazakhstan; Kenya; Kiribati; Kuwait; Kyrgyzstan; Lao People's Democratic Republic; Latvia; Lebanon; Lesotho; Liberia; Libya; Lithuania; Luxembourg; Madagascar; Malawi; Malaysia; Maldives; Mali; Malta; Mauritania; Mauritius; Mexico; Micronesia (Federated States of); Mongolia; Montenegro; Morocco; Mozambique; Myanmar; Namibia; Nepal; Netherlands; New Zealand; Nicaragua; Niger; Nigeria; Norway; Oman; Pakistan; Panama; Papua New Guinea; Paraguay; Peru; Philippines; Poland; Portugal; Qatar; Republic of Korea; Republic of Moldova; Romania; Russian Federation; Rwanda; Saint Lucia; Saint Vincent and the Grenadines; Samoa; Sao Tome and Principe; Saudi Arabia; Senegal; Serbia; Seychelles; Sierra Leone; Singapore; Slovakia; Slovenia; Somalia; South Africa; South Sudan; Spain; Sri Lanka; Sudan; Suriname; Swaziland; Sweden; Switzerland; Syrian Arab Republic; Thailand; The Former Yugoslav Republic of Macedonia; Timor-Leste; Togo; Tonga; Trinidad and Tobago; Tunisia; Turkey; Turkmenistan; Uganda; Ukraine; United Arab Emirates; United Kingdom of Great Britain and Northern Ireland; United Republic of Tanzania; United States of America; Uruguay; Uzbekistan; Vanuatu; Venezuela (Bolivarian Republic of); Viet Nam; Yemen; Zambia; Zimbabwe.
Modelling approach to the paediatric TB burden

This meant that a total of 180 countries, with a combined population of 7,057,102 thousand could be included in our analysis. These countries are mapped, together with their estimated TB incidence in Figure 2.

Infection rates

Karel Styblo first empirically estimated the coefficient between the prevalence of smear-positive adult TB and the annual risk of infection, $\beta$ (6). Using European data on ARIs from TST surveys combined with prevalence surveys, he inferred from a rate of 10 infections per year for a typical smear positive TB case. This quantity has been revisited since (7–9), as lower average case severity from improved case-detection, the rise of HIV-related tuberculosis, and social changes may all have led to departures from Styblo’s estimate; indeed, more recent estimates tend to be lower. It should be noted that from a modelling perspective, use of this community relationship between smear positive TB and the ARI does not imply that infections caused by smear negative TB have been altogether neglected. Rather, it can be thought of as modelling the infections due to smear positive TB cases plus the product of the number of smear negative cases per smear positive case and the relative infectiousness of a smear negative case, i.e. as inflated to indirectly include the smear negative contribution to transmission.

We used data from the reviews of Bourdin Trunz and colleagues (9) and van Leth and colleagues (8) (see Table 1) and fitted a log-normal distribution to these estimates of $\beta$ (see Figure 3 and Table 2), which was used to model the values of this coefficient in the community ARI approach to estimating infection incidence.

Table 1: Literature available used to inform community transmission parameters.

<table>
<thead>
<tr>
<th>Country</th>
<th>year</th>
<th>ARI</th>
<th>year</th>
<th>smr+ TB prevalence</th>
<th>$\beta$</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cambodia</td>
<td>2002</td>
<td>2.06</td>
<td>2002$^i$</td>
<td>269</td>
<td>7.8</td>
<td>T21</td>
</tr>
<tr>
<td>China</td>
<td>1979</td>
<td>1.1</td>
<td>1979$^j$</td>
<td>187</td>
<td>5.7</td>
<td>T22</td>
</tr>
<tr>
<td>China</td>
<td>1990</td>
<td>1.0</td>
<td>1990$^j$</td>
<td>134</td>
<td>7.2</td>
<td>T22</td>
</tr>
<tr>
<td>China</td>
<td>2000</td>
<td>0.7</td>
<td>2000$^j$</td>
<td>110</td>
<td>6.5</td>
<td>T23</td>
</tr>
<tr>
<td>The Philippines</td>
<td>1981-83</td>
<td>2.5</td>
<td>1981-83$^j$</td>
<td>660</td>
<td>3.8</td>
<td>T25/5</td>
</tr>
<tr>
<td>The Philippines</td>
<td>1997</td>
<td>2.3</td>
<td>1997$^j$</td>
<td>310</td>
<td>7.4</td>
<td>T24/5</td>
</tr>
<tr>
<td>South Korea</td>
<td>1965</td>
<td>5.3</td>
<td>1965$^j$</td>
<td>668</td>
<td>7.9</td>
<td>T26</td>
</tr>
</tbody>
</table>
### Modelling approach to the paediatric TB burden

<table>
<thead>
<tr>
<th>Country</th>
<th>Year</th>
<th>Survey</th>
<th>T26 Estimate</th>
<th>T26 Value</th>
<th>T27 Estimate</th>
<th>T27 Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>South Korea</td>
<td>1970</td>
<td>3.9</td>
<td>1970(^1)</td>
<td>559</td>
<td>7.0</td>
<td>T26</td>
</tr>
<tr>
<td>South Korea</td>
<td>1975</td>
<td>2.3</td>
<td>1975(^1)</td>
<td>480</td>
<td>4.8</td>
<td>T26</td>
</tr>
<tr>
<td>South Korea</td>
<td>1980</td>
<td>1.8</td>
<td>1980(^1)</td>
<td>309</td>
<td>5.8</td>
<td>T26</td>
</tr>
<tr>
<td>South Korea</td>
<td>1985</td>
<td>1.2</td>
<td>1985(^1)</td>
<td>239</td>
<td>5.0</td>
<td>T26</td>
</tr>
<tr>
<td>South Korea</td>
<td>1990</td>
<td>1.1</td>
<td>1990(^1)</td>
<td>143</td>
<td>7.7</td>
<td>T26</td>
</tr>
<tr>
<td>South Korea</td>
<td>1995</td>
<td>0.5</td>
<td>1995(^1)</td>
<td>93</td>
<td>5.4</td>
<td>T26</td>
</tr>
<tr>
<td>Kenya</td>
<td>1994-96</td>
<td>1.1</td>
<td>1995(^2)</td>
<td>132</td>
<td>8.3</td>
<td>T27</td>
</tr>
<tr>
<td>Tanzania</td>
<td>1993-98</td>
<td>0.9</td>
<td>1996(^1)</td>
<td>172</td>
<td>5.2</td>
<td>T28</td>
</tr>
<tr>
<td>Tanzania</td>
<td>2000</td>
<td>0.68</td>
<td>2000(^2)</td>
<td>214</td>
<td>3.1</td>
<td>T29</td>
</tr>
<tr>
<td>Malawi</td>
<td>1994</td>
<td>1.0</td>
<td>1995(^2)</td>
<td>248</td>
<td>4.0</td>
<td>T30</td>
</tr>
<tr>
<td>Madagascar</td>
<td>1991-94</td>
<td>1.2</td>
<td>1995(^2)</td>
<td>145</td>
<td>8.4</td>
<td>T31</td>
</tr>
<tr>
<td>Egypt</td>
<td>1995-97</td>
<td>0.3</td>
<td>1996(^2)</td>
<td>24</td>
<td>13.2</td>
<td>T32</td>
</tr>
<tr>
<td>Laos</td>
<td>1996</td>
<td>1.1</td>
<td>1996(^2)</td>
<td>165</td>
<td>6.7</td>
<td>T33</td>
</tr>
<tr>
<td>India</td>
<td>2000-03</td>
<td>1.5</td>
<td>2002(^2)</td>
<td>155</td>
<td>9.6</td>
<td>T34</td>
</tr>
<tr>
<td>South Korea</td>
<td>1975</td>
<td>1.78</td>
<td>1975(^1)</td>
<td>480</td>
<td>3.7</td>
<td>V7/8</td>
</tr>
<tr>
<td>South Korea</td>
<td>1980</td>
<td>1.24</td>
<td>1980(^1)</td>
<td>310</td>
<td>4.0</td>
<td>V7/8</td>
</tr>
<tr>
<td>South Korea</td>
<td>1985</td>
<td>1.12</td>
<td>1985(^1)</td>
<td>240</td>
<td>4.7</td>
<td>V7/8</td>
</tr>
<tr>
<td>South Korea</td>
<td>1990</td>
<td>0.6</td>
<td>1990(^1)</td>
<td>144</td>
<td>3.2</td>
<td>V7/8</td>
</tr>
<tr>
<td>China</td>
<td>1979</td>
<td>0.64</td>
<td>1979(^1)</td>
<td>187</td>
<td>3.4</td>
<td>V9</td>
</tr>
<tr>
<td>China</td>
<td>1985</td>
<td>0.59</td>
<td>1985(^1)</td>
<td>156</td>
<td>3.8</td>
<td>V9</td>
</tr>
<tr>
<td>China</td>
<td>1990</td>
<td>0.64</td>
<td>1990(^1)</td>
<td>134</td>
<td>4.8</td>
<td>V10</td>
</tr>
<tr>
<td>China</td>
<td>1996</td>
<td>0.7</td>
<td>1996(^1)</td>
<td>121</td>
<td>5.8</td>
<td>V10</td>
</tr>
<tr>
<td>The Philippines</td>
<td>1982</td>
<td>2.48</td>
<td>1982(^1)</td>
<td>950</td>
<td>2.6</td>
<td>V11</td>
</tr>
<tr>
<td>The Philippines</td>
<td>1994</td>
<td>2.30</td>
<td>1994(^1)</td>
<td>525</td>
<td>4.4</td>
<td>V11</td>
</tr>
</tbody>
</table>

1. Survey
2. WHO estimate

Tn = Bourdin Trunz and colleagues (9) - ref n; Vn = van Leth and colleagues (8) - ref n;
T21-T32 = (10), (11), (12), (13), (14), (15), (16), (17), (18), (19), (20), (21), (22);
V7-V11 = (23), (15), (11), (12), (13)
Modelling approach to the paediatric TB burden

Figure 3: Log-normal fit to literature estimates of the transmission parameter, beta.

Table 2: Choice of transmission parameters describing the rates of infection for a given exposure, modelled as log-normal distributions.

<table>
<thead>
<tr>
<th>quantity</th>
<th>mu</th>
<th>sigma</th>
<th>median</th>
<th>LQ</th>
<th>UQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>community $\beta(y^{-1})$</td>
<td>1.678</td>
<td>0.371</td>
<td>5.354</td>
<td>4.168</td>
<td>6.879</td>
</tr>
</tbody>
</table>

Disease progression

Disease progression risks were based on the review of the pre-chemotherapy literature by Marais and colleagues (24). BCG vaccination had not been used in the populations reviewed in this study. We take these risks as approximating the risk of progression within a year. The risks of developing TB disease in (24) were used to parameterise the mean of beta distributions for each age category, with the variance taken as 0.125 x mean (equivalent to taking a range of plausible variation as +/-25% of the mean, and the variance as a quarter of the plausible range). The risk of extrapulmonary disease (miliary TB or TB meningitis) was modelled as a random proportion of
the risk of any disease, modelled by a beta distribution with mean equal to the range midpoints in (24) and the variance taken as a quarter of these ranges. Where no ranges were given, +/-25% of the mean was used as the range. Where the probability was expressed as <x, x/2 was used as the mean. Modelling the two types of disease in this way ensured that the random total probability for any kind of disease never exceeded unity. These parameters and their associated modes, lower- and upper-quartiles are presented in Table 3.

Table 3: Choice of parameters for TB disease progression risks, modelled as beta distributions.

<table>
<thead>
<tr>
<th>quantity</th>
<th>alpha</th>
<th>beta</th>
<th>median</th>
<th>LQ</th>
<th>UQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>probability TB disease, age [0,1) years</td>
<td>1.500</td>
<td>1.500</td>
<td>0.500</td>
<td>0.298</td>
<td>0.702</td>
</tr>
<tr>
<td>probability TB disease, age [1,2) years</td>
<td>1.250</td>
<td>3.750</td>
<td>0.250</td>
<td>0.108</td>
<td>0.360</td>
</tr>
<tr>
<td>probability TB disease, age [2,5) years</td>
<td>0.330</td>
<td>6.270</td>
<td>0.050</td>
<td>0.002</td>
<td>0.064</td>
</tr>
<tr>
<td>probability TB disease, age [5,10) years</td>
<td>0.137</td>
<td>6.703</td>
<td>0.020</td>
<td>0.000</td>
<td>0.013</td>
</tr>
<tr>
<td>probability TB disease, age [10,15) years</td>
<td>0.870</td>
<td>4.930</td>
<td>0.150</td>
<td>0.043</td>
<td>0.219</td>
</tr>
<tr>
<td>proportion TB EP, age [0,1) years</td>
<td>0.960</td>
<td>2.240</td>
<td>0.300</td>
<td>0.112</td>
<td>0.451</td>
</tr>
<tr>
<td>proportion TB EP, age [1,2) years</td>
<td>0.712</td>
<td>1.322</td>
<td>0.350</td>
<td>0.107</td>
<td>0.557</td>
</tr>
<tr>
<td>proportion TB EP, age [2,5) years</td>
<td>0.620</td>
<td>5.580</td>
<td>0.100</td>
<td>0.017</td>
<td>0.145</td>
</tr>
<tr>
<td>proportion TB EP, age [5,10) years</td>
<td>0.750</td>
<td>5.250</td>
<td>0.125</td>
<td>0.029</td>
<td>0.183</td>
</tr>
<tr>
<td>proportion TB EP, age [10,15) years</td>
<td>0.112</td>
<td>6.615</td>
<td>0.017</td>
<td>0.000</td>
<td>0.008</td>
</tr>
</tbody>
</table>

**Protection from BCG**

The reported efficacy of BCG in protecting against different forms of TB disease varies substantially. Rodrigues and colleagues (25) conducted a meta-analysis of protection against TB meningitis and miliary TB, finding a protection of 86% [65%,95%] from RCTs and 75% [61%,84%] from case-control studies. Colditz and colleagues (26) meta-analysis of BCGs protection for infants found protections against all TB of 74% [62%,83%] from RCTs and 52% [38%,64%] from case-control studies. They found a protection of against TB meningitis of 64% [12%,86%] and a protection of 78% [58%,88%] against disseminated TB. Bourdin Trunz and colleagues (9) updated these meta-analyses to arrive at protections of 73% for TB meningitis and 77% for miliary TB. We choose to model the protection against extra-pulmonary TB by a hazard
ratio (i.e. 1-protection) following a beta distribution, with parameters presented in Table 4. We chose to model the protection against PTB as a certain fraction of the protection against EPTB, with this fraction beta distributed with parameters given in Table 4. Taken together, this approach gave a protection against PTB of median [LQ-UQ] = 54% [38%-69%] and a protection against EPTB of median [LQ-UQ] = 70% [52%-84%].

There has been suggestion (e.g. (27,28)) that the observed protective effect of BCG varies with latitude, as a result of differences in the prevalence of non-tuberculous mycobacteria in the environment. We carried out separate analysis as a sensitivity analysis where we allowed 41% (28) of the protection from BCG to vary with latitude, so that the protection at the equator was only 59% of that at the poles. The latitude of each country’s capital was used.

**Table 4: Choice of parameters describing the protective effect of BCG vaccination, modelled as beta distributions.**

<table>
<thead>
<tr>
<th>quantity</th>
<th>alpha</th>
<th>beta</th>
<th>median</th>
<th>LQ</th>
<th>UQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>hazard ratio for EPTB, all ages</td>
<td>1.250</td>
<td>2.500</td>
<td>0.301</td>
<td>0.155</td>
<td>0.484</td>
</tr>
<tr>
<td>fractional protection for PTB, all ages</td>
<td>4.000</td>
<td>1.000</td>
<td>0.841</td>
<td>0.707</td>
<td>0.931</td>
</tr>
</tbody>
</table>

**The effect of HIV**

For each country, the proportion of children who were HIV-infected had their TB incidence multiplied by an incidence rate ratio (IRR). There is limited evidence around the effect of HIV infection in children on the rate of progression to TB disease following infection with *Mycobacterium tuberculosis*. Seddon and colleagues (29) reported an HIV prevalence of 29% among paediatric TB cases diagnosed in Cape Town during 2007-2009. Taken with the UNAIDS estimate of 3% HIV prevalence for those under 15 years of age in South Africa, this gives an IRR of 13, suggesting that children with HIV are 13 times more likely to develop TB than those HIV negative. Madhi and colleagues (30) estimated a higher incidence rate ratio (RR, 22.5; 95% CI, 13.4–37.6) among children hospitalised with TB in Johannesburg, South Africa between 1996-1997 as did Hesseling and colleagues (31) (RR 24.2 95% CI, 17–34) among infants aged up to 12 months in the Western Cape of South Africa between 2004-2006. We modelled the IRR with a log-normal distribution, described in Table 5.
Table 5: Choice of parameter describing the IRR for developing TB disease given HIV infection, modelled as a log-normal distribution.

<table>
<thead>
<tr>
<th>quantity</th>
<th>mu</th>
<th>sigma</th>
<th>median</th>
<th>LQ</th>
<th>UQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRR for TB given HIV infection, all ages</td>
<td>2.996</td>
<td>1.000</td>
<td>20.000</td>
<td>10.118</td>
<td>39.261</td>
</tr>
</tbody>
</table>

**Numerical methods**

The main analyses were based on sample sizes of 10,000 runs, using a Latin hypercube sampling scheme to improve coverage of the parameter space. Gaussian kernel density estimation was used to estimate distributions, unless otherwise stated. All analyses were carried out using the R environment for statistical computing (32).

**Limitations**

This section details the limitations of the model, some of the assumptions that have been made and some of the omissions.

1. The model of infection risk is based on homogeneous mixing assumptions that are likely to apply best to countries where TB is common in the general population, rather than countries where TB is most often found in special groups.
2. We have neglected geographical (and other) heterogeneities within countries; in fact, TB is frequently very heterogeneous at this scale. Similarly, we have looked at TB only in the general population and not in special populations such as prisons, mines etc. Although these populations do not usually include children, adults from these populations are very high risk and when returning home are likely to contribute to childhood infections.
3. The model uses as a starting point the estimated TB prevalence data for adults with TB. These estimates are themselves based on a simple model of the duration of TB disease; biases in this would reflect in our results.
4. It is assumed that the proportions of adults with smear positive and non-smear positive disease in the notification data is representative of these proportions in the prevalence data.
5. We assume the same distribution for the transmission parameter $\beta$ in all countries. This parameter may differ systematically and be a contributing factor to variation in country TB burden.
6. We have assumed that children are only infected by adults and not by other children. Although it is likely that the majority of infections do result from adults, children, and especially older children, can certainly be infectious.

7. We have not included elements of the degree of smear positivity - certainly the more positive the sputum the more likely children are to become infected. Smear status has been considered binary - smear positive or smear negative.

8. We have not included any measure of the effects of drug resistance on whether children are more or less likely to be exposed, become infected or develop disease.

9. In common with almost all infectious disease modelling, we have assumed a consistent and linear relationship between adult TB prevalence and ARI. In reality this relationship may be non-linear and variable.

10. The model ignores anergic children. We have used a positive test of infection as a surrogate for TB infection and assumed that it is a necessary stage to pass through towards disease. However, many children do not mount response to TST or IGRA and in cohorts of children with TB disease a significant proportion of children have negative tests of infection.

11. We have looked at HIV status in the child as a binary issue (HIV or no HIV). The impact of degree of immunosuppression and the use of ART has not been included. HIV prevalence in children has been assumed not to vary with age.

12. BCG has been challenging to model. The literature surrounding its effects is hugely variable, with a protective effect described as ranging from 0 to 70%. The implications of geographical location are also unclear. However, we have included a summary protective effect for both pulmonary and nonpulmonary disease and also modelled the effect of the latitude, based on the best available data. We have not included any element of BCG strain used, the mode of delivery, efficacy of cold chain etc. We also neglected the possibility that BCG vaccination protects against infection, as distinct from disease.

13. When modelling the risk of disease progression following infection we have not included factors such as *M. tuberculosis* strain type, host genetics, host nutritional status, exposure to sunlight (and vitamin D levels) or any other factors other than age, HIV status and BCG vaccination history. It is clear that factors other than these can play a part in risk of disease progression.

14. We have assumed that risk of infection and risk of disease progression once infected are unrelated. However, in reality, it is likely that the degree of exposure (and likely bacillary load exposed to), will affect both the risk of infection as well as the risk of disease progression once infected.

15. From an infectious disease modelling perspective we have assumed that all children at the beginning of their exposure were non-immune and had not been infected before. We have assumed that the risk of disease progression following infection is the same for a child who has recently been infected for the first time as it is for a child who has recently been infected but who has also been previously infected.
16. Literature risks of progression refer to a life-long time-frame, whereas we have interpreted these risks as operating on a timescale shorter than the width of our age categories. This is necessary with the underlying model structure to allow for age-dependent risks of progression, since we do not represent time-since-infection. However, where our age-categories are narrowest, the progression is also likely to be most rapid, which ameliorates this approximation.

17. The model does not include IPT. While IPT significantly reduces the likelihood of disease progression following infection, in the high burden countries child contacts are rarely identified and IPT is rarely provided.

**Summary of differences from previous work**

The methodology in this paper is very much that described in Dodd et al. (1) and its online appendix. In this section, we briefly highlight the differences.

The largest difference is the set of countries to which the method is applied. Here, we apply the model to a set of 180 countries, using data from 2013; whereas in Dodd et al. (1), we considered only the 22 highest burden TB countries (HBCs), using data largely from 2010.

We only consider the ‘community’ model of infection in this document, as data to inform the household method were not available for a large enough number of countries. We also shifted from using the latitude of a country’s capital to the latitude of a country’s centroid in the model variant with latitude variation in BCG efficacy.

**Results**

In this section, we present model estimates of the absolute TB incidence aggregated over extra-pulmonary and pulmonary TB disease, for younger and older age groups ([0,5) and [5-15), respectively). Results are also presented from the latitude-independent (labelled ‘nolat’) and latitude-dependent (labelled ‘lat’) approaches to modelling the protection from BCG vaccination.
Figure 4: Model estimates for all TB incidence in those aged under 15 years in all 180 countries; both model variants shown. The lines represent a Gaussian kernel density estimate.

In Figure 4 and Table 6 we compare the results from the different model variants aggregated over all age groups and all 180 countries. Modes were somewhat lower than corresponding medians. The latitude-dependent model variants of BCG protection consistently estimated higher numbers than the latitude-independent model variants.
Table 6: Model estimates for all TB incidence in those aged under 15 years in all 180 countries; both model variants shown. (LQ=lower quartile; UQ=upper quartile.)

<table>
<thead>
<tr>
<th>LAT</th>
<th>median</th>
<th>LQ</th>
<th>UQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>lat</td>
<td>827,427</td>
<td>548,562</td>
<td>1,244,865</td>
</tr>
<tr>
<td>nolat</td>
<td>592,770</td>
<td>379,301</td>
<td>914,330</td>
</tr>
</tbody>
</table>

In Figure 5 we present a visualisation of the mean proportions of the total due to ages 0-4 and 5-14, split by type of disease. The mean estimated incidence in ages 0-4 is similar to that in ages 5-14 (evidenced by similar bar widths). Extrapulmonary TB is more common in the younger age groups (approximately 25% of incident cases), and makes up approximately 10% of the total incidence in both age groups. The estimated summaries of model incidence in each age-group across all 180 countries are presented in Table 7.

Figure 5: A mosaic plot of the mean proportion of the estimated burden that is pulmonary vs extrapulmonary (here denoted PTB and DTB, respectively), and in younger vs older age groups (0-4 and 5-14). This is a diagrammatic representation of a cross-tabulation: both the widths and heights are meaningful.
Figure 6: Model estimates for all TB incidence in those aged under 15 years in all 180 countries, stratified by age; both model variants shown. The lines represent a Gaussian kernel density estimate.

Table 7: Model estimates of total TB incidence in children 0-4 years and 5-14 years, in all 180 countries for the BCG model variant with no variation of efficacy by latitude. (LQ = lower quartile; UQ = upper quartile.)

<table>
<thead>
<tr>
<th>age group</th>
<th>LAT</th>
<th>median</th>
<th>LQ</th>
<th>UQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>lat</td>
<td>421,400</td>
<td>268,531</td>
<td>630,061</td>
</tr>
<tr>
<td>0-4</td>
<td>nolat</td>
<td>290,443</td>
<td>176,550</td>
<td>455,100</td>
</tr>
<tr>
<td>5-14</td>
<td>lat</td>
<td>354,536</td>
<td>171,307</td>
<td>661,682</td>
</tr>
<tr>
<td>5-14</td>
<td>nolat</td>
<td>255,034</td>
<td>120,651</td>
<td>489,440</td>
</tr>
</tbody>
</table>
We previously found that, for the 22 HBCs, a small proportion of incident paediatric TB cases were infected with HIV (median [LQ-UQ] = 5.0% [2.4%-10.1%]), which is lower than the corresponding statistic for adults (13%, (33)). However, this is likely to vary substantially by country: the proportion of co-infected children was predicted to be over 30% in South Africa and Zimbabwe.

Our approach also allows the estimation of the incidence and prevalence of *M.tb* infection (as opposed to disease – see Table 8), which has relevance, e.g., to preventive therapy interventions. As these quantities do not depend on the natural history of progression, they do not depend on the model variant used, and are proportionately more precise.

**Table 8: Estimates of the prevalence and incidence of *M.tb* infection in all 180 countries, by age group for the ‘lat’ model variant.** (LQ=lower quartile, UQ=upper quartile)

<table>
<thead>
<tr>
<th><em>M.tb.</em> infection</th>
<th>age group</th>
<th>median</th>
<th>LQ</th>
<th>UQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>prevalence</td>
<td>0-14</td>
<td>62,896,959</td>
<td>48,897,936</td>
<td>80,649,253</td>
</tr>
<tr>
<td>incidence</td>
<td>0-14</td>
<td>8,997,410</td>
<td>6,936,881</td>
<td>11,696,688</td>
</tr>
<tr>
<td>prevalence</td>
<td>0-4</td>
<td>7,861,145</td>
<td>6,080,713</td>
<td>10,173,695</td>
</tr>
<tr>
<td>incidence</td>
<td>0-4</td>
<td>3,191,801</td>
<td>2,459,671</td>
<td>4,145,840</td>
</tr>
<tr>
<td>prevalence</td>
<td>5-14</td>
<td>55,063,579</td>
<td>42,804,117</td>
<td>70,500,851</td>
</tr>
<tr>
<td>incidence</td>
<td>5-14</td>
<td>5,809,956</td>
<td>4,477,617</td>
<td>7,553,058</td>
</tr>
</tbody>
</table>

The countries with the highest overall burdens of paediatric TB include very large countries with relatively low per capita TB rates (e.g. China) and relatively smaller countries with higher per capita TB rates (e.g. South Africa). Figure 7 shows the 50 countries with the highest overall paediatric TB incidences, and is dominated by India, which has 18% [IQR: 14-22%] of the total burden. Most of these countries have high BCG vaccination coverage, with Pakistan, Ethiopia, Nigeria and Somalia having relatively lower coverage.
The proportion of TB burden occurring in children is expected to correlate positively with overall TB incidence, both because of the ecological association between high TB incidence and young populations, and because higher incidence leads to earlier average infection with higher chances of progression (34). The proportion of TB burden occurring in children is also a key quantity in other methods to paediatric TB burden estimation (35). In Figure 8 we show the proportion of TB incidence predicted to occur in children by the ‘lat’ model variant for each country.
country against the overall TB incidence that year. Panel B shows this data on a logit-log scale and a trend line corresponding to a summary rule:

\[
\frac{p}{1-p} = \frac{1}{50 \times I^{1/4}}
\]

where \( I \) is the TB incidence per 100,000 per year, and \( p \) is the proportion of TB incidence occurring in children. In (1), we note that countries with high HIV incidence tend to follow a different pattern as HIV associated TB has a different relationship between incidence and the prevalence that drives infection.

Figure 8: Proportion of TB incidence in children predicted by the ‘lat’ model variant against total TB incidence for each country. Panel A is on a linear scale; panel B has a logit scale for the y-axis and log-scale for the x-axis and shows a linear model trend line corresponding to the equation on the plot.

In Figure 9 we compare the TB incidence predicted in children by the ‘lat’ model variant with the number of cases reported by each country. This figure is on a log-log scale and the preponderance of countries under the diagonal line therefore represents a considerable gap between notifications and incidence. Some countries do report numbers that are larger than those predicted by the model however. While the model is expected to be more appropriate in settings with generalized TB epidemics, it is interesting to note that the US and the UK (countries where the notifications are thought to be relatively good estimates of the true incidence) lie close to and slightly above the diagonal. The ratio of total notifications to predicted incidence was 37%.
Figure 9: Incidence predicted by ‘lat’ model variant against reported cases in children for 180 countries. Predicted incidence and notifications are both on log-scales. The dashed line indicates equality between reported and estimated cases; below the line notifications are lower than predicted incidence. The blue color scale depicts country GDP in USD for 2012 (from the World Bank) with a square-root transformation. Red error-bars depict inter-quartile ranges. Data for the US and UK are highlighted with green text and arrows.

**Funding**

This work was funded by a grant from UNITAID and the US Agency for International Development to the Global Alliance for TB Drug Development.
Abbreviations used in text

ARI  annual risk of infection
ART  antiretroviral therapy
BCG  Bacille Calmette Guérin
EPTB extrapulmonary tuberculosis
HBC  high burden country
HIV  human immunodeficiency virus
IGRA interferon gamma release assay
IPT  isoniazid preventive therapy
IQR  interquartile range
IRR  incidence risk ratio
LQ  lower quartile
PTB  pulmonary tuberculosis
M.tb  Mycobacterium tuberculosis
TB  tuberculosis
TST  tuberculin skin test
UN ESA United Nations Economic and Social Affairs
UQ  upper quartile
WHO World Health Organisation
References


Modelling approach to the paediatric TB burden


