Questions to the Task Force

1. What are your general comments on the challenges to estimating the burden of TB by age and sex?
2. Does the Task Force agree with the proposed next steps? Is there anything important that is missing or should be deleted?
Background Paper No. 9

Estimating the burden of TB by age and sex: availability of data, gaps and next steps

Background
Each year WHO produces estimates of TB incidence, prevalence and mortality for each country. These estimates are presented in the Stop TB Department's annual report. However, as age- and gender-stratified data are limited, age- and gender-stratified estimates have never been produced. The development of new methods for this analysis must take into account the quality and completeness of the age/sex-disaggregated National TB control Programme (NTP) notification data available at WHO-HQ.

Gender
A resolution of the Sixtieth World Health Assembly in 2007 asked WHO to "give priority to the use of sex-disaggregated data and gender analysis in WHO’s publications, in an effort to strengthen health-information systems in order to ensure that they reflect awareness of gender equality as a determinant of health". However, it is often difficult and burdensome for countries to report these data, and the quality of the data received is of questionable reliability.

The data that are available suggest that, although globally men account for a higher proportion of notified TB cases (around 60-65%), data on sex differentials in some regions and particular countries show very unusual patterns which need to be further explored and explained. In most countries, there is no significant difference in TB between sexes in childhood and pre-adolescence. However, during adolescence rates begin to diverge and there is a much higher incidence in men than in women during adulthood. The relative risks reduce again in the older age groups. The reasons for these differences are likely to result from various factors, including access to care, ethnicity, the influence of the HIV co-epidemic, as well as other biological, social and cultural variables. Whether, and the extent to which, identifying fewer women with TB globally is due to sex (as a biological determinant) or gender (as a socio-cultural determinant reflecting barriers women may face in accessing TB care) have been issues for discussion and debate. The analysis of the trends in age/sex-disaggregated notification data, coupled with the analysis of changes in their completeness and quality, may be particularly significant for the understanding of the overall determinants of TB disease.

Age
The childhood TB subgroup of the Stop TB Partnership’s DOTS expansion working group has requested that the burden of TB in children be properly estimated. This is made more difficult by the fact that childhood TB is a difficult to diagnose and poorly reported disease. What we do see in general is a differential risk of developing TB by age. It is common to see a J-shaped curve of TB incidence rates by age, with higher rates in younger children from infancy to pre-adolescence. Rates increase abruptly during the
adolescent years and remain high throughout adulthood, with a tendency to increase as age progresses, which is particularly notable for countries with low incidence rates. Case notifications of childhood TB, for example, depend on several factors. These include the magnitude of the epidemic; higher incidence leads to higher transmission which, in turn, leads to a higher proportion of TB in children than would be seen where incidence is low and disease primarily due to reactivation of infection acquired in the past.

The age structure of the population, the availability of diagnostic tools, the extent of routine contact tracing, and TB case definitions and recording and reporting norms and practices are also very influential. Significant under-diagnosis, misdiagnosis and under-notifications may be due to the following:

- Difficulty in establishing a definitive diagnosis due to its paucibacillary nature, particularly in young children, and the high frequency of extrapulmonary disease. These factors make bacteriological confirmation difficult and may lead to higher mortality among children;
- Lack of a standard case definition for diagnosis and notification of TB cases in children. The recording and reporting systems of many countries do not use the same broad groupings for smear-positive, smear-negative and extrapulmonary TB than are currently used for adults. In addition, in a few countries, children with a positive tuberculin reaction but no other sign of active TB disease may be notified as disease cases;
- Relatively low public health priority given to tuberculosis in children as compared to adults, as it is usually smear-negative and thus not thought to contribute to the spread of TB. This low priority status is evident in the fact that many countries do not have age breakdowns of notification data, and, where they do, usually include only smear-positive cases, although such cases represent only a small subset of the total burden. Until 2006, the WHO global tuberculosis notification database included age breakdowns only for smear-positive tuberculosis cases. Since 2006, data on smear-negative and extrapulmonary cases broken down by age categories 0-4 and 5-14 years have been requested;
- Pediatricians are often unaware of reporting requirements (reporting from hospitals is typically nonexistent in many HBCs);
- Some studies suggest that countries reporting a high proportion of extrapulmonary TB in children may, in fact, be under-diagnosing cases of childhood pulmonary TB. Other countries lack the capacity to diagnose extrapulmonary TB entirely.

Despite this, in young children, TB is an event that denotes recent transmission, and thus tells us about the underlying epidemiology more broadly. Also, infected children of today become TB cases in the future, so they should not be ignored.

**Methods**

In order to gain a better understanding of the data and its reliability, we undertook an analysis of all the notification data available at WHO. From this, we calculated notification rates disaggregated by age and sex for smear positive cases and the proportion of notifications of children of 0-14 years of age to the total number of notified cases separately for each case types, for the 9 epidemiological regions and for individual countries. Furthermore, 95% confidence intervals were calculated and trend lines
established for all rates and proportions. We also calculated the ratio of smear-negatives and extrapulmonary cases to smear-positives for the age-group of 0-14 years of age. The subsequent discussion provides some examples of selected regions and countries in tables and time-series graphs, but an Excel file containing the data and means by which one can draw automatic graphs is attached. With these analyses, we aim to (1) better understand the data limitations and (2) prompt a discussion of the best approach to estimating TB incidence by age and sex.

Data availability
Until recently WHO only requested that countries report TB notification data by age for smear positive cases, even though age-disaggregated data for smear-negative and extrapulmonary cases was available in many countries. The data availability is as follows:

- Smear-positive cases, disaggregated by sex and age: 0-14, 15-24, …65+
  - 1995: 55% countries, 74% of global burden
  - 2007: 69% countries, 98% of global burden
- Smear-negative and extrapulmonary by age: 0-14
  - 2006: 51% countries, 30% of global burden
  - 2007: 52% countries, 38% of global burden
- All cases by age: 0-4, 5-14
  - 2006: 36% countries, 22% of global burden
  - 2007: 41% countries, 25% of global burden

Availability of smear+ age/sex-disaggregated data, % of regional burden (2007)

<table>
<thead>
<tr>
<th>Year</th>
<th>Global</th>
<th>AFR</th>
<th>AMR</th>
<th>EMR</th>
<th>EUR</th>
<th>SEA</th>
<th>WPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>85.0</td>
<td>81.2</td>
<td>28.5</td>
<td>90.0</td>
<td>44.9</td>
<td>95.5</td>
<td>91.7</td>
</tr>
<tr>
<td>2006</td>
<td>98.3</td>
<td>95.6</td>
<td>96.8</td>
<td>99.7</td>
<td>94.4</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>2007</td>
<td>97.8</td>
<td>93.5</td>
<td>99.4</td>
<td>99.9</td>
<td>94.9</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Approach to estimating TB incidence by age and sex
Given that a reasonably long time-series of notification data is currently available only for smear-positive cases and the current smear-negative/extra-pulmonary data are very recent and unstable, our initial approach to estimating TB incidence by age and sex was to use a very simplistic approach. It is described below.

To estimate by sex, the approach uses the estimated incidence for all cases and applies the sex distribution of notified smear-positive cases (or a trend line, so as to be more stable) to all incident cases. Inherent in this are the following assumptions:

- Case detection rate is constant for men and women
- Case detection rate is constant for smear-positive, smear-negative and extrapulmonary cases
- The notifications trends in men and women reflect the true underlying gender distribution of incident cases
To estimate by age, the approach uses the estimated incidence for all cases and applies the age distribution of notified smear-positive cases (or a smoothed series) to all incident cases. Inherent in this are the following assumptions:

- Case detection rate is constant for all age groups
- Case detection rate is constant for smear-positive, smear-negative and extra-pulmonary
- The notifications trends by age reflect the true underlying age distribution of incident cases

To estimate by age and type, the approach uses the estimated incidence of smear-positive cases, stratified using the age distribution of notified smear-positive cases to estimate smear-positive incidence by age. Then, it extrapolates the ratio of smear-negative and extrapolmonary to smear-positive cases from the last two years of data to all previous years. Inherent in this is the following assumption:

- Relationship between case detection rate for smear-negative and extra-pulmonary TB has been constant over time

Below we review the potential validity of these assumptions.

Can we assume that the case detection rate has been constant for men and women and by age group over time? And can we assume that the gender and age-stratified notifications reflect the true underlying incidence patterns?

**Trends in TB sex disaggregated notification rates – by type**

![Graphs showing trends in TB sex disaggregated notification rates](image)

Globally, we see increasing rates of smear-positive notifications, but these primarily represent increases in case detection and the influence of HIV. However, what is interesting to note is that, as case detection increases, the gap between notifications of smear positive cases among males vs. females is increasing, with more notifications in men. Is this the epidemiology of the disease or does it reflect differences in access to care? At this stage, we do not know.
With limited data on smear-negative and extra-pulmonary cases, we cannot say much. The data are very unstable which is natural in the early stages of data collection. It is difficult to make assumptions based on these data.

Regionally, we see that the gap between male and female notifications is decreasing, and was already small, in the Eastern Mediterranean. In Africa-high HIV countries, the relationship between notifications among men and women has been stable over time. In the Established Market Economies, the gap between male and female notifications has been decreasing but is still relatively wide. And in the Western Pacific (largely due to the influence of China), we see that the gap is widening, with male notifications increasing relative to that of females.

When drilling down to specific countries, we see even more variation. In Afghanistan, we see rates in women increasing at a much higher rate than for men. Could it be that men in the military are not notified by the NTP? In Pakistan, the rates are identical, and we will need to check if there is a problem with data coding. In Russia, the gap between
male and female notifications has been and remains very wide. Is this due to other underlying health issues in men that affect TB risk or under-notification in women? In China, the gap was fairly consistent, but the rates have diverged in recent years when notifications increased. Is there a differential case-finding effort? Are men seeking employment in urban areas increasing notifications?

As is shown above, there is considerable variation between individual countries, between countries and their associated regions, and between regions in terms of the sex distribution of TB notifications over place and time. In no case is the reason for these temporal and spatial variations obvious, thus making it difficult to utilize broad assumptions about the true underlying pattern in countries or regions regarding the relationship between male and female incidence.

Similar uncertainties emerge when we add the age variable to the analysis.

Globally and regionally, we see some consistent patterns, such as very low rates in children, in which female notifications are similar or a bit higher than male, followed by an abrupt increase in adolescence with an increasing male/female gap with age. In the Eastern Mediterranean, female notifications surpass male notifications in certain age groups, though it is not known why. One hypothesis is that countries where a high proportion of young men are engaged in military services (which often have TB services that are distinct from the NTP), male cases are under-reported. However, this pattern is influenced primarily by the smear-positive notifications. As we also add case type to the analysis, more differences become evident.

In looking at age alone - particularly the proportion of childhood TB - there remains considerable difficulty in interpreting the data.
In the Africa-high HIV region, we see a higher overall incidence which naturally leads to a higher incidence in children. However, Africa-low HIV has higher rates of TB than Latin America but the proportion of childhood TB is less. Is this due to a reduced capacity to diagnose kids than that found in Latin American countries? Or is it because it is not the norm to notify childhood cases of TB? The reasons for this are not known as yet. Established Market Economies have lower incidence and, thus, lower transmission than other regions; however, their notifications show a higher proportion of childhood cases than the South-East Asian Region or the Western Pacific. And what could explain the differences between Central and Eastern European countries? There is anecdotal information to suggest that some Eastern European countries notify TB infection in children as cases. This is an issue with case definition rather than the underlying epidemiology. And finally, what could explain the remarkably low proportion of childhood cases notified in the Western Pacific? Is it simply because pediatricians and pediatric hospitals do not report? The answers to these questions are currently unknown.

Because we only have these data over time for smear-positive notifications, we can also see the time series for each region of the proportion of these cases that are aged 0-14.
Globally, we see the proportion of smear-positive cases in children under 15 is decreasing, but that is influenced by high-burden countries such as China and India. There has been a notable decrease in Africa-low HIV countries over time and a small increase in the Eastern Mediterranean Region in recent years.

Here we see the distribution of the proportion of case in children under 15 across a selection of European countries.

**Smear + notification trends of % 0-14/all, for European countries, 2007**
Russia and Poland are close to zero - do they notify childhood cases in these countries? And why is Azerbaijan so different from the others? Again, we do not yet have the answers to these questions.

We can look more closely at a couple of countries where we have additional data.

**Smear + notification trends of % 0-14/all, for Sudan and Japan.**

Sudan has a relatively high proportion of smear-positive childhood notifications, but this proportion has been decreasing over time. Japan reports a very low proportion of smear-positive cases among children, and this rate has been decreasing in the late 90s and since then relatively stable.

We can try to find explanations for the pattern seen in Sudan by looking more closely at the overall TB/HIV situation there.

**Sudan- estimates of TB incidence and HIV prevalence**

Sudan first began notifying TB cases to WHO in the early 1990s. In the initial years of reporting, it is likely that many prevalent cases were reported as incident cases. The notifications have been reasonably stable over the last decade, but they are thought to have problems with case detection and it is estimated that notifications represent only
about 50% of incidence cases. We must also consider war and conflict in the country and the degree to which different zones are affected. Although HIV prevalence is not as high in adults as it is in some neighboring African countries, it is not decreasing as in some other African countries, and the HIV prevalence among TB cases still increasing. Given all of this, it is very unlikely that the observed decrease in the proportion of reported TB in children reflects a true decrease in incidence and transmission.

As mentioned previously, a common feature in TB notifications is that the initial years of are usually the most unreliable. If projections are to be built on which to base our estimation process for earlier years, influential outliers have to be removed beforehand.

**Proportion of all smear-positive notifications aged 0-14, Netherlands and Greece: effect of removing influential outliers**

![Graph showing proportion of all smear-positive notifications aged 0-14, Netherlands and Greece: effect of removing influential outliers.](image)
Can we assume that the case detection rate has been constant for smear-positive, smear-negative and extra-pulmonary cases over time?

In Africa, the influence of HIV is likely reducing the gap between male and female cases and maintaining high rates of extrapulmonary TB across multiple age groups. However, in Eastern Europe, the male/female gap is not as obvious for extra-pulmonary cases as it is for pulmonary TB. The extra-pulmonary rates decline gradually with increasing age.

Again, we can take a more in-depth look at childhood cases in particular, where the data are even more difficult to interpret.

Here again we see that the reliability of data collection in the early years of reporting can be quite variable. Below is the percent of smear-negative and extra-pulmonary cases reported in children under 15 for selected countries in the initial two years of notifications. While the proportions seem reasonable for Belgium and Germany, the data from Canada and Greece are questionable.

Proportion of all smear-negative and extrapulmonary notifications aged 0-14, for selected countries.
Uneven ratio smear- and extrapulmonary to smear+ for children of 0-14 years. Box plot by country.

Perhaps the most striking feature of childhood TB data is the ratio of the combined smear-negative and extra-pulmonary to smear-positive notifications. These data represent the last two years which are the only years for which the data are available. The median ratio is reasonably constant around 10 to 15 for all regions except the Eastern European countries. However, there is quite a lot of variation within the regions, in particular for the Eastern European countries where the median ratio is around 35. One possible explanation for this is the notification of TB infection among children as TB cases. The Established Market Economies, which have better means of diagnosing smear-negative and extra-pulmonary TB among children would be expected to have a narrower range of ratios.

Given this discussion, the paucity of data, and the obvious questions they elicit, it is clear that there are many factors which must be considered when analysing TB notification data by age and sex. These include:

- Demographic changes in the population over time
- Changes in the proportion of TB in different age groups over time due to underlying changes in TB epidemiology
- Over-reporting of some age and sex groups (reporting of infection as disease in Eastern Europe; misdiagnosis of TB in older age groups)
- Under-reporting of childhood TB

Nelson et al. attempted to calculate estimates of TB incidence in children. The authors first used estimates of the incidence of new smear-positive cases in each country. They then stratified cases by age group using the age distribution of notified new smear positive cases in each sub-region. These estimates assumed that the age distribution of actual smear-positive cases was the same as that of notified smear-positive cases. The authors noted that "published estimates of the proportion of cases expected to be smear-positive by age were then used to calculate the estimated total number of cases in each age group"; however, no reference for these published estimates was provided.
Next steps

The examples above highlight the value and the need to study routine records and reports that sharpen the focus on gender and age issues for TB control. The reasons for these differences need to be explained, and they are likely to result from various factors, including access to care, ethnicity, particular forms of TB, the HIV co-epidemic, especially in Africa and similar high prevalence settings, as well as other diverse biological, social and cultural variables.

1. The possibility exists to evaluate trends in case detection of childhood TB making use of data already available within some NTPs. Further in-depth analyses of notification data, including treatment outcome data, could also be performed. Additional data do exist in some countries, but it would require a special study to extract these data. As it is not part of routine reporting, one could make a special request for notification data by age, sex and case type for a series of years (unlike the annual data collection form which asks only for the current year). This could be accompanied by a questionnaire asking about the following:
   - Case definition and changes over time

<table>
<thead>
<tr>
<th>Country</th>
<th>Children aged 0–14 years</th>
<th>% of TB occurring in children</th>
<th>Estimated TB case rates &lt;15 years of age*</th>
<th>Estimated case rate†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afghanistan</td>
<td>17,540</td>
<td>25.3</td>
<td>189</td>
<td>324</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>33,166</td>
<td>10.2</td>
<td>61</td>
<td>236</td>
</tr>
<tr>
<td>Brazil</td>
<td>23,520</td>
<td>20.7</td>
<td>47</td>
<td>66</td>
</tr>
<tr>
<td>Cambodia</td>
<td>3,966</td>
<td>5.3</td>
<td>70</td>
<td>571</td>
</tr>
<tr>
<td>China</td>
<td>86,978</td>
<td>5.3</td>
<td>27</td>
<td>129</td>
</tr>
<tr>
<td>Democratic Republic of Congo</td>
<td>24,052</td>
<td>16.1</td>
<td>106</td>
<td>306</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>28,675</td>
<td>16.1</td>
<td>95</td>
<td>272</td>
</tr>
<tr>
<td>India</td>
<td>185,233</td>
<td>10.2</td>
<td>53</td>
<td>179</td>
</tr>
<tr>
<td>Indonesia</td>
<td>15,691</td>
<td>2.7</td>
<td>23</td>
<td>263</td>
</tr>
<tr>
<td>Kenya</td>
<td>22,124</td>
<td>16.1</td>
<td>167</td>
<td>450</td>
</tr>
<tr>
<td>Mozambique</td>
<td>7,703</td>
<td>16.1</td>
<td>98</td>
<td>268</td>
</tr>
<tr>
<td>Myanmar</td>
<td>8,007</td>
<td>10.2</td>
<td>51</td>
<td>165</td>
</tr>
<tr>
<td>Nigeria</td>
<td>32,310</td>
<td>12.4</td>
<td>63</td>
<td>228</td>
</tr>
<tr>
<td>Pakistan</td>
<td>61,905</td>
<td>25.3</td>
<td>103</td>
<td>172</td>
</tr>
<tr>
<td>Philippines</td>
<td>12,167</td>
<td>5.3</td>
<td>43</td>
<td>304</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>7,778</td>
<td>4.2</td>
<td>30</td>
<td>126</td>
</tr>
<tr>
<td>South Africa</td>
<td>35,449</td>
<td>16.1</td>
<td>237</td>
<td>501</td>
</tr>
<tr>
<td>Thailand</td>
<td>2,317</td>
<td>2.7</td>
<td>15</td>
<td>141</td>
</tr>
<tr>
<td>Uganda</td>
<td>12,099</td>
<td>16.1</td>
<td>103</td>
<td>320</td>
</tr>
<tr>
<td>United Republic of Tanzania</td>
<td>18,890</td>
<td>16.1</td>
<td>118</td>
<td>337</td>
</tr>
<tr>
<td>Vietnam</td>
<td>7,559</td>
<td>5.3</td>
<td>29</td>
<td>183</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>12,267</td>
<td>16.1</td>
<td>221</td>
<td>603</td>
</tr>
<tr>
<td>Total for the 22 high-burden countries</td>
<td>659,397</td>
<td>9.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Case rates per 100,000 children. United Nations population estimates used for denominator.
† Case rates per 100,000 population.
TB = tuberculosis.
• Policy of notifying smear-positive, smear-negative, extra-pulmonary and TB infection by age and changes over time
• Use of clinical algorithms for:
  o TST
  o X-ray
  o Smear and culture examination
• BCG coverage historical trends
• Contact investigation policy and practice
• Chemoprophylaxis

A previous study by IUATLD/CDC in 2002 attempted to extract this information with limited success, but the questionnaire could form the basis of a new data collection instrument with some modifications.

2. An "onion model for kids" exercise, in which childhood TB experts from the countries would be asked to make a quantitative assessment about the percent of children with TB who may be missed by the surveillance system at the various layers of the onion, could be eventually incorporated to the above mentioned questionnaire, which is to be sent to countries for completion. An alternative would be to use the exercise in the context of regional workshops and country missions, so that it is done in an interactive way, where country participants are challenged about their opinions and prompted to substantiate them with evidence from various sources.

3. Another possibility which deserves further attention is the analysis of the age-sex disaggregated TB mortality data available at the WHO mortality database. It may be possible to better understand the epidemiology of TB and to assess the number of incident and prevalent cases at different age groups and for men and women by analysing TB mortality data.

4. A further idea would be to compare the TB notification rates in women of reproductive age and in young children, which depends on the assumption that transmission preferentially occurs from mother to child (as opposed to from other index cases).

Conclusions
We have shown the lack of historical data for the breakdown by age and sex of smear-negative and extra-pulmonary TB, as well as the variation in the age and sex distribution of all TB cases across countries, regions and the world. In addition, there exist uneven policies and practices with regard to the following:
• Collection of smear from children with pulmonary TB for sputum and culture examinations;
• Case definitions for childhood TB - infection may be notified as disease;
• Case definitions for childhood TB in its clinical forms - smear-negative, extra-pulmonary
We also lack information on the duration of TB and case fatality rates across different age groups.

Until we have a better understanding of these phenomena and how they vary over time and place, we should not expect to draw reasonable conclusions from the data in their current form. Additional qualitative and quantitative information is required.

**Bibliography**