Questions to the Task Force

1. What are your general comments on progress to date in applying the Task Force framework?

2. Do you agree with the proposed next steps? Is there anything important that is missing or should be deleted?

3. Which of the standards listed in the annex do you agree or disagree with, and do you have any suggestions for additional standards that could be used?

4. Would you be willing to be part of a small working group responsible for further development of standards and benchmarks that would need to be met for surveillance data to be considered a direct measurement of disease burden?
Background Paper No. 2

Application of Task Force framework for systematic assessment of surveillance data via regional workshops and country missions: Progress to date, challenges faced, lessons learned, next steps

This paper summarizes the main elements of the Task Force framework that was endorsed in September 2008, describes the methods used by WHO to implement it in practice in 2009, explains the challenges faced and lessons learned to date, and sets out the proposed next steps for discussion in the fourth Task Force meeting in March 2010.

1. Summary of Task Force framework

A substantial amount of surveillance data are available to assess the burden of disease (cases and deaths) due to TB. The Task Force developed and endorsed a conceptual framework for the systematic assessment of these data in September 2008 (Figure 1), with two main objectives:

- To improve estimates of the burden of disease due to TB by making better use of surveillance data;
- To develop clear recommendations for how to strengthen national TB surveillance systems and associated plans to implement these recommendations.

The ultimate goal is that surveillance data should be of sufficient quality and coverage that TB incidence can be measured directly from TB notifications and TB mortality can be measured directly from vital registration records. For this to be possible, surveillance data and systems must meet certain (pre-defined as far as possible) standards.

Figure 1. Framework for the assessment of TB surveillance data.
The framework is made up of three major interrelated components (Figure 1):

1. Assessment of the quality and completeness of notification data for TB cases (in the TB routine notification system) and deaths (in the vital registration system);
2. Assessment of the extent to which notification and vital registration data reflect trends in TB incidence and mortality; and
3. Assessment of the number of incident TB cases and deaths that are “missing” from surveillance data (“undetected” cases and deaths).

These components are explained in more detail below. For a comprehensive review, please refer to the Policy Paper.

1.1 Data quality

Assessing data quality involves analyses of the completeness and reliability of data reported within the existing notification system.

Completeness is assessed by comparing the number of reports expected versus the number of reports received at national and subnational levels. For such assessments to be possible, countries must have systems in place to monitor the expected and actual movement of data from one reporting level to the next (e.g. reporting unit to district, district to province, province to central unit) within the country.

Assessment of data reliability includes analysis of whether cases are classified correctly (for example, whether retreatment cases are classified correctly) and whether duplicate records exist. It also includes assessment of the internal and external consistency of data. This can be done by exploring the variability in TB notification and death patterns geographically (looking at sub-national variation) and over time (sub-nationally and nationally). External consistency is assessed by comparing observed values with those that would be expected based on existing knowledge of TB epidemiology. For example, reliable data exist on the fraction of pulmonary cases that are sputum smear-positive, and these data can be compared with figures reported at national and subnational levels.

1.2 Trends

The purpose of the trends assessment is to examine whether or not - or the extent to which - changes in notifications are due to real changes in TB incidence and changes in reported deaths are due to real changes in TB mortality. This involves assessment of changes in case-finding efforts, case definitions, the coverage and quality of the vital registration and notification system, and policy changes affecting reporting of cases and deaths, all of which could cause changes in notifications and registered TB deaths without an underlying change in TB incidence and mortality; and assessment of changes in factors that may cause real changes in TB incidence and mortality, such as the HIV epidemic, an influx of immigrants from endemic countries, and the prevalence of risk factors for TB such as smoking. There may also be increasing or decreasing numbers of outbreaks or cases in specific settings such as prisons due to changes in control efforts; the effects of these changes are particularly noticeable in low incidence countries where outbreak control has a more profound effect on overall incidence, and thus notifications. Changes in the overall performance of the TB control programme can also influence notifications. In situations where infectious cases are quickly removed from the community, TB transmission can be decreased and cure rates may be higher.
1.3 Are all cases and deaths captured in surveillance data?

The framework also aims to evaluate the extent to which all cases and deaths are captured in surveillance data. This can be done as follows:

- Application of the "onion model". This helps to assess where, why and to what extent cases are missing from surveillance data, using quantitative and qualitative data as much as possible. The major reasons why cases are missed from official notification data include lack of notification of cases by public and private providers, failure to identify cases that seek health-care as TB suspects and/or to properly diagnose them, and lack of access to health services;
- Through evidence generated by inventory and capture-recapture studies;
- Through evidence generated by prevalence surveys; and
- Cross-validation of notification and vital registration data.

**Figure 2. The onion model.**
When applying the "onion" model (Figure 2), examples of the quantitative and qualitative data that can provide direct or indirect evidence of the fraction of cases missing from surveillance data include:

- Access to health care from demographic and health surveys data
- Overall performance of health systems as measured by:
  - Infant mortality ratio
  - Number of primary health care units or doctors per population
  - % of assisted births
  - Quality of care
  - Level of training/qualifications/competency of staff in primary health care services
- Performance of TB diagnostic systems
  - % people who died from TB (vital registration data) and had never accessed TB diagnosis and treatment
  - Results from external quality assurance (EQA) of laboratories
  - Data from surveys of the knowledge, attitudes and practices (KAP) of patients and health staff
- Contribution of different TB care providers
  - Health expenditure in the private or nongovernmental organization sectors, out-of-pocket expenditure
  - Regulations about reporting of cases and their enforcement in practice
- TB drug distribution: countries where anti-TB drugs are exclusively distributed by the NTP, in particular those where drug disbursement is dependent on case notification, have fewer cases diagnosed and unreported by non-NTP providers.

Inventory studies are studies in which cases are identified from one data source and then sought in other data sources to see if the reporting is consistent and accurate across the appropriate registries. This can be done by cross-checking existing registers (national TB register, hospital records, laboratory registers, etc.), or by introducing registers for a defined period of time (for example to record cases diagnosed in private hospitals or by private practitioners).

Capture-recapture studies aim to estimate how many cases may be missed from the various layers of the onion. It remains to be determined whether the methodology used in these studies can really account for frequent violations to the underlying assumptions.

TB prevalence surveys can provide not only estimates of the burden of disease, but also important insights into treatment seeking behaviour. By characterizing and quantifying the cases identified through a prevalence survey (most of whom are typically identified exclusively by the survey, while a minority had already been identified as a result of routine programme case detection), useful information is obtained on where and why cases are being missed.

Vital registration data can be used to identify the extent to which people who died from TB accessed TB care and were recorded in notification data.
2. WHO regional workshops and country missions: progress to date, methods used and case studies

2.1 Overview of progress to date

In 2009, the framework was used in four regional workshops. These covered 18 countries in the European Region, 9 countries in the South-East Asia Region, 22 countries in the Eastern Mediterranean and 13 countries in Latin America (Figure 3). In addition, the framework was applied in three country missions to the Philippines, Tanzania and Viet Nam.

![Figure 3. Progress in applying the framework in 2009.](image)

In all workshops, application of the framework led to the generation, analysis and interpretation of a large amount of useful information, and it was possible to update estimates of disease burden, to identify major ways in which surveillance systems need to be strengthened, and to develop action plans for strengthening surveillance. At the same time, several challenges were faced and lessons learned.

The next subsection explains the methods used to translate the framework into practice in these regional workshops and country missions. This is followed by a set of case studies.

2.2 Methods used

Using standard Excel templates, representatives from each country were asked to provide national and sub-national data on TB case notifications, infrastructure and case-finding efforts for the years 1995–2008. The requested data were as follows:

- Population by age and sex
- TB case notifications by case type (smear-positive, smear-negative, extrapulmonary)
- TB case notifications by age and sex
- Active case finding efforts
  - Number of people screened for TB through active case finding
  - Number of new pulmonary TB patients detected through active case finding
- Number of new and re-treatment TB cases reported by non-NTP/non-MoH providers
• Number of foreign-born or non-citizen new and re-treatment TB cases reported
• MDR-TB
  o Number of new TB cases for whom diagnostic drug susceptibility testing (DST) was done
  o Number of MDR-TB cases among new TB cases
  o Number of re-treatment TB cases for whom DST was done
  o Number of MDR-TB cases among re-treatment TB cases
• TB/HIV
  o Number of new and re-treatment TB patients for whom HIV status was known
  o Number of HIV-positive TB cases (among new and retreatment cases)
  o HIV prevalence in the general population (%)
• Laboratory infrastructure
  o Number of laboratories performing smear microscopy and culture examination
  o Number of smear microscopy laboratories for which EQA was carried out
  o Of smear laboratories for which EQA was carried out, number with adequate/satisfactory performance
• Number of dispensaries, clinics and hospitals
• Number of non-NTP providers and hospitals
• Number of staff
  o Number of core managerial staff dedicated to TB control at each major administrative level
  o Number of doctors working full-time for TB control
  o Number of nurses working full-time for TB control
• Other (chronic respiratory cases, TB suspects examined, slides examined, % culture-positive/smear-positive)

These data were then used to complete workbooks\(^1\) designed to answer the following questions:
• Are TB notifications from the existing recording and reporting system as complete as possible (are all reports from the lowest reporting unit upwards reaching the central unit)?
• Are TB notifications reliable (limited or negligible misclassifications)?
• What proportion of incident cases are missing from routine notifications and why?
• How have TB notifications changed over time?
• Do changes in notifications over time reflect operational changes (e.g., improved case finding)?
• Do changes in notifications over time reflect changes in the underlying epidemiology?
• What specific studies/activities are suggested to improve TB surveillance and programme monitoring & evaluation?

\(^1\) Workbook used in the workshops "Improving estimation of TB disease burden via systematic assessment of surveillance data". Available at http://www.who.int/tb/advisory_bodies/impact_measurement_taskforce/meetings/improving_estimates/en/index.html
2.3 Case studies

Five case studies are provided below, to illustrate how the Task Force framework has been implemented in practice. These case studies are for Tanzania, Bahrain, Latvia, Mexico and the Russian Federation.

Case study 1: Tanzania

The Task Force conducted a mission to Tanzania in September/October 2009, a detailed report of which is available upon request. With one of the oldest TB control programmes in Africa, there were considerable amounts of data to review.

The main findings from this mission according to the different components of the framework are shown in Figure 4. Further details are illustrated in Graph 1, Graph 2 and Table 1.

![Figure 4: Results from application of framework in Tanzania.](image)

The national data show a big increase in total cases since 1979, with a more marked increase in smear-negative and extra-pulmonary cases from the mid-1980s until 2001 followed by a plateau and a slight decrease in recent years (Graph 1). At the national level, the data appeared internally consistent with no major anomalies, and and mostly within the range expected according to TB epidemiology at national level (Graph 1 and 2). However, at the subnational level there was considerably variability by region, some but not all of which could be explained by the distribution of HIV across the country and variations in laboratory
capacity for smear examination (Graph 2). It appeared that there were some data quality issues and/or differences in notification policies that accounted for some of the regional variation observed.

**Graph 1. Tanzania national trends in TB notifications by case type, from 1979.**

![Graph 1](image1)

**Graph 2. Proportion of smear-positive/pulmonary, by year of notification and by region in 2007.**

![Graph 2](image2)

When trends were analysed, there was evidence that both the HIV epidemic and case-finding efforts had contributed to increases in notifications. The challenge was to disentangle these effects, and the data required to do this - notably data on TB notifications by case type disaggregated by HIV status for different time periods - were insufficient. There were survey data on the proportion of notified cases infected with HIV for 1992 and 1996, but no survey or surveillance data for more recent years. Surveillance data on notifications of TB cases disaggregated by type of case and HIV status are now being collected at the district level, but these data were not being reported from the districts upwards and were therefore unavailable at the national level at the time of the mission. This highlighted the need for better and more timely reporting of data between levels.
In the absence of more comprehensive data, the revision of estimates of disease burden relied mainly on application of the "onion" model for 3 different years. Conducting this exercise in detail with the country produced several interesting insights.

Country representatives were asked to make initial assessments of the percent of all cases that were unreported from each layer of the onion model for three years - 1997, 2003 and 2008 - with ranges around each estimated proportion. For 2008, country representatives initially estimated that about 86% of all new cases were detected by the TB programme (Table 1). According to this initial estimate, of the 13.8% of cases that were missing, 1.5% were diagnosed but not reported to the NTP; 0.9% were detected by the non-NTP public or private providers but not reported to the NTP; 2.4% presented themselves to the health services but were not diagnosed; 2.7% had access to health services but did not seek care; and 6.3% were estimated to have no access to health services.

The initial figures indicated by country representatives were challenged by the mission team, and further discussions and analyses led to their revision:

- The mission team agreed that the number of cases diagnosed by NTP or private and public non-NTP providers but not notified to the NTP register was probably small, contributing less than 2.5% of the total number of cases (layers 2 and 3). The main argument in support of this was the centralized distribution of first line anti-TB drugs by the Ministry of Health, with no purchase or distribution of first-line drugs possible in the private sector.
- The mission team agreed that the percentage of cases without access to health services is probably less than 7%. The evidence to support this argument was that 93% of the population lives within 10 km of a basic health care facility (as referenced in the grey literature - please see the mission report).
- The mission team felt that the estimated proportion of cases missed in layers 4 and 5 (cases seeking health care but not being diagnosed, and cases with access to health-care but not seeking health care) was too optimistic (too low). This was based on data showing that many new TB cases have been diagnosed following the introduction of recent interventions, such as screening for TB among people living with HIV (PLHIV) and recent improvements to the laboratory network (notably a large increase in the number of smear microscopy units throughout the country). These interventions have not yet been expanded to all clinics and districts in the country, suggesting that there is still room for improvement in terms of case-finding efforts. In addition, studies have shown average treatment delays as long as 3 months in some areas of the country, which may not only contribute to increased mortality but also increased transmission. This information was not compatible with an estimate that only 5% of cases nationally were being missed in layers 4 and 5. The proposed revisions based on further discussions with country representatives were that 9% of cases presented to the health services but were not diagnosed, and 5% had access to health services but did not seek care. These revisions were agreed to by the country representatives.

The mission was made up of findings and discussions that led to clear recommendations for how to strengthen surveillance, some with immediate effect. These included:

- District level data should be provided and analysed nationally in order to reconstruct historical notification data from this level as far back as possible, in particular data on notifications of cases disaggregated by HIV status;
• The form used to collect data on HIV status from districts should also include TB case type (smear status, extra-pulmonary), and record treatment outcomes by HIV status;
• There is an urgent need to roll out the electronic recording and reporting system which is currently only being tested in some limited areas. Without such a system, it is difficult or impossible to adequately assess many aspects of data quality;
• More widespread adoption of updated recommendations on recording and reporting is required;
• Surveillance data should be analysed routinely to inform programme performance at all levels. This will require strengthened capacity in analysis of surveillance data, initially at the national level. It will also be necessary to improve feedback of data analysis and interpretation to TB and other staff working at the peripheral levels;
• Strengthened supervision of programme activities at the district level is urgently needed, with particular attention to the supervision of M&E and laboratory activities.

Table 1. Expert opinion about the proportion of missing cases out of all new cases using the onion model. Comparison of the country representatives' initial opinion and the revised opinion based on further discussions and analyses.

<table>
<thead>
<tr>
<th>Onion layers</th>
<th>Country representatives' expert opinion % of missing out of all new cases</th>
<th>Revised opinion % of missing out of all new cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lowest possible value</td>
<td>Most likely value</td>
</tr>
<tr>
<td>6. No access to health care</td>
<td>4.1</td>
<td>6.3</td>
</tr>
<tr>
<td>5. Access but do not present themselves</td>
<td>1.4</td>
<td>2.7</td>
</tr>
<tr>
<td>4. Presenting but not diagnosed</td>
<td>1.6</td>
<td>2.4</td>
</tr>
<tr>
<td>3. Diagnosed by public non-NTP</td>
<td>0.7</td>
<td>0.9</td>
</tr>
<tr>
<td>2. Diagnosed by NTP but not notified</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>Sum of % of missing cases: layers 2 to 6</td>
<td>8.8</td>
<td>13.8</td>
</tr>
<tr>
<td>Country's estimates of CDR</td>
<td>91.2</td>
<td>86.2</td>
</tr>
<tr>
<td>WHO estimates of CDR (all cases - 2007)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference (country's - WHO estimates)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Country representatives were receptive to these recommendations, and it is anticipated that they will be implemented in the near future, thus facilitating subsequent data analyses.

Experience in regional workshops indicates that the recommendations above are applicable in many other countries.
**Case study 2: Bahrain**

Surveillance data from Bahrain was reviewed during a regional workshop held in Cairo in October 2009. The most notable outcome of the discussion and analyses related to the national level notifications. As can be seen from the Graph 3, notifications jumped considerably after 1997 and then stabilized around 2000.


Bahrain is a low-incidence country with no known epidemiological, socioeconomic or other factors that could explain this increase, although in recent years at least part of the increase could be due to TB cases found among migrants. In discussion, the NTP manager explained that revisions to the M&E system and data collection procedures in the late 1990s had greatly improved efficiency. Among other measures that were introduced, anti-TB drugs started to be distributed only in two pharmacies in the country (MoH Hospital and Military Hospital), and notification became a requirement for drug distribution. This appeared to account for the sharp increase in notifications after that time, with the suggestion that notifications are now very close to incidence.

**Case study 3: Mexico**

Mexican surveillance data were reviewed during the regional workshop for the Americas in Brazil in September 2009. DOTS implementation began in 1995 and brought with it a unified notification system, the incorporation of new reporting units and active case-finding in prisons and high TB and HIV burden settings. It appears that these changes explained the increase in notifications during the period 1995–97. This was followed by a stabilization in notifications until around 2004, when another increase was evident. This increase corresponded to the implementation of Public-Private Mix activities. Furthermore, in 2006 13% (4 out of 32) states did not have at least one laboratory performing culture, whereas by 2008 only one of them still lacked a culture laboratory. These data on infrastructure and case-finding initiatives helped us to understand changes in TB notifications. At the same time, it was difficult to quantifying their individual contribution to changes in notifications, although the overall conclusions were a) that the increase in notifications reflected changes in case-finding and reporting and not changes in underlying TB incidence, and b) that with
the recently achieved high coverage and quality of the laboratory network and the TB notification system, notifications are likely to be very close to incidence.

**Graph 4. Mexico national trends in TB notifications and incidence estimates, from 1995.**

![Graph 4](image)

**Case study 4: Latvia**

Surveillance data from Latvia were reviewed as part of the regional workshop for Europe held in April 2009, and incidence estimates were revised based on this review. In Graph 5, the blue line represents notifications and the red line estimated incidence with confidence intervals. There was a rapid increase in notifications 1993–97 followed by a stabilization between 1997 and 2000 and then a slow decline. In workshop discussions, the main explanations for the increase in notifications were the economic disruption that followed the break up of the Soviet Union combined with major outbreaks of TB in the subsequently inflated prison population, and better reporting. It was difficult to quantify the relative contribution of real increases in incidence and better reporting. Therefore, we assumed a flat trend before 1997 with decreasing confidence bounds over time reflecting increased confidence in the notifications, and a steady rate of decline from 2001.

**Graph 5. Latvia national trends in TB case notification, incidence estimates and TB mortality (excluding HIV/TB deaths), from 1990.**

![Graph 5](image)
When studying the decrease in notifications after 2000, it was suggested that this was the result of TB control efforts. The NTP felt it was finding, reporting and treating nearly all incident cases. This was supported by the following observations: intensive case-finding measures were put in place, particularly in high TB-burden settings; there was an overall improvement in the economic situation that could have reduced incidence; and a simultaneous decrease in TB mortality recorded in VR data, which reinforces the assumption that the decline in incidence was real. Despite this, the MDR situation seems to have deteriorated over the same period. This is not expected to have influenced incidence overall, but might have contributed to higher case fatality.

**Case study 5: Russian Federation**

Russian surveillance data provide another example of the need to disentangle several factors that contribute to increased notifications and the difficulties involved in doing this. **Graph 6** shows the distribution of civilian and non-civilian cases reported cases (left axis) and their rates (right axis). As in Latvia, we know that there were multiple reasons for the increase in notifications that occurred in the 1990s. The performance of the NTP in terms of case detection and cure rates had been deteriorating since the mid-1980s and there was a significant increase in the prison population where TB incidence rates are extraordinarily high. This was thought to be a major contributing factor to the overall increase in incidence. However, in the 1990s one simple measure resulted in a rapid increase in reporting: this was the reporting of cases detected and treated outside of Ministry of Health institutions - notably cases among prisoners, soldiers, migrants and the homeless who were treated in "non-civilian" hospitals and had previously been reported to the Ministry of Justice. In the early 1990s, as the number of reports coming from the Ministry of Justice institutions increased, so did the proportion of non-civilian cases out of the total number of cases. This highlights the point that data on the source of referral is a very useful - and simple to collect - piece of information that can shed light on where cases occur, how incidence is changing over time in various sub-populations, and where interventions can be best targeted. To date, however, it is rare that such data have been available for analysis.

**Graph 6. Russia national trends in TB case notification, by source of referral.**

[Graph source: Russian surveillance data analysed by Dr. Florian Marx, during his internship at WHO STB/TME in 2008.]
3. WHO regional workshops and country missions: challenges faced and lessons learned

3.1 Challenges faced

During the workshops and country missions described above, several common challenges have emerged. These are listed below. The list starts with general challenges followed by challenges ordered according to the logic of the Task Force framework (i.e. data quality, assessment of trends, extent to which cases are missing from surveillance data).

- **Availability of data.** No country was able to provide all of the data requested (i.e. the data listed on pages 5–6), although data availability was better for more recent years. A particular weakness was historical programmatic data on case-finding efforts. Such data were more readily available for recent years, when countries have been asked to provide information to funding agencies, but even then it is rarely in a standardized format and often difficult to analyse. One important missed opportunity is the lack of data on the yield of cases found when Public-Private Mix (PPM) initiatives are introduced. It was rare to find information on the source of TB patients found in this way, such as the provider or type of health care facility in which the patient was diagnosed. This information could be included in routine recording and reporting forms as recommended by WHO, but most countries have not yet done this;

- **Understanding of TB epidemiology.** Workshop participants may have a relatively limited knowledge and/or training in TB epidemiology;

- **Analytical capacity, ownership and use of data.** Some country teams had strong analytical capacity and were using data for decision-making. This was particularly evident among countries that produce their own annual reporting documents. However, in general, access to tools and methods for analysis are fragmented and use of data limited. Many NTPs do not routinely evaluate their own data for anomalies or expected changes. Most countries have only one person - either an M&E officer or a data manager - at the central level who is responsible for compiling, organizing, analysing and disseminating the data. Few countries analyse sub-national data. Also, performance assessments are conducted infrequently, and data analysis is rarely done for the purposes of assessing data quality. There also appear to be problems regarding country ownership of data, with most data collection efforts viewed as necessary for external uses. While reporting is usually viewed as a necessary task, surveillance data are not commonly used to inform programmatic decisions;

- **Most countries still have paper-based recording and reporting systems.** Without electronic recording and reporting systems, it is difficult or impossible to assess several aspects of data quality. For example, it is impossible to assess if there are misclassifications or duplications as a patient cannot be followed through the various registries that exist. Electronic recording and reporting would allow automated or semi-automated checks for completeness of reporting, reliability, and duplications/misclassifications. This should facilitate the follow-up of patients who enter and re-enter the reporting system many times in the same or different health care units; these patients are more likely to have poor treatment outcomes and, as such, their care requires special attention. All analyses of internal and external consistency as well as trends would also be much easier to undertake if data at all levels were available in electronic format. Also, inconsistent definitions for disaggregated notifications by case type and source of referral are being used.
Electronic systems can produce automated analyses and reports, thus facilitating the work of M&E officers and supervisors who would then require training only in interpretation of findings, rather than in conducting the analysis and interpreting its output;

- **Difficulty in disentangling the effects of case-finding and other factors on changes in TB notifications over the past decade.** Case-finding efforts have improved broadly over the last decade but it remains challenging to disentangle the effects of these efforts from any changes in underlying TB epidemiology. It is particularly difficult in high HIV-prevalence settings, where data on notifications disaggregated by HIV status are required to analyse whether trends in notifications reflect underlying trends in TB incidence. Such data are usually not available, as explained above for Tanzania;

- **Too much reliance on expert opinion and too little objective evidence about the fraction of cases missing from surveillance data.** To date, assessment of the fraction of cases missing from surveillance data has been based on expert opinion among a few individuals (which may be biased), due to insufficient evidence based on data. Generally, only one or two participants per country have attended regional workshops, usually the NTP manager and the monitoring and evaluation (M&E) officer. Other things being equal, these participants are more likely to be biased in favor of their programme performance;

- **Initial over-estimation of the proportion of cases that are captured by TB notifications.** This is a particular problem given that assessing the fraction of cases missing from surveillance data is equivalent to estimating the case detection rate, an indicator that has long been used as a measure of programme performance. For instance, most countries initially report that only a small numbers of cases are found by the NTP but not reported. When guided through the onion model, participants were often taking their first-ever view of the various ways in which cases may be missed by surveillance.

- **Infrequent availability and use of vital registration data.** Currently, only a few countries with a high TB burden have vital registration systems of sufficient quality to directly measure TB deaths. Moreover, most NTPs do not view the analysis and interpretation of data on TB deaths from vital registration systems, when available, as part of their responsibility.

3.2 Lessons learned

Linked to the challenges described above, various lessons about how to improve application of the Task Force framework have been learned in the regional workshops and country missions conducted to date. These are listed below:

- **Electronic recording and reporting systems are urgently needed and are essential for assessing many aspects of data quality;**

- **Data on notifications by case type disaggregated by HIV status need to be routinely collected, especially in high HIV-prevalence settings;**

- **Need for better capacity to use data within NTPs.** This could include developing access to and use of profiles, dashboards, interactive graphics and mapping tools to strengthen country analyses and capacity to use these analyses to inform key decision making processes;
• **Need for better supervision of M&E activities at country level.** Although a supervisor figure is generally present within NTPs, their activities rarely relate to M&E terms of reference as many have simply not been trained to supervise these activities. There is a need to strengthen the supportive supervision mechanisms from national to regional, regional to district and district to facility/community levels. Countries need help in training supervisors, understanding the terms of reference for M&E supervisors, and developing easy tools to be used in the field. Related to this is the need for improved feedback regarding data analysis and interpretation to TB and other health care staff working at the peripheral level. These staff rarely have the opportunity to see the broader picture of the TB control efforts to which they contribute. The supervisor would be the best person to provide this feedback. When staff are able to appreciate the importance of the data they are collecting, data quality is likely to improve;

• **Expert opinion needs to be challenged to prompt country participants to substantiate their assessments based on external evidence.** One partial solution would be to bring national TB epidemiology experts from outside the NTP to workshops and country missions, to participate in and help guide discussions;

• **Need for pre-defined standards that can be used for self-assessment or external assessment of the quality and coverage of surveillance data;**

• **Most countries are far from the standards required for disease burden to be measured directly from notifications and vital registration data.** This means that, given the diagnostic tools currently available, incidence is not directly measurable in most countries. Some countries may, however, meet the standards required to directly measure *trends* in incidence using notification data and to directly measure *trends* in mortality using vital registration records;

• **Need to progressively increase reliance on data to assess cases missed by surveillance and progressively reduce use of expert opinion;**

• **Systems for recording and reporting TB deaths within vital registration systems should be developed and better use made of vital registration data.** Strengthening vital registration systems can only be achieved as part of wider efforts to increase death registration and improve health information systems. However, NTPs could make better use of available data to better understand the epidemiology of TB and, in turn, to identify how deaths from TB could be reduced;

• **Need for more inventory studies and other operational research to provide evidence on where and to what extent cases are missing from surveillance data.** Although it is unrealistic to expect that countries with poor notification systems for TB cases and deaths will be able to provide high quality, widely representative evidence of case finding efforts and information on where cases might be missing, operational research could fill in some of the gaps. This research can be used to study the number of cases that are missing from TB notification data. These studies typically involve prospectively or retrospectively collecting data from places where TB cases may 1) be diagnosed but not notified, 2) seek care but not be diagnosed, and 3) experience symptoms but not seek care. Data from well-designed studies, with study sites sampled in a representative manner, may be a cheap but robust way to assess the situation of the country as a whole. Further details are provided in the recently published policy paper. These studies can be used to assess the quality of
notification data from places where it is thought to be particularly good or bad and to draw lessons from those with high quality data, while testing the implementation of measures aimed at improving data quality in others.

4. Next steps required to close data and conceptual gaps

Considerable progress has been made in applying the Task Force framework for systematic assessment of surveillance data in practice. The next steps required to make further progress are listed below.

Development of framework and related guidance

- Define the standards/benchmarks that need to be attained for a country's surveillance data to be considered a direct measure of TB disease burden (cases and deaths) and the evidence required to demonstrate these standards have been met. To date, clear standards/benchmarks for self-assessment or external assessment of the quality and coverage of surveillance data have not been defined (though some ideas are included in the Policy Paper). Such standards and related sources of evidence need to be developed and included as an integral part of the framework and the tools used to implement it in practice; some suggestions (related both to assessment of the surveillance system from which the data are generated and assessment of the surveillance data themselves) are listed in the Annex. It is clear that a good TB surveillance system is one in which all the components of the framework are routinely evaluated. It is also clear that such evaluations need to be done not only at the national but also the sub-national levels, as national level analyses may mask inconsistencies and anomalies that are only apparent sub-nationally. It is hoped that the Task Force will be able to define a few "marker" indicators that can be used to assess whether or not the standards required for direct measurement are met. A possible next step would be to establish a small group from within the Task Force to take forward this work, drawing in particular on experience from countries with strong surveillance systems. Identifying how to demonstrate that these systems meet the standards required for direct measurement of cases and deaths should help to define standards/benchmarks that would be generally applicable;

- Provide guidance on how to conduct inventory studies and other forms of operational research that can provide evidence about the fraction of cases missing from surveillance data. To date, assessment of the fraction of cases not captured in surveillance data has been based mostly on expert opinion. This needs to be progressively replaced with assessments based on routine data or research. Inventory studies and innovative operational research are included in the Task Force framework, but specific guidance on how to conduct inventory studies and what kinds of operational research are likely to be of greatest use is not yet available;

- Provide clear guidance on whether or not capture-recapture methods should be used to assess the coverage of surveillance data. Capture-recapture studies are part of the Task Force framework but relatively few such studies have been carried out. When studies have been conducted, there are concerns about whether underlying assumptions required for the methods to be valid are met. Clear guidance on whether or not such studies should be conducted, and if so then under what circumstances, is needed.
Development of capacity to use framework

- **Train a pool of consultants in the methods available to implement the Task Force framework.** A workshop to train consultants is planned for the last quarter of 2010. People who should be invited to this workshop need to be identified;

- **Provide training and related support to NTPs in the supervision of M&E work.** Training materials and standard operating procedures (SOPs) could provide terms of reference and guidelines for the M&E component of programme review missions, something which is often overlooked at present;

Practical application

- **Further workshops and country missions with more participation from Task Force members.** Regional workshops are planned for the 7 regional priority countries in the Western Pacific Region (Cambodia, China, Laos, Mongolia, Papua New Guinea, the Philippines) in June 2010 and for 20 countries in the African Region in September 2010. Other countries where use of the framework should be prioritized need to be identified;

- **Promote systematic assessment of surveillance data using the framework and associated tools in the context of Global Fund grants.** The Global Fund is strongly committed to assessments of data quality and strengthening of monitoring and evaluation systems. Tools developed by the Global Fund can be used alongside, or merged with, existing tools developed by WHO to implement the Task Force framework: for example, it has already been agreed that in future the WHO "workbook" will be used alongside the Global Fund's "MESS" tool\(^2\) (see also background paper number 4). With recent changes to the Global Fund "architecture", programme reviews will occur in countries with Global Fund grants every three years. The Task Force framework could be applied to assess the quality and coverage of surveillance data as part of these reviews, or ideally just prior to a review so that findings inform the full programme review and can be used as the basis for practical recommendations about how to strengthen surveillance using funding available from the Global Fund;

- **Promote implementation of electronic recording and reporting and provide associated guidance and support.** ERR is essential to improve assessment of data quality and to analyse whether trends in notifications and recorded deaths are a good proxy of trends in TB incidence and mortality. A small working group could be established to define standards for TB surveillance and the ERR system specifications required to achieve them. This should include consideration of whether or not the TB surveillance system should be integrated into the general (or infectious disease) health information system.

Disease burden indicators

- **Increase the availability and use of quality vital registration data for regular assessments of tuberculosis burden.** These data provide a relatively inexpensive direct measurement of disease burden that can be used in many settings, although it is rarely available in countries with a high TB burden. Development and use of mortality measurements from vital registration and validated interim mortality measurement systems need to be given high priority and used as a primary indicator

\(^2\) Monitoring and evaluation systems strengthening assessment tool.
of progress in TB control, to be analysed in combination with time-changes in routine notification rates (replacing the case detection rate). For further discussion, see also background paper number 11.

5. Questions to the Task Force

- What are your general comments on progress to date in applying the Task Force framework?
- Do you agree with the proposed next steps? Is there anything important that is missing or should be deleted?
- Which of the standards listed in the annex do you agree or disagree with, and do you have any suggestions for additional standards that could be used?
- Would you be willing to be part of a small working group responsible for further development of standards and benchmarks that would need to be met for surveillance data to be considered a direct measurement of disease burden?
Annex: Standards or benchmarks that should be attained for TB incidence and mortality to be measured directly from surveillance data: some suggestions

Standards or benchmarks to be attained by the TB surveillance system:

- NTPs (or equivalent) routinely assess the completeness of reporting of their TB notification data from the lowest administrative level to the national level (e.g. routine data quality audits, inventory studies cross-checking national databases with primary records).
- Mechanisms are in place to enter the results of laboratory examinations and other procedures, as well as treatment outcome data to the system as soon as they are available.
- Data on case finding efforts (e.g. number of TB diagnostic units, number of clinical staff working full-time for TB control, number of slides examined for diagnosis of pulmonary TB patients) and on the main determinants of TB (HIV, TB in prisons, TB in immigrants, …) are available for several years.
- Clear documentation of changes to policies regarding case definitions and reporting of cases for at least the last 10 years.
- NTPs (or equivalent) have a case-based electronic TB database for notifications.
- NTPs (or equivalent) check and correct for duplications and misclassifications in notifications (in particular of whether retreatment cases are erroneously classified as new cases). Linkage of national TB notifications coupled with the implementation of procedures to distinguish new and re-treatment or transfer-in records should have been performed within the last three years covering a period of at least five years, with mistakes corrected.
- NTPs (or equivalent) perform inventory studies using routinely collected data (in particular on TB laboratory data, TB data in HIV notification systems, TB hospitals and dispensaries, TB mortality from VR), and through special studies where appropriate (e.g. where there are a large number of providers of TB care not formally linked to the NTP).
- NTPs (or equivalent) analyse TB mortality data from vital registration systems wherever minimum standards are met together with TB notification data.

Standards or benchmarks to be attained by the TB surveillance data:

- Limited missing data for at least the main variables in the system, for at least 10 years.
  - Personal identifiers
  - Demographic information
  - Type of entry (new or retreatment, by categories)
  - Clinical form
- HIV status
- Smear and culture results
- Treatment outcome

- Limited evidence of duplications and misclassifications in full audits and spot checks, for at least 5 years.
- Notification data are confirmed as showing limited variability (need to be defined) over time and space for national and subnational data and are consistent with values expected according to current knowledge of TB epidemiology for at least 10 years, or notification data show considerable variability over time and space, which can be explained.

**Standards or benchmarks to be attained for direct measurement of TB trends:**

- Trends in TB notifications assessed to mirror trends in incidence, as demonstrated by:
  - Consistency of trends across different sources of notifications and/or,
  - No evidence of changes in case definitions, policies in reporting and case finding over around 10 years and/or,
  - Trends in TB notifications mirror trends in incidence after corrections for changes in case definitions, policies in reporting and case finding and/or,
  - Trends in TB notifications mirror trends in TB mortality, and no evidence in changes in case fatality rates.
- Trends in TB deaths assessed to mirror trends in mortality, as demonstrated by:
  - No evidence of changes in consistency in coding, policies in reporting and coverage of the vital registration system over around 10 years.

**Standards or benchmarks to be attained for surveillance systems to be considered as missing negligible number of cases and TB deaths:**

- Vital registration systems meets minimum standards of quality and coverage (e.g. proportion of ill defined causes ≤ 20% and coverage of ≥ 70%).
- Excellent geographical and financial access to health care facilities where TB diagnosis is available (e.g. ≥90% of population lives within X Km or X hours of a health care unit where TB diagnosis is available, diagnosis available free of charge and/or reimbursable through health care insurance, close to 100% of deliveries assisted by trained health care worker, low infant and maternal mortality ratio, low estimated out-of-pocket expenditures)
- Staff have proven competency to diagnose TB in all health care facilities where TB diagnosis is considered to be available (need to define how competency would be proven)
- Bacteriology laboratories are quality assured
- Limited or negligible undernotification from the private sector as evidenced by:
  - Limited health care expenditures in the private sector (e.g. ≤ 5%) and/or,
o Private sector only has access to anti-TB drugs following notification of a TB case (e.g. much of Latin America) and/or,

o Share of notifications reported by different types of care providers is consistent with known patterns of health care expenditure of health care seeking behavior and/or,

o Inventory studies (e.g. cross sectional or cohort surveys of TB patients diagnosed by private providers) in a random sample of providers demonstrate that cases diagnosed in the private sector have been notified and/or,

o Limited or negligible anti-TB drug sales in the private sector and/or,

o Policies/regulations requiring notification of TB cases which are enforced (need to define how this could be proved).

• Limited or negligible undernotification from the NTP (or equivalent) and non-NTP public sector as evidenced by:

  o All public providers only have access to anti-TB drugs following notification of a TB case and/or,

  o Policies/regulations requiring notification of TB cases which are enforced (need to define how this could be proved) and/or,

  o Inventory studies or similar in a random sample of providers demonstrate that cases diagnosed in the public sector have been notified.