Improving the estimation of the tuberculosis burden in India

Krycia Cowling, Rakhi Dandona & Lalit Dandona

Public Health Foundation of India, Plot 47, Sector 44, Gurgaon – 122002, National Capital Region, New Delhi, India.

Correspondence to Lalit Dandona (email: lalit.dandona@phfi.org).

(Submitted: 2 September 2013 – Revised version received: 7 May 2014 – Accepted: 18 June 2014 – Published online: 20 August 2014)

Abstract

Although India is considered to be the country with the greatest tuberculosis burden, estimates of the disease’s incidence, prevalence and mortality in India rely on sparse data with substantial uncertainty. The relevant available data are less reliable than those from countries that have recently improved systems for case reporting or recently invested in national surveys of tuberculosis prevalence. We explored ways to improve the estimation of the tuberculosis burden in India. We focused on case notification data – among the most reliable data available – and ways to investigate the associated level of underreporting, as well as the need for a national tuberculosis prevalence survey. We discuss several recent developments – i.e. changes in national policies relating to tuberculosis, World Health Organization guidelines for the investigation of the disease, and a rapid diagnostic test – that should improve data collection for the estimation of the tuberculosis burden in India and elsewhere. We recommend the implementation of an inventory study in India – to assess the underreporting of tuberculosis cases – as well as a national survey of tuberculosis prevalence. A national assessment of drug resistance in Indian strains of Mycobacterium tuberculosis should also be considered. The results of such studies will be vital for the accurate monitoring of tuberculosis control efforts, in India and globally.

Introduction

India accounts for an estimated 2.2 million of the 8.6 million new cases of tuberculosis that occur each year globally and harbours more than twice as many cases as any other country. A national programme for the control of tuberculosis achieved nationwide coverage in 2006 but this programme has limitations in terms of disease surveillance. All attempts to estimate the burden of tuberculosis in India are based on indirect methods characterized by substantial uncertainty and a lack of subnational detail. For a country of over 1.2 billion people, including, probably, more than 500 million individuals latently infected with Mycobacterium tuberculosis,
weaknesses in the available estimates of the tuberculosis burden are disappointing and limit effective policy-making.\(^3\) China – the country with the second-highest number of tuberculosis cases – has drastically improved its tuberculosis estimates since implementing a web-based system of mandatory case reporting, in 2005.\(^4\) More accurate estimates of the tuberculosis burden in India are needed to guide national policy-making, to improve the assessment of control efforts and to understand global trends in the incidence of tuberculosis. Here, we review the data currently used to estimate the tuberculosis burden in India and their limitations, and discuss options for the collection of new data that could yield improved estimates.

**Estimating the tuberculosis burden**

**Possible sources**

Ideally, in any country, tuberculosis surveillance is based on a comprehensive monitoring system to which all new cases are reported and a vital registration system that collects accurate data on the causes of all deaths. Such case monitoring and vital registration systems allow evaluation of the incidence of new infections and the levels of tuberculosis-related mortality, respectively. In countries lacking such systems, estimates of the tuberculosis burden are typically based on national tuberculosis prevalence surveys and national mortality surveys that assess causes of death. Although many countries have either undertaken prevalence surveys since 2002 or are planning to undertake such surveys by 2017,\(^1\) India is not one of them (see Appendix A, available at http://www.phfi.org/images/Publications/journals/2014_who_bulletin%20_tb_estimation_india_web_appendix_a.pdf). India has not implemented a national survey of tuberculosis prevalence since 1955.\(^5\) Tuberculin surveys – which allow the prevalence of latent *M. tuberculosis* infection to be estimated – are no longer recommended by the World Health Organization (WHO) for estimating the prevalence of active tuberculosis.\(^6\,7\)

**Data availability in India**

India’s current efforts and investments in generating reliable data for estimating its tuberculosis burden are inadequate, especially when compared with the corresponding efforts in countries with similar levels of wealth and tuberculosis endemicity. India does not yet have comprehensive systems for the reporting of tuberculosis cases or vital registration and it has only conducted scattered subnational surveys on the prevalence, incidence and mortality of tuberculosis.
Although the Revised National Tuberculosis Control Programme is working to improve the system for the notification of tuberculosis cases in India, it will be many years before every relevant provider is participating. The improvement of India’s vital registration system is an important larger goal that will also take a long time. In addition, there is no commitment to conduct a national survey on the prevalence of tuberculosis in India, even though it is 59 years since such a survey was last conducted.

Currently, estimates of the tuberculosis burden in India are predominately based on the numbers of cases that are notified and expert opinion on the corresponding level of underreporting. Such estimates could be made more accurate and less subjective if expert opinions could be replaced with empirical estimates of the level of underreporting in the case notification system.

The most widely-used estimates of the national burden of tuberculosis in India are produced by the Revised National Tuberculosis Control Programme, WHO and the Global Burden of Disease Study. The uncertainty in these estimates is illustrated by comparing the values for tuberculosis incidence (Fig. 1), prevalence (Fig. 2) and mortality (Fig. 3) in India from these three sources for the period 1990–2011.. The Revised National Tuberculosis Control Programme and WHO produce their estimates of incidence by dividing the number of case notifications by 1 minus the estimated proportion of all cases that are not reported. The level of underreporting has been estimated from expert opinions and from the results of two subnational studies – which indicated that only about 60% of tuberculosis cases in the study areas were notified. The Global Burden of Disease Study used substantially different methods and several covariates to try to strengthen estimates based on sparse data. Greater details for each methodology – and a discussion of the difficulties of measuring the success of the Revised National Tuberculosis Control Programme using rates of case detection – are provided in Appendix A.

**Underreporting**

**The private sector**

Although case notification data are routinely collected in all of India’s districts, the corresponding levels of underreporting – and the geographical variation in those levels – are unknown. The Revised National Tuberculosis Control Programme has nationwide coverage of
the public providers of health care but only limited engagement with the private sector – where at least 50% of tuberculosis patients are estimated to seek treatment.\textsuperscript{13–16} In a 2011 survey in 30 low-performing districts representative of those receiving support from the Global Fund to Fight AIDS, Tuberculosis and Malaria, about half of the members of study households needing tuberculosis treatment went to private providers.\textsuperscript{10} Smaller studies assessing the proportion of cases treated in the private sector in India are discussed in Appendix A. In 2012, India declared tuberculosis a notifiable disease – i.e. it made the reporting of tuberculosis mandatory. However, there remains much scope to improve notifications from private providers, partly because of the generally poor regulation of the private sector.\textsuperscript{17,18} Public–private mix initiatives are interventions intended to educate and engage private providers in diagnosing, reporting and treating tuberculosis in accordance with national guidelines.\textsuperscript{19} Such initiatives have been implemented in India but only currently cover 14 major cities and about 50 million people. This represents the lowest coverage of the 20 countries in which coverage with such initiatives has been reported.\textsuperscript{20} In areas of India covered by these initiatives, 45% of all new smear-positive cases of tuberculosis are reported by private providers.

**Framework to assess underreporting**

In addition to cases not reported by private providers, other factors also contribute to underreporting of tuberculosis cases. In 2002, WHO developed the “onion” model as a framework for assessing the percentage of tuberculosis cases that go unreported (Fig. 4).\textsuperscript{22} The model’s six rings range from cases in people with no access to the health system – representing the sixth and outermost layer of the “onion” – to those diagnosed and reported by providers affiliated to a national tuberculosis programme – representing the first and innermost layer and the only cases captured in case notification data. A variety of data sources can be used to estimate the proportion of cases in each ring. Proportions differ by country because of variations in national health systems.

**Data to assess underreporting**

*Cases outside the health system*

The two outermost rings of the “onion” model correspond to the proportion of the population without access to the health system and the proportion that does not utilize health services even though they have access. Together, these two rings are populated by people who do not seek
treatment from the health system. In India, in the 2007–2009 District Level Household Survey, 2584 (0.34%; 95% confidence interval, CI: 0.33–0.36%) of the 717,692 study households reported that they did not seek any form of medical treatment when their members were sick. The subnational variation in the proportion of cases not accessing the health system can be evaluated from data collected in District Level Household Surveys.

**Underdiagnosis**

Different types of proxy data currently represent the best available options for estimating the proportion of cases in the fourth ring of the onion model – i.e., cases that present to providers but are not correctly diagnosed. Although all undiagnosed cases are combined in the onion model, in India it is useful to distinguish between the undiagnosed cases who present to providers affiliated to the Revised National Tuberculosis Control Programme and those who present to other providers. There is more relevant information available on the Programme-affiliated providers and these providers should have relatively lower rates of underdiagnosis because they are expected to adhere to certain diagnostic algorithms. In general, providers who are not affiliated to the Programme rely heavily on radiology for investigating a potential case of tuberculosis, even though radiology has low specificity and is more appropriate as a screening tool than as a diagnostic test. The Programme’s diagnostic algorithms prescribe the use of sputum-smear microscopy, which has a sensitivity of about 64% and a specificity of approximately 98%. Diagnostic performance can be enhanced by repeating such microscopy for all symptomatic patients who have been initially found smear-negative – a procedure that is also recommended by the Programme. The level of underdiagnosis may be reasonably approximated from Programme data that are released quarterly for each district. These data include the number of suspects examined per smear-positive case diagnosed as well as scores based on performance monitoring – for the district’s case finding efforts. Although WHO has suggested that underdiagnosis might be assessed by evaluating laboratory capacities or the knowledge and practices of health staff, there is sparse literature to support the accuracy or feasibility of these methods, which need further development.

**Inventory studies**

Individuals in the second and third rings of the onion model – i.e., cases that are correctly diagnosed but not reported – can be investigated through inventory studies. The need for such
studies in India was highlighted in the 2011 WHO global tuberculosis report. WHO recently published a guide for countries undertaking inventory studies, which can be used for one of three objectives: to quantify the level of underreporting of diagnosed cases, to estimate tuberculosis incidence by using capture–recapture methods, or to demonstrate that the underreporting of diagnosed cases is negligible. The WHO guide describes four possible designs for an inventory study, with the choice depending on the chosen objective or objectives and the available data (Table 1): a survey of all providers in randomly sampled areas; a survey of all providers in large self-contained areas – with at least two additional case databases; a retrospective analysis with no new data collection; and a survey of a sample of all providers selected using lot-quality assurance sampling. Each of these types of inventory study can be greatly facilitated by an existing national database of diagnosed cases with unique identifiers and standard case definitions. India is currently scaling up its web-based system of case reporting and planning the use of unique identification numbers for all diagnosed cases. However, the limited engagement of the Revised National Tuberculosis Control Programme with the private sector may limit the usefulness of these new initiatives. If individuals in multiple databases are assigned more than one unique identification number, cases may be linked by probabilistic matching. Though error-prone, this has been used successfully in related studies.

For India, given the current status of the national case reporting system, a survey of all providers in randomly sampled areas is the most applicable of the feasible types of inventory study. The capture–recapture analysis using a survey of all providers in large, self-contained areas has fairly stringent data requirements, including at least three independent registries across which record linkage is possible, the right degree of overlap between these registries – ideally 15–30% – and a population with little to no migration and an equal probability of a case being recorded in each registry. These requirements cannot be met at the national level in India at this time. India’s National Tuberculosis Institute is conducting a capture–recapture study in Tumkur district, Karnataka state, and the methods used in that study should be assessed for their possible use at the national level. The methods used in recent capture–recapture studies conducted in other low- and middle-income countries (Appendix A) may also provide indications of the possibility of such studies in India.

Table 2 summarizes the data sources that might be used in India to estimate each component of underreporting. In addition to quantifying the underreporting of diagnosed cases
by all providers, an inventory study would also generate empirical estimates of the proportions of tuberculosis cases that are treated by providers affiliated and not affiliated to the Revised National Tuberculosis Control Programme. Given the limitations of the existing data sources for estimating underdiagnosis in India, the value of any inventory study is likely to be enhanced by the collection of new data on underdiagnosis by all providers (Table 1 and Table 2). The providers followed to assess underreporting could also be evaluated for underdiagnosis, either by tracking the proportion of patients presenting with symptoms indicative of tuberculosis for whom providers order diagnostic tests or using medical vignettes. Although such studies are reliant on the cooperation of the relevant providers, such assistance has been obtained in earlier studies in India.

### National prevalence surveys

At this time there appears to be no widespread governmental support for a national survey of tuberculosis prevalence in India. In an analogous situation less than a decade ago, a national human immunodeficiency virus (HIV) prevalence survey was not recommended for India but, when undertaken, led to a huge adjustment in the estimated HIV burden in India.

### Embedded or stand-alone surveys?

A national survey of tuberculosis prevalence could be embedded within one of the nationwide household surveys periodically conducted in India, which would probably be more cost–effective than a stand-alone prevalence survey. Although existing household surveys cover a small proportion of Indian households, India’s vast population ensures that such surveys provide adequate samples for assessing tuberculosis prevalence, even at subnational levels. For example, the 2005–2006 National Family Health Survey sampled almost 110,000 households. Furthermore, the nationwide household surveys use cluster designs, which are recommended for prevalence surveys. Recent National Family Health Surveys and District Level Household Surveys have collected data on self-reported tuberculosis but these are considered too inaccurate for estimating tuberculosis burdens (Appendix A).

One advantage of a stand-alone survey is that it can be scheduled for the near future. Plans for the next National Family Health Survey, in 2014–2015, are too advanced for the inclusion of a tuberculosis prevalence survey to be considered. The subsequent National Family Health Survey will probably not be implemented until at least 2019. A stand-alone tuberculosis
prevalence survey could also allow tuberculosis-related risk factors and issues of health-care access to be explored more completely than might be feasible in a general national health survey. India’s size and diversity pose challenges for all national surveys; however, successful periodic household surveys on many health topics and a successful nationwide assessment of HIV prevalence indicate the general feasibility of a national survey of tuberculosis prevalence.

**Drug resistance**

The planners of any future nationwide prevalence survey should consider the use of the new Xpert MTB/RIF assay (Cepheid, Sunnyvale, United States of America), which detects both pulmonary and extrapulmonary tuberculosis, provides results in less than 2 h and simultaneously tests for drug resistance. In field tests among smear-negative but culture-positive patients, the assay demonstrated a sensitivity of 77% and a specificity of 99%. Although the assay is easy to use with sputum samples and provides rapid results, it is costly and its use may pose logistical challenges in some settings – at least initially. The current form of the assay only detects resistance to rifampicin but data on the distribution among tuberculosis cases in India of resistance to just this drug may still provide useful insights. However, there has been some recent concern about the assay’s accuracy in detecting rifampicin resistance in India.

A better understanding of the prevalence and distribution of drug-resistant tuberculosis is an emerging priority. Escalating prevalences of multidrug-resistant and extensively drug-resistant tuberculosis are among the greatest challenges to tuberculosis control globally – and India has the greatest number of cases of these forms of tuberculosis. Subnational studies have revealed alarmingly high and increasing prevalences of multidrug-resistant tuberculosis in some areas of India. Any comprehensive investment in the collection of better data for estimating the tuberculosis burden in India should therefore include some support for the evaluation of the role of drug resistance.

**Recommendations**

This discussion outlines several options for the collection of new data to improve estimates of the tuberculosis burden in India. We recommend that both a study of underreporting in the case notification data and a national tuberculosis prevalence survey – possibly including an assessment of drug resistance – be implemented soon.
An inventory study to assess underreporting of tuberculosis cases – based on the new WHO manual for such studies\textsuperscript{28} – should be planned. Although substantial data collection would be required, an inventory study could be relatively short. WHO recommends only three months of follow-up for such a study. We believe a survey of all providers in randomly sampled areas, including an assessment of underdiagnosis, would be the best option because it would generate empirical estimates for the greatest number of relevant parameters. Regardless of the study design used, the sampling of a large number of diverse areas is critical to elucidating any subnational variation in India’s tuberculosis burden.

While logistically more demanding than an inventory study, a prevalence survey would be complementary. If both types of study were implemented, our understanding of tuberculosis epidemiology and control efforts in India would improve further. The costs and benefits of embedded and stand-alone prevalence surveys and the value of a simultaneous assessment of drug resistance should be carefully weighed.

Investments in new data are particularly important for understanding the subnational variation in tuberculosis epidemiology within India’s large population. Improved estimates for India would greatly contribute to a better understanding of the global tuberculosis epidemic. Further development of the methods used to assess underreporting –through their application in India – would also benefit other countries wishing to assess the quality of their case reporting systems. Finally, the need for these studies is timely given the recent goals set by the Indian government in the 2012–2017 Revised National Tuberculosis Control Programme Strategic Plan, which require reliable data for planning and evaluation.

**Acknowledgements**

We thank the Institute for Health Metrics and Evaluation at the University of Washington for sharing the tuberculosis estimates for India from the 2010 Global Burden of Disease Study. LD is also affiliated with the Institute for Health Metrics and Evaluation, University of Washington, Seattle, USA.

**Funding:**

This work was supported by a grant from the Indian Council of Medical Research.

**Competing interests:**

None declared.
References


Table 1. **Possible study designs for estimating the tuberculosis burden in India**

<table>
<thead>
<tr>
<th>Study design</th>
<th>Possible objectives</th>
<th>Existing data used</th>
<th>New data collection required</th>
<th>Current feasibility in India</th>
<th>Application in India</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inventory study</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survey of all providers in sampled areas</td>
<td>Quantification of underreporting of diagnosed cases. Demonstration of negligible underreporting</td>
<td>National tuberculosis surveillance database</td>
<td>Provider survey of all tuberculosis providers in random sample of geographical areas</td>
<td>Feasible for quantifying underreporting of diagnosed cases</td>
<td>Of the feasible studies, relatively inexpensive because fewer data need to be collected</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Proportion of cases with no health system utilization estimated from self-reported household survey data. Level of underdiagnosis estimated from other new data collection or existing data with limitations</td>
</tr>
<tr>
<td>Survey of all providers in large areas suitable for capture-recapture analysis</td>
<td>Quantification of underreporting of diagnosed cases. Estimation of tuberculosis incidence</td>
<td>National tuberculosis surveillance database plus two other case-based databases for each geographical area selected</td>
<td>Provider survey of all tuberculosis providers in random sample of large, self-contained geographical areas</td>
<td>Needs to be assessed</td>
<td>Generates comprehensive, direct estimate of underreporting at all levels</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Assumptions regarding migration and probability of inclusion in each database. Error-prone because of reliance on probabilistic matching across multiple databases</td>
</tr>
<tr>
<td><strong>Retrospective analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retrospective analysis</td>
<td>Quantification of underreporting of diagnosed cases. Estimation of tuberculosis incidence. Demonstration of negligible underreporting</td>
<td>National tuberculosis surveillance database plus one or two national case-based databases – the exact number depending on objectives</td>
<td>None</td>
<td>Not feasible because multiple national case-based databases not available in India</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survey of sample of all providers,</td>
<td>Demonstration of negligible</td>
<td>National tuberculosis</td>
<td>Provider survey of random sample of</td>
<td>Not appropriate because</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survey of all providers with assessment of underdiagnosis</td>
<td>Quantification of underreporting and underdiagnosis by RNTCP and non-RNTCP providers</td>
<td>National tuberculosis surveillance database</td>
<td>Provider survey of all tuberculosis providers in random sample of geographical areas, including assessment of underdiagnosis</td>
<td>Feasible</td>
<td>Generates direct estimates of the greatest number of the parameters contributing to underreporting</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>National prevalence survey</td>
<td>Estimation of national prevalence of active tuberculosis in adults. Assessment of the proportion of tuberculosis cases which are drug-resistant</td>
<td>None</td>
<td>For nationally representative sample of adults aged ≥ 15 years: either Xpert MTB/RIF assay or X-ray screening plus two sputum samples if symptomatic or X-ray abnormal</td>
<td>Feasible</td>
<td>Generates direct estimate of national tuberculosis prevalence, with potential to assess extent of drug resistance</td>
</tr>
</tbody>
</table>

RNTCP: Revised National Tuberculosis Control Programme. * Described in detail in the WHO guide for conducting inventory studies.  

Page 15 of 20
Table 2. **Data needed, in each of the feasible types of inventory study, to estimate underreporting of tuberculosis cases in India**

<table>
<thead>
<tr>
<th>Characteristics of tuberculosis case</th>
<th>Data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Survey of all providers in sampled areas</td>
</tr>
<tr>
<td>Not reported to the case notification system</td>
<td>Household surveys</td>
</tr>
<tr>
<td>No access to health system</td>
<td>Household surveys</td>
</tr>
<tr>
<td>No utilization of health system</td>
<td>Assessment of laboratory capacity or knowledge and practices of health staff</td>
</tr>
<tr>
<td>Using non-RNTCP providers:  Not diagnosed</td>
<td>RNTCP case finding efforts score or the number of patients examined per case</td>
</tr>
<tr>
<td>Diagnosed but not reported Using RNTCP providers:  Not diagnosed</td>
<td>Provider survey</td>
</tr>
<tr>
<td>Diagnosed but not reported</td>
<td>Provider survey</td>
</tr>
<tr>
<td><strong>Diagnosed and reported to the case notification system</strong></td>
<td></td>
</tr>
</tbody>
</table>

RNTCP: Revised National Tuberculosis Control Programme.
Estimates of the mean incidence of tuberculosis, India, 1990–2011

GBD: Global Burden of Disease Study; RNTCP: Revised National Tuberculosis Control Programme; WHO: World Health Organization.

Notes: The estimates from the Global Burden of Disease Study exclude HIV-positive cases. Error bars indicate 95% confidence intervals.
Fig. 2. *Estimates of the mean prevalence of tuberculosis, India, 1990–2011*

![Graph showing the mean prevalence of tuberculosis in India from 1990 to 2011. The graph includes data from the Global Burden of Disease Study (GBD), Revised National Tuberculosis Control Programme (RNTCP), and the World Health Organization (WHO). Error bars indicate 95% confidence intervals.](image)

GBD: Global Burden of Disease Study; RNTCP: Revised National Tuberculosis Control Programme; WHO: World Health Organization.

Notes: The estimates from the Global Burden of Disease Study exclude HIV-positive cases. Error bars indicate 95% confidence intervals.
Fig. 3. **Estimates of the mean level of tuberculosis-attributable mortality, India, 1990–2011**

GBD: Global Burden of Disease Study; RNTCP: Revised National Tuberculosis Control Programme; WHO: World Health Organization.

Notes: Deaths caused by co-infection with HIV and tuberculosis have been excluded. Error bars indicate 95% confidence intervals.
Fig. 4. World Health Organization “Onion” model for assessing the fraction of tuberculosis cases missed by routine notification data

1. Cases recorded in TB notification data

2. Cases diagnosed by the NTP or by providers collaborating with the NTP, but not recorded/reported

3. Cases diagnosed by public or private providers, but not reported

4. Cases presenting to health facilities, but not diagnosed

5. Cases with access to health services that do not go to health facilities

6. Cases with no access to health care

Note: Reproduced, with the permission of the World Health Organization.21