Estimating the Annual Risk of Tuberculosis Infection and Disease in Southeast of Iran Using the Bayesian Mixture Method

Ali Akbar Haghdoost 1; Mahdi Afshari 1, 7; Mohammad Reza Baneshi 1; Mohammad Mehdi Gouya 3; Mahshid Nasehi 3; Mahtab Movahednia 4

Background: Tuberculosis is still a public health concern in Iran. The main challenge in monitoring epidemiological status of tuberculosis is to estimate its incidence accurately.

Objectives: We used a newly developed approach to estimate the incidence of tuberculosis in Sistan, an endemic area in southeast of Iran in 2012-13.

Patients and Methods: This cross-sectional study was conducted on school children aged 6-9 years. We estimated a required sample size of 6350. Study participants were selected using stratified two-stage cluster sampling method and recruited in a tuberculin skin test survey. Indurations were assessed after 72 hours of the injection and their distributions were plotted. Prevalence and annual risk of tuberculosis infection (ARTI) were estimated using the Bayesian mixture model and some traditional methods. The incidence of active disease was calculated using the Markov Chain Monte Carlo technique.

Results: We assumed weibull, normal and normal as the best distributions for indurations due to atypical reactions, BCG (Bacillus Calmette-Guérin) reactions and Mycobacterium tuberculosis infection, respectively. The estimated infection prevalence and ARTI were 3.6% (95%CI: 3.1, 4.1) and 0.48%, respectively. These estimates were lower than those obtained from the traditional methods. The incidence of active tuberculosis was estimated as 107 (87-149) per 100000 population with a CDR of 54% (40%-68%).

Conclusions: Although the mixture model showed slightly lower estimates than the traditional methods, it seems that this method might generate more accurate results for deep exploration of tuberculosis endemicity. Besides, we found that Sistan is a high endemic area for tuberculosis in Iran with a low case detection rate.

Keywords: Tuberculosis; Infection; Survey

1. Background

Despite effective efforts implemented to control tuberculosis (TB), this disease is a leading public health concern in the world (1, 2). Nearly a third of the world’s population is estimated to be infected with Mycobacterium tuberculosis (3). This pathogen is often transmitted through air from patients with pulmonary disease (1) and in some cases from laboratory animals (4). Of these infected population, about 8.7 million (ranged 8.3-9 million) new patients with tuberculosis were detected in 2011 (5). Iran is a developing country in the eastern Mediterranean region with an incidence of 21 tuberculosis patients per 100000 population in 2011 (6). During the last decades, southeast part of this country, especially Sistan area which is close to the border of Afghanistan (country with high burden of tuberculosis), has been the most affected area in Iran (7). A main target within the eight goal of the Millennium Development Goal, 2000, is to "stop the increase and start reversing the incidence of tuberculosis by 2005", i.e. incidence should be halved by 2015 compared to 1990 (8). To monitor the national response and assess the effectiveness of national TB program, it is crucial to accurately estimate the incidence of Mycobacterium tuberculosis (MBT) infection and TB disease and monitor it over the time. TB incidence indicates new TB cases developing active disease during a given period, usually one year (5). The annual risk of tuberculosis infection (ARTI) is one of the most sensitive but hard to measure indicators, which reveals the current epidemiological situation of tuberculosis (9). It is defined as the average probability that a group of individuals acquire new infection with Mycobacterium tuberculosis during one year (10). Generally, it is calculated from the prevalence of MBT infection estimated through tuberculin skin test (TST) surveys. In areas where exposure to non-tuberculosis mycobacteria (NTM) or previous
ous BCG (Bacillus Calmette-Guérin) vaccine is common, interpretation of TST results is challenging (11). Different techniques were proposed to interpret the results of TST surveys to estimate the prevalence of MBT infection. For example, cut-off point method (COP) considering a conditional induration (usually 10 mm) to distinguish infected and noninfected persons (11), antimode method (AM) using the antimode between the two modes of distribution of the indurations as the cut-point and mirror image method (MI) assuming that reaction sizes caused by MBT infection are symmetrically distributed around the second mode of indurations. Therefore, area under this mode is considered as the frequency of MBT infected subjects. The accuracy of such (traditional) methods is still under question (12). Recently, the Bayesian mixture model (BMM) was proposed to improve the interpretation of TST results (9). This technique is used to analyze data of different subgroups when the real subgroup which every individual is belonging to is unknown (13). Actually, this technique assumes that observed distribution of the tuberculin reactions are combination of three components including NTM, reactions due to BCG vaccinations and MBT infection. Therefore, this method tries to estimate parameters to distinguish these components accurately. Our prior knowledge about the characteristics of TST results is combined with our observed data (indurations) and yields an updated probability as well as parameters of the skin reactions following TST. Results of this method, not only estimate the posterior prevalence and distribution of TST skin reactions, but also represent similar information regarding cross reactions due to NTM infection and BCG vaccination (13). Several studies estimated the prevalence and incidence of TB infection and ARTI using the BMM method. According to the results of these studies, estimated ARTI varies between 0.2% to more than 4% in different regions (12, 14). Many TST surveys were conducted in different parts of Iran during the last decade. All of them estimated the prevalence of infection using only cut-off point method. The estimated ARTI in those studies on school children varied between 0.22% and 0.9% (15, 16). The results of an interim analysis of approximately 2000 registered participants of this survey have been reported as an article to compare the estimates of the first three above methods (17). The current study includes more sample size and deeper and more precise analyses, also aims to estimate the prevalence and incidence of MBT infection and active disease.

2. Objectives

Despite the importance of tuberculosis and the efficiency of TB control programs, no accurate studies have been performed to estimate the prevalence and incidence of TB infection and disease in Iran, particularly southeast of this country. Few conducted TST surveys estimated these indicators using traditional methods in different settings. Reported challenges of these methods, make it necessary to use more accurate methods of estimation. The current survey was performed to estimate the annual risk of TB infection and case detection rate using the Bayesian mixture model method.

3. Patients and Methods

3.1. Study Population and Sample Size

This cross-sectional study was performed in Sistan area located in southeast of Iran with a long border with Afghanistan. For decades, this region has been an endemic area for tuberculosis (7) and has more than 400000 inhabitants living in four districts. Our study population was elementary school pupils attending classes 1-3 and aged 6-9 years. Those pupils with high fever at the time of recruitment, history of active tuberculosis, immunocompromising condition, skin rash and those whom parents did not provide informed consent were excluded. To estimate the required sample size, we used the results of previous surveys in other parts of Iran (16). Considering the higher prevalence of tuberculosis in the study area compared to other sub-national regions in Iran, the expected prevalence (p) was assumed to be at minimum of 10%. We used the following formulae to estimate the sample size given the clinical importance level of 14% of the true value, significant level of 5% and design effect (DE) of 3 (11). This estimate was increased by 20% to consider exclusions, absenteeism and dropouts. Finally, the sample size was estimated as 6350 students.

\[ n = \frac{(1.96)^2 \times p \times (1-p) \times DE}{(0.14 \times p)^2} \]

3.2. Ethical Considerations

The study protocol and procedures were reviewed and approved by the Research Review Board of Kerman University of Medical Sciences (K/9/243). After obtaining approval from the local school authorities, we distributed an information brochure in the community and pupils and their parents by local social mobilizers. The brochure included the objective of the study, procedures, potential benefits and harms of participation. Moreover, a passive consent was provided from the parents. In addition, the brochure included some questions about the history of tuberculosis or any other immunocompromising disorders in children.

3.3. Survey Design

We selected the required clusters (schools) using stratified 2-stage cluster sampling according to the WHO recommended sampling techniques (11). Initially, we divided our sample size into rural/urban strata proportional to size method. Then, we selected districts using systematic random sampling allocating equal sample size for districts in each stratum. Finally, 78 schools (25 urban and 53 rural primary schools) were selected and all eligible students studying in classes 1-3 were tested.
3.4. Tuberculin Skin Testing

We trained nine local health workers for two weeks to perform TST and complete the checklists. Before injection, each student was examined for BCG vaccine scar (presence or absence of scar on the usual vaccination site), skin rashes (presence or absence of rashes) and body temperature (temperature more than 39°C using an oral thermometer was considered as fever) by an experienced general physician. Tuberculin skin reactions were assessed after 72 hours of administration of the test. For each child, the maximum transverse diameter of indurations in mm was measured by a flexible transparent ruler calibrated in millimeters after identifying its margins by palpation. To reduce the observer bias, the reader was blinded regarding the status of BCG scar of children. All TST induration results were measured by only one experienced examiner to improve the consistency and minimize inter-observer measurement errors. Details of the testing and reading programs were reported elsewhere (17).

3.5. Statistical Analysis

3.5.1. Bayesian Mixture Model Method

We performed the mixture model analysis assuming different combinations of distributions for mixture components (MBT infection, NTM infection and cross reaction due to BCG vaccination) namely normal, lognormal, and Weibull as suggested in the literature and according to our observed data (10). To conduct the Bayesian techniques, we used a procedure within the empirical Bayes framework (EM algorithm) using R software (version 2.15.2, http://lib.stat.cmu.edu/R/CRAN) (14).

The prior parameters for the three mixture components in the Bayesian technique were estimated according to our observed data assuming different combinations of prior proportions for these components. We assumed that infection prevalence belonging to these components were the same in children with non-zero induration. Then we assumed different proportions for each component and designed mixture models for each assumption. For each model, we estimated the posterior distribution and probability for all mixture components combining the prior distributions with the observed data. Best models were selected using the Akaike information criterion (AIC) indicator.

3.5.2. Annual Risk of Tuberculosis Infection

This indicator was calculated using the following formula:

$$ARTI = (1-p)^{age}$$

Where $p$ stands for the prevalence of MBT infection and age indicates the average age of study children. The mean age was calculated after adding 0.5 year to the age in completed years. This estimated ARTI was the average annual risk of TB infection in the study children from birth to the time of survey, and these children were always exposed to infection, we assumed that this calculated ARTI corresponded closely to the midpoint of the average lives of the participants i.e. 2008-2009.

3.5.3. Tuberculosis Incidence

The incidence of smear positive TB disease was estimated using the Stoblo coefficient (each 1% ARTI corresponds to an incidence of 50 patients per 100000 population). We also calculated some adjusted coefficients by means of parameters reported in other studies from China, Korea and Philippine. These parameters were ratios of the number of infections per prevalent TB cases considering the current situation of TB and HIV influencing the mathematical association between ARTI and incidence of tuberculosis (8).

To estimate TB incidence with reasonable 95% uncertainty level, we applied the Markov Chain Monte Carlo technique to simulate different ratios of infections per prevalent TB cases derived from the literature assuming TB prevalence-incidence ratio compatible with the study region.

The following formula was used for the above purpose:

Incidence = ARTI coefficient

3.5.4. Case Detection Rate (CDR)

First, the expected incidence of tuberculosis for 2012-13 was estimated from what we estimated as TB incidence for 2008-09 (above) assuming 4% decrease in the incidence annually (18) using the exponential function formula:

$$\text{Expected TB incidence in 2012} = \text{TB Incidence in 2009} \times (1-0.04)^3$$

Then, we divided notified incidence of tuberculosis in the study area by the expected incidence to calculate the CDR.

3.5.5. Traditional TB Prevalence Estimation Methods

We also applied some traditional methods to estimate the TB infection prevalence. We considered different indurations (COP method) and anti-mode point of the distribution of the indurations (AM method) as cut-off points to distinguish infected subjects and uninfected ones. In addition, we used the area under the curve symmetrically distributed around the second mode as the number of infected persons (MI method) (9-11). Then, we compared the results with those estimated using BMM method. P value less than 0.05 was considered as statistically significant. All statistical analyses were performed using R software 2.15.2 and STATA SE V.11 software.

4. Results

A total of 6403 school children registered for the survey. Of these, 1013 were excluded from the analysis because of age outlying, absenteeism in the testing and reading days and unwillingness of participants or their parents. Finally, the participation rate was 84%. Of the 5390 satisfactorily test-read children, 97.6% had a BCG scar. On average, they were 7.0 (SD 0.9) years old, 53.8% were female...
and 60.8% living in rural areas (Table 1). In addition, 86.4% (86.3% in pupils with BCG scar and 87.9% without it) had no reaction to TST.

Table 1. Demographic and Clinical Variables of School Children Participated in Tuberculin Skin Test Survey in Southeast of Iran in 2012-13.a,b

<table>
<thead>
<tr>
<th>Clinical and Demographic Variables</th>
<th>Urban</th>
<th>Rural</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BCG Scar</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes (n = 5257)</td>
<td>2038 (38.8)</td>
<td>3219 (61.2)</td>
</tr>
<tr>
<td>No (n = 127)</td>
<td>73 (57.5)</td>
<td>54 (42.5)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (n = 2876)</td>
<td>1229 (42.7)</td>
<td>1647 (57.3)</td>
</tr>
<tr>
<td>Male (n = 2471)</td>
<td>842 (34.1)</td>
<td>1629 (65.9)</td>
</tr>
<tr>
<td><strong>Age, y</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-7 (n = 3589)</td>
<td>1341 (37.4)</td>
<td>2248 (62.6)</td>
</tr>
<tr>
<td>8-9 (1799)</td>
<td>772 (42.9)</td>
<td>1027 (57.1)</td>
</tr>
</tbody>
</table>

a Abbreviation: BCG, bacillus calmette-guérin.
b Data are presented as No. (%).

4.1. Bayesian Mixture Model Method

We observed the best fit for model including weibull-normal-normal distributions for reactions due to NTM, BCG vaccine and MBT, respectively when the prevalence in non-zero indurations were considered as 20%, 50% and 30% for these three components (AIC: -8626). Using BMM method, the prevalence of NTM, BCG vaccine cross-reaction and MBT infection were estimated as 2.3%, 6.8%, and 3.6%, respectively. Moreover, the mean (SD) sizes of indurations for these three components were 2.7 (2.4), 6.1 (3.6) and 16.9 (4.4) millimeters, respectively (Figure 1). The prevalence values of MBT infection were estimated as 4.7% (95%CI, 4.1-5.5), 3.7% (95%CI, 3.1-4.4) and 4.7% (CI95%, 4.1-5.5) when different traditional methods including cut-off point 10 mm, anti-mode 12 mm and mirror image method (second mode 15 mm) applied, respectively. The corresponding ARTIs obtained from these methods were 0.64%, 0.50% and 0.64%, respectively (Table 2).

Table 2. Prevalence and Annual Risk of Tuberculosis Infection Using Different Methods a,b

<table>
<thead>
<tr>
<th>Infection Prevalence (95% CI)</th>
<th>All Children</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cut-Off Point = 10</td>
<td></td>
</tr>
<tr>
<td>Prevalence</td>
<td>4.7 (4.1-5.5)</td>
</tr>
<tr>
<td>ARTI</td>
<td>0.64</td>
</tr>
</tbody>
</table>

a Abbreviation: CI, Confidence Interval; ARTI, Annual Risk of Tuberculosis Infection.
b Data are presented as percentage.
4.2. Incidence of Tuberculosis and Case Detection Rate

Using the ARTI of 0.48% (BMM method) and the Styblo rule, the incidence of smear positive TB in the study area in 2012-13 was estimated as 21 per 100000 for the target population. After applying the correction factors (details in the method sections), the corresponding figure was 107 (95% credibility interval: 87-149) per 100000 population. Case detection rates in 2012-13 using different corrections are presented in Table 3.

<table>
<thead>
<tr>
<th>Method</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Based on Styblo rule (7)</strong></td>
<td></td>
</tr>
<tr>
<td>Incidence</td>
<td>21</td>
</tr>
<tr>
<td>CDR (%)</td>
<td>273</td>
</tr>
<tr>
<td><strong>Based on adjusted rules and assumptions (7)</strong></td>
<td></td>
</tr>
<tr>
<td>Incidence (95% CI)</td>
<td>107 (87-149)</td>
</tr>
<tr>
<td>CDR (95% CI)</td>
<td>54 (40-68)</td>
</tr>
</tbody>
</table>

Abbreviation: CI, credibility interval.

5. Discussion

Our study as one of the few TST surveys in southeast of Iran revealed an ARTI value of about 0.48% using the BMM method, corresponding to an estimated TB incidence of 107 per 100000 population. We also found a relatively high prevalence of intermediate reactions due to nonspecific sensitivity to tuberculin. There are several reasons for challenging the traditional techniques to interpret TST survey results, where the population was exposed preliminary to BCG vaccination or NTM (11). For example, variable specificity in different populations (11) and data reduction of the COP technique and difficulty in identifying visible modes and anti-modes in AM and MI methods could be mentioned (9). Considering the above limitations and complexity in determining the source of indurations (14), we used the Bayesian Mixture method in our study as a suitable and satisfying (11) solution for interpretation of the TST results. In addition to the estimates of MBT infection prevalence, applying BMM method provided information about the prevalence and distribution of other components including cross reactions to tuberculin due to the effect of NTM and BCG vaccination (13). Compared to the traditional methods, BMM provided the lowest estimate of infection prevalence and ARTI, which was very close to the AM method results. Similar to our findings, in TST surveys conducted in Sumatra, Indonesia (vaccinated children) (9) and Zambia (12), the results obtained by the traditional methods were higher than those estimated by the BMM method. However, in TST survey of unvaccinated children in Sumatra (9), and also a survey performed among children living in Java, Indonesia (10), BMM provided higher estimates than the traditional methods. TST surveys in East Nusa Tenggara, Indonesia (10) and South Africa (12) had comparable estimates from traditional methods to BMM technique. A single survey conducted in Spine, reported that the results of BMM method were intermediate between those provided by other methods (14). It seems that it is not possible to easily obtain a standard and global calibration coefficient to change the results of different methods to each other. Because, prevalence rates estimated by different methods differ considerably among various studies, further studies, particularly meta-analysis are needed to explore the heterogeneous factors such as ratio adjustment coefficients. In this study, to accurately estimate the incidence of active smear positive TB disease, we had to apply coefficients and adjustments rather than those suggested in the Styblo rule. This might be due to the difference between Sistan and Styblo study settings regarding TB control programs and HIV status. Using the Monte Carlo Marcov Chain techniques, we concluded that each ARTI in our study corresponded to an incidence of 200-337 cases per 100000 population. Therefore, the incidence of smear positive TB cases in our study region for 2012-2013 was estimated as 107 (87-149) per 100000 population. Considering the survey results and reported incidence of active TB disease in other subnational regions, it seems that Sistan is still the highest affected endemic area for TB in Iran. The incidence was also more than the Eastern Mediterranean regional average (46.5). Among 22 high burden countries, our study region had a higher incidence more than Brazil, China, Russia, Nigeria, Thailand and Tanzania and very similar to Ethiopia, India and Pakistan. Other countries in this group had higher incidences than Sistan area (6). For 2012, we calculated the CDR about 54% (40%-68%). This means that local TB control program was able to only detect about a half of the expected smear positive pulmonary tuberculosis cases. This indicator which reveals the sensitivity of TB control system for diagnosis and supervision, was significantly lower than the level estimated by the World Health Organization (71%) (6). With ongoing activities, it is not expected that Iran could reach the target of the eight Millennium Development Goals (detecting 70% of smear positive TB patients) by 2015 (19). However, these results do not discredit the case finding activities in this region. High notification rate of the total TB cases in Sistan area supports this probability that a relatively high portion of smear positive TB cases have been diagnosed as smear negative or extrapulmonary cases. An explanation for this misdiagnosis of smear positive cases could be due to low sensitivity of routine diagnostic procedures, especially sputum smear laboratory tests. It may also be due to low quality of sputum specimens transferred to lab centers. We should acknowledge the limitation of our study. Most children who were not attending school or tested for TST but not returned for follow-up visit, were from remote
areas. Such subpopulation might be affected more by TB, so our results might underestimate the true TB prevalence and incidence. We tried to include the test results of all absent children as possible. More reliable tests such as Quantiferon-TB (20) could be used in further surveys, but administration of such tests is more difficult compared to TST. The Bayesian mixture model technique needs prior knowledge obtained from previous studies. We had to use adjusted or empirical parameters, because no survey in similar setting was performed in the study area. It is recommended to conduct further TST surveys based on standard protocols in different parts of Iran with regular intervals as basic models to assess the trends of infection. Despite the above limitations, reading the test results by only one person and using standard protocols of TST survey recommended by WHO, decreased many probable sources of errors. In addition, we did not limit our estimations to traditional methods and for the first time, we applied newly developed and more sophisticated statistical methods to estimate the incidence of both infection and disease in southeast part of Iran. In conclusion, even with slightly lower risk obtained by the BMM technique, still Sistan is a high endemic area for TB in Iran with a low smear positive case detection rate. Active case finding using more reliable and sensitive screening tests should be seriously considered and implemented by health authorities in this region. Our results also highlighted the importance of TST cross reactions with non-tuberculosis mycobacteria.

Acknowledgements
We wish to thank the chief and personnel of the Iranian Center for Disease Control, Zabol University of Medical Sciences and Zabol Education Office for their kindly cooperation. The authors declared no conflict of interest.

Author’s Contributions
Study concept and design: Afshari, Haghdoost; Analysis and interpretation of data: Afshari, Baneshi, Haghdoost; Drafting of the manuscript: Afshari, Movahhed Niya; Critical revision of the manuscript for important intellectual content: Haghdoost, Baneshi. Statistical analysis: Afshari, Baneshi, Haghdoost; Administrative, technical, and material support: Afshari, Gouya, Nasehi, Movahhed Niya. Study supervision: Gouya, Nasehi, Afshari, Movahhed Niya.

Funding/Support
This project was supported by Kerman University of Medical Sciences.

References
