A clinical prediction rule to identify patients with tuberculosis at high risk for HIV co-infection

S.K. Sharma, Tamilarasu Kadhiran & Amit Banga

Department of Medicine, All India Institute of Medical Sciences, New Delhi, India

Received April 17, 2008

Background & objective: Many patients presenting with tuberculosis (TB) have underlying human immunodeficiency virus (HIV) co-infection. Routine HIV testing, however, is not a component of the national TB control programme in India. We sought to derive and validate a clinical prediction rule, based on clinical and laboratory parameters, to identify patients at high risk for HIV co-infection among those treated for active TB.

Methods: Case records of adult patients with active TB treated between 1997 and 2003 at the All India Institute of Medical Sciences hospital, New Delhi were retrospectively reviewed. The data set was randomly split into a training set and a testing set. First a clinical prediction rule was derived by multivariable logistic regression on the training set and was subsequently validated on the testing set.

Results: The study group comprised 1074 patients [training set 711 (66%), HIV co-infected 66 (9%); testing set 363 (34%), HIV co-infected 30 (8%)]. In the training set, male gender [odds ratio (95% CI) 5.31(1.52-18.61)], axillary lymphadenopathy [9.71 (3.24-29.10)], anaemia [7.56 (2.48-23.05)], hypoalbuminaemia [3.67(1.31-10.26)], and reduced triceps skinfold thickness [2.91(0.95-8.89)] were independently associated with HIV co-infection. In the testing set, presence of any two of these five features was 94 per cent (95% CI 84-100%) sensitive and 54 per cent (49-60%) specific for predicting HIV co-infection; negative predictive value was 99 per cent (98-100%). Area under the receiver-operating characteristic curve was 0.93 (0.86-1.0) in the testing set.

Interpretation & conclusions: A simple clinical prediction rule based on clinical and laboratory parameters could be used to identify a subgroup of patients, among those treated for active TB in a hospital setting, for targeted HIV testing.

Key words Clinical prediction rule - HIV infection - tuberculosis

Tuberculosis (TB) is the most common opportunistic infection among human immunodeficiency virus (HIV)-infected persons living in developing countries, and is the most common cause of mortality. Incident TB, often, is the index illness leading to a diagnosis of underlying HIV infection in these patients. In fact, the World Health Organization (WHO) recommends that in countries with a generalised or concentrated HIV epidemic, all patients with active TB should be offered HIV testing. Even with optimal anti-TB treatment, HIV co-infected patients with active TB (HIV-TB) are at a higher risk of death and recurrent TB as compared to HIV-negative patients with TB. Apart from early access to antiretroviral treatment, if detected at the time
of presentation with TB, patients with HIV-TB might benefit from interventions such as co-trimoxazole and post-treatment isoniazid preventive therapies10,11.

Recently, the algorithmic approach to the diagnosis of sputum-smear negative TB in HIV-infected individuals has been revised12. In view of this change, in addition to HIV testing of patients diagnosed having TB (i.e., TB cases), the WHO has extended its recommendations for routine HIV testing to include persons with suspected TB (i.e., TB suspects) as well12. However, routine HIV testing is not a component of the national TB control programmes in India and elsewhere, due to operational and cost constraints. Currently, the Revised National TB Control Programme (RNTCP) of India recommends risk-based referral for HIV testing of only those patients with high risk behaviour, other sexually-transmitted infections, or opportunistic conditions suggestive of HIV infection13. However, the RNTCP envisages implementing routine HIV testing for all patients with TB in a phased manner13. In this scenario, the strategy of identifying a subgroup of patients presenting with TB for targeted HIV testing needs to be considered as a bridge approach until the WHO recommendations are operationalised. To demonstrate in principle the feasibility of this strategy, we sought to develop and validate a clinical prediction rule, based on simple clinical and laboratory findings, to identify the patients at high risk for HIV co-infection among those treated for active TB.

Material & Methods

Study population: This study was conducted at the All India Institute of Medical Sciences hospital, New Delhi, a tertiary level teaching hospital catering predominantly to low- and middle-income groups. The catchment area of this hospital includes the National Capital Territory of Delhi and the neighbouring states, mainly Uttar Pradesh, Bihar, Uttarakhand, Jharkhand, and Haryana. All patients aged 13 yr and above, treated for active TB during the period from January 1997 through December 2003, in the outpatient clinics of the Department of Medicine and those hospitalised during this period for the treatment of TB were eligible for inclusion in the study. Patients were excluded if HIV testing was not performed at the time of diagnosis of TB or the result of HIV testing was unavailable.

A diagnosis of TB was made as per criteria described previously14. At the time of diagnosis, details of presenting symptoms, physical findings, anthropometry, radiological features, and laboratory findings were collected using a pre-designed instrument. For the present study, these case records were reviewed retrospectively. The study protocol was reviewed and approved by the Institute Ethics Committee at the All India Institute of Medical Sciences, New Delhi.

HIV testing: After pretest counselling and obtaining informed consent, patients underwent HIV testing at the time of diagnosis of TB. HIV testing was performed using commercially available enzyme-linked immunosorbent assay (ELISA) kits (Detect HIV-1/2, BioChem Immunosystems Inc., Montreal, Canada; UBI HIV 1/2 EIA, Beijing United Biomedical Co. Ltd., Beijing, China) to detect antibodies to HIV-1 and HIV-2 in the serum, as per the WHO strategy II15. Initially, all serum samples were tested using one ELISA. Serum that was non-reactive on the first test was considered HIV-negative. All serum samples found reactive on the first test were retested with a second ELISA based on a different antigen preparation and/or test principle. Serum that was reactive on both tests was reported HIV-positive. In case of discordant results, tests were repeated with the two assays and concordant results after repeat testing were considered final.

Statistical analysis: Statistical analyses were performed using a statistical software package (SPSS for Windows, version 10.0.1, SPSS Inc., Chicago, USA). The study group was split once randomly a posteriori, using a computer-run random selection algorithm, into a training set (two-thirds) and a testing set (one-third). Continuous variables were presented as mean ± standard deviation (SD) or as median [interquartile range (IQR)]. Categorical variables were expressed as numbers with proportions, n (%).

In the training set, we compared the baseline clinical, laboratory, and radiological characteristics of patients with HIV-TB with that of HIV-negative patients with TB. Continuous variables were compared using independent-samples t-test or Mann-Whitney U test, as appropriate. Categorical variables were compared by chi-squared test or Fisher’s exact test. The variables associated with HIV serostatus at P<0.1 significance level on univariate analyses were selected for multivariable analysis. To avoid overfitting16, among the selected variables only those deemed potentially useful for clinical prediction (based on potential availability under field conditions) were chosen for inclusion in the multivariable model. Likewise, only variables having ≥10 outcome-points were considered16.
The chosen variables were entered as covariates to develop a multivariable logistic regression equation, by conditional backward step-wise elimination, with HIV serostatus as the outcome variable. For easy usage, all continuous covariates were dichotomised [using cut-offs identified from receiver-operating characteristic (ROC) curves17] before entering into the equation. Then weights were assigned to the variables retained in the final equation – the assigned weights were equal to their β coefficients rounded to the first decimal. The aggregate of these weighted-variables was expressed as a summary variable (SV), for each patient individually. An ROC curve was plotted with the SV as the test variable and HIV serostatus as the state variable. We identified the optimal cut-off of SV from the ROC curve as the point corresponding to the best trade-off between sensitivity and specificity. This value of SV constituted the prediction rule.

In the testing set, the SV was calculated using the equation derived on the training set. Area under the ROC curve was used to assess the overall predictive performance of the SV in the testing set17. The prediction rule was applied on the testing set, and the corresponding sensitivity, specificity, positive and negative predictive values, and likelihood ratios were calculated using standard methods18. Further, to define the appropriate use of the prediction rule, the posterior probabilities were estimated19 for several hypothetical populations with different levels of HIV seroprevalence (prior probability), ranging from 1 to 50 per cent, among patients with TB.

Results

A total of 1074 patients with TB, treated over the study period of seven years, in whom the result of HIV testing was available, were included. The random split yielded a training set of 711 (66%) patients including 66 (9%) with HIV-TB and a testing set of 363 (34%) patients [HIV-TB 30 (8%) patients]. Characteristics of the patients in the two data sets were comparable except for a slight excess of pulmonary involvement, cervical lymphadenopathy, and miliary TB among patients with HIV-TB in the testing set; the body-mass index, mid-arm circumference, total lymphocyte count, and serum albumin of patients with HIV-TB in the testing set were comparatively lower as compared to that with HIV-TB in the training set (Table I).

In the training set, on univariate analysis, the following variables were associated with the presence of HIV co-infection at P<0.1 significance level: increasing age; male gender; lower body-mass index, mid-arm circumference, and triceps skinfold thickness; absence of pulmonary involvement, miliary disease, and cavity formation on chest radiograph; negative sputum-smear and tuberculin skin test status; presence of axillary, inguinal, mediastinal, and intra-abdominal lymphadenopathy, pleural effusion, and meningitis; and low levels of haemoglobin, total lymphocyte count, and serum albumin (Table I). Of these 19 variables, presence of pulmonary involvement, miliary disease, cavity formation, pleural effusion, meningitis, intra-abdominal and mediastinal lymphadenopathy, sputum-smear status, and tuberculin skin test status were not considered for inclusion in the multivariable model; the remaining 10 variables were included in the multivariable model. In the training set, complete data for these 10 variables were available for 528 (74%) of 711 patients, including 29 (5.5%) with HIV-TB.

After stepwise elimination, five variables remained in the final multivariable equation (Table I). These variables were assigned their corresponding weights (as described under Methods), and the SV was expressed as, SV = (1.7 × male gender) + (2.3 × axillary lymphadenopathy) + (1.1 × triceps skinfold thickness < 8 mm) + (2 × haemoglobin <11 g/dl) + (1.3 × albumin < 3.5 g/dl); all variables were scored as 1 if present and 0 if not. In the training set, the area under the ROC curve was 0.87 (95% CI 0.79-0.95; Fig. A). The optimal trade-off between sensitivity and specificity was at a cut-off of SV ≥ 2.4 (i.e., presence of any two of the five variables in the equation), and the corresponding sensitivity and specificity for predicting HIV co-infection were 90 per cent (95% CI 79-100%) and 49 per cent (45-54%), respectively. This cut-off of SV constituted the prediction rule.

In the testing set, complete data to calculate the SV were available for 286 (79%) patients, including 18 (6%) with HIV-TB. After calculating the SV for these patients, the prediction rule was applied. The rule had a sensitivity of 94 per cent (17 of 18 patients; 95% CI 84-100%) and a specificity of 54 per cent (146 of 268 patients; 49-60%). Corresponding positive and negative predictive values were 12 per cent (17 of 139 patients; 7-18%) and 99 per cent (146 of 147 patients; 98-100%). The positive and negative likelihood ratios were 2.07 (95% CI 1.75-2.46) and 0.10 (0.02-0.69), respectively. The area under the ROC curve in the testing set was 0.93 (95% CI 0.86-1.0; Fig. B).

The effect of varying levels of prior probability of HIV co-infection among patients with TB on the
Table I. Comparison of characteristics between HIV co-infected (TB+ HIV+) and HIV-negative (TB+ HIV-) patients with tuberculosis

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Training data set (n = 711)</th>
<th>Testing data set (n = 363)</th>
<th>Unadjusted odds ratio (95% CI)</th>
<th>Adjusted odds ratio*†† (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TB+ HIV+ n = 645</td>
<td>TB+ HIV+ n = 66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, yr †‡</td>
<td>32 ± 14</td>
<td>34 ± 9</td>
<td>31 ± 13</td>
<td>37 ± 12</td>
</tr>
<tr>
<td>Male gender †</td>
<td>362 (56)</td>
<td>57 (86)</td>
<td>194 (58)</td>
<td>25 (83)</td>
</tr>
<tr>
<td>Duration of symptoms, wk‡</td>
<td>12 (6-52)</td>
<td>12 (3-63)</td>
<td>16 (6-52)</td>
<td>6.5 (4-24)</td>
</tr>
<tr>
<td>Body-mass index, kg/m²‡‡</td>
<td>18.9 ± 3.6</td>
<td>17.8 ± 2.9</td>
<td>18.9 ± 3.8</td>
<td>16.0 ± 2.7</td>
</tr>
<tr>
<td>Mid-arm circumference, cm</td>
<td>22.6 ± 4.1</td>
<td>19.2 ± 4.4</td>
<td>22.9 ± 4.0</td>
<td>17.4 ± 3.3</td>
</tr>
<tr>
<td>Triceps skinfold thickness, mm</td>
<td>8.8 ± 3.3</td>
<td>6.5 ± 1.6</td>
<td>9.1 ± 3.2</td>
<td>6.3 ± 1.7</td>
</tr>
<tr>
<td>Pulmonary involvement ††</td>
<td>447 (69)</td>
<td>26 (39)</td>
<td>234 (70)</td>
<td>18 (60)</td>
</tr>
<tr>
<td>Cavity formation †‡‡</td>
<td>121 (27)</td>
<td>2 (8)</td>
<td>66 (28)</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Positive sputum-smear ††</td>
<td>208 (32)</td>
<td>3 (5)</td>
<td>120 (36)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Lymphadenopathy †††</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cervical</td>
<td>213 (33)</td>
<td>18 (27)</td>
<td>102 (31)</td>
<td>14 (47)</td>
</tr>
<tr>
<td>Axillary</td>
<td>51 (8)</td>
<td>17 (26)</td>
<td>22 (7)</td>
<td>8 (27)</td>
</tr>
<tr>
<td>Inguinal</td>
<td>38 (6)</td>
<td>11 (17)</td>
<td>19 (6)</td>
<td>6 (20)</td>
</tr>
<tr>
<td>Intra-abdominal</td>
<td>30 (5)</td>
<td>10 (15)</td>
<td>21 (6)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Mediastinal</td>
<td>77 (12)</td>
<td>13 (20)</td>
<td>45 (14)</td>
<td>5 (17)</td>
</tr>
<tr>
<td>Pleural effusion †</td>
<td>45 (7)</td>
<td>11 (17)</td>
<td>20 (6)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Ascites ††</td>
<td>23 (4)</td>
<td>2 (3)</td>
<td>7 (2)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Pericardial effusion ††</td>
<td>14 (2)</td>
<td>1 (2)</td>
<td>10 (3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Meningitis †</td>
<td>16 (2)</td>
<td>5 (8)</td>
<td>7 (2)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Miliary tuberculosis †</td>
<td>90 (14)</td>
<td>4 (6)</td>
<td>39 (12)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Haemoglobin, g/dl ††</td>
<td>11.7 ± 2.2</td>
<td>10.1 ± 2.3</td>
<td>11.6 ± 2.4</td>
<td>8.8 ± 2.8</td>
</tr>
<tr>
<td>Total lymphocyte count, cells/μl ††</td>
<td>2454 ± 1132</td>
<td>1850 ± 707</td>
<td>2369 ± 820</td>
<td>1444 ± 974</td>
</tr>
<tr>
<td>Serum albumin, g/dl ††</td>
<td>4.1 ± 0.8</td>
<td>3.1 ± 0.7</td>
<td>4.1 ± 0.8</td>
<td>2.8 ± 0.7</td>
</tr>
<tr>
<td>Positive TST †‡‡</td>
<td>273 (42)</td>
<td>7 (11)</td>
<td>141 (42)</td>
<td>1 (3)</td>
</tr>
</tbody>
</table>

Mid-arm circumference and triceps skinfold thickness were measured at midpoint between the acromion and olecranon processes of right arm using a non-elastic measuring tape and Lange skinfold calipers, respectively; mean of three consecutive readings was taken; anthropometric data were available for 917 (85%) of 1074 patients; data on haemoglobin and serum albumin were available for 996 (93%) and 917 (85%) of 1074 patients, respectively

* = estimated in the training data set; † = only those variables of potential use in field settings with P < 0.1 were included in the multivariable analysis (see Methods); ‡ = data presented as mean ± SD; § = continuous variables were dichotomised for estimating odds ratios, using following cut-offs: age ≥ 28 yr, body-mass index < 19 kg/m², mid-arm circumference < 23 cm, triceps skinfold thickness < 8 mm, haemoglobin < 11 g/dl, total lymphocyte count < 2400/μl, and serum albumin < 3.5 g/dl; ‡‡ = data expressed as number (%) of patients; ¶ = data presented as median (IQR); ** = denominator was no. of patients with pulmonary involvement; †† = lymph node enlargement of at least 1 cm across was considered significant; †‡† = defined as induration ≥ 10 mm in HIV-negative and ≥ 5 mm in HIV co-infected patients; -- = not included in the multivariable model; TST = tuberculin skin test (performed with 5 TU)

The performance of the prediction rule is presented Table II. The prediction rule was found to have a good negative predictive value in excess of 98 per cent in populations with prevalence rates of HIV co-infection ranging 1-10 per cent among patients with TB (Table II).

Discussion

We demonstrated in the present study that a simple clinical prediction rule based on clinical and laboratory parameters could be used to identify a subgroup of patients at high-risk for HIV co-infection among those treated for TB. Earlier, another prediction rule has been derived in patients from Africa for similar purpose. However, the earlier rule was based on radiographic findings, which are usually unavailable in the field setting. In contrast, the current prediction rule is based only on clinical and laboratory parameters that are readily available to the clinician at the time of initial evaluation of patients with TB. Further, it was found that the prediction rule derived in the current study had a good negative predictive value and could thus potentially obviate the need for HIV testing in more than half the patients under resource-poor settings, especially in the context of large scale national TB control programmes.

As shown here, this would hold true particularly in populations with a prevalence of HIV co-infection ranging 1-10 per cent among patients with TB (The estimated prevalence of HIV co-infection among patients in the training data set being close to 10%).
with TB in India is about 1.2%²². However, for obvious reasons, the subgroup of patients identified as high risk by the prediction rule would need further HIV testing for a diagnosis of HIV infection to be made. We would like to stress, where conditions permit, a non-selective testing strategy is the most preferred one for identifying patients with HIV-TB, and a convincing argument can be made in favour of non-selective testing over that of a targeted testing strategy. Notwithstanding, in reality only about 4 per cent of patients with TB in India get tested for HIV co-infection.²²

As all consecutive patients with TB were included, there was no obvious selection bias. However, being a hospital-based study, the study population is likely to differ from that seen in primary care settings, where most of the patients with TB receive their treatment. It is possible that when compared to the primary care setting, patients might present in a tertiary care setting late during the course of illness and with comparatively advanced disease. But, this delay is likely to have been there in both the groups namely HIV-TB as well as HIV-negative patients with TB, and it is unclear how this delay would influence the performance of the prediction rule. In the present study, we sought to demonstrate only the feasibility of identifying a subgroup of patients with TB at high risk for HIV co-infection. It needs to be emphasised that the findings of the present study cannot be directly extrapolated to primary care settings. This approach has to be first validated specifically in a primary care setting before being incorporated into practice.

Overall, cavitary pulmonary disease, sputum-smear positivity, and tuberculin skin test positivity were less common among patients with HIV-TB, and these findings were in consonance with the published literature.²³ On the contrary, miliary TB was found to be less common among patients with HIV co-infection in the present study. It is likely that due to the severity of the underlying disease, many of these patients succumbed to their illness before they could seek medical attention.

The factors found to be independently associated with HIV co-infection in the present study were reflective of either the risk factor for HIV infection (male gender) or the consequences of HIV infection (malnutrition and anaemia). It is well known that in India, unlike the African countries, HIV infection is more common among men than women. Similarly, it is known that weight loss and malnutrition are more pronounced in patients with HIV-TB.²⁴ Also, a progressive decline in haemoglobin level occurs as HIV disease advances.²⁵ In the present study, presence of axillary lymphadenopathy was found to be associated with HIV infection. An earlier study from Africa had made a similar observation.²⁶ Generalised non-specific

---

**Fig.** Empirical receiver-operating characteristic curve for the summary variable to predict HIV co-infection in the training data set (A) and the testing data set (B). (1-specificity) represents the rate of false-positive error.
immune activation occurs in patients with HIV-TB\textsuperscript{23}, and it probably accounts for this observed association.

Certain methodological issues merit attention. First, a positive Western blot assay is considered the gold standard for the confirmation of HIV infection. Diagnosis of HIV infection was made using sequential ELISA, in the present study. However, this is the strategy recommended by the WHO for symptomatic patients with suspected HIV infection\textsuperscript{15}, and this strategy has been validated and found to have a positive predictive value similar to Western blot-based testing\textsuperscript{27}. Second, we did not evaluate the predictive value of clinical clues such as a history of high risk sexual behaviour and presence of mucosal candidiasis, since these were considered as compelling indications on their own to offer HIV testing to a patient with TB. Third, in the derivation of the multivariable model, we dichotomised the continuous variables before inclusion in the model. This could have resulted in a modest loss of statistical power to detect a relation between the variables and the outcome\textsuperscript{28}. However, dichotomisation was necessary to facilitate easy use of the prediction rule under clinical settings.

Finally, since the study was a retrospective analysis, validation was done by the split-group method. This represents a narrow validation of the prediction rule, and in the hierarchy of evidence proposed by the Evidence-Based Medicine Working Group\textsuperscript{29}, this would be considered Level 4 evidence. As pointed out earlier, it needs to be validated prospectively and independently in broader clinical settings, specifically in populations where it is intended to be used. The cut-offs for the various parameters in the prediction rule might need to be adjusted to suit the population of interest, so as to achieve the best predictive performance.

In conclusion, male gender, presence of axillary lymphadenopathy, anaemia, hypoalbuminemia, and reduced triceps skinfold thickness were associated with the presence of underlying HIV co-infection among patients with TB presenting to hospital setting. A simple clinical prediction rule based on these findings (presence of any two of these features) had a good negative predictive value and could be used to identify a subgroup of patients with TB for targeted HIV testing in resource-poor settings where universal access to HIV testing remains limited.

Acknowledgment

One of the authors (TK) was supported by the Scientists’ Pool Scheme of the Council of Scientific & Industrial Research (CSIR), New Delhi, Government of India.

References

9. Sonnenberg P, Murray J, Glynn JR, Shearer S, Kambashi B, Godfrey-Faussett P. HIV-1 and recurrence, relapse, and


