A simple, valid, numerical score for grading chest x-ray severity in adult smear-positive pulmonary tuberculosis

Anna P Ralph,1,2 Muhamed Ardian,3,4,5 Andri Wiguna,4,5 Graeme P Maguire,6 Niels G Becker,2 Glen Drogumuller,7 Michael J Wilks,8 Govert Waramori,4 Emiliana Tjitra,9 Sandjaja,9 Enny Kenagalem,3,10 Gysje J Pontororing,10 Nicholas M Anstey,1,11 Paul M Kelly1,2

ABSTRACT

Background The grading of radiological severity in clinical trials in tuberculosis (TB) remains unstandardised. The aim of this study was to generate and validate a numerical score for grading chest x-ray (CXR) severity and predicting response to treatment in adults with smear-positive pulmonary TB.

Methods At a TB clinic in Papua, Indonesia, serial CXRs were performed at diagnosis, 2 and 6 months in 115 adults with smear-positive pulmonary TB. Radiographic findings predictive of 2-month sputum microscopy status were used to generate a score. The validity of the score was then assessed in a second data set of 139 comparable adults with TB, recruited 4 years later at the same site. Relationships between the CXR score and other measures of TB severity were examined.

Results The estimated proportion of lung affected and presence of cavitation, but not cavity size or other radiological findings, significantly predicted outcome and were combined to derive a score given by percentage of lung affected plus 40 if cavitation was present. As well as predicting 2-month outcome, scores were significantly associated with sputum smear grade at diagnosis (p<0.001), body mass index, lung function, haemoglobin, exercise tolerance and quality of life (p<0.02 for each). In the validation data set, baseline CXR score predicted 2-month smear status significantly more accurately than did the proportion of lung affected alone. In both data sets, CXR scores decreased over time (p<0.001).

Conclusion This simple, validated method for grading CXR severity in adults with smear-positive pulmonary TB correlates with baseline clinical and microbiological severity and response to treatment, and is suitable for use in clinical trials.

INTRODUCTION

Sputum smear microscopy, and culture where available, are standardised modalities for diagnosing and monitoring treatment response in pulmonary tuberculosis (TB). Chest radiography (CXR) provides useful information regarding disease extent and progress, but there is no agreed-upon, validated system for grading the severity of CXR abnormalities in bacteriologically proven pulmonary TB. Several methods were devised for this purpose at the time of early TB treatment trials, such as those described by the Madras TB Chemotherapy Centre in 1960,1 Simon in 19662 and the National TB and Respiratory Disease Association of the USA in 1969.3

Despite this, no system has been validated in predicting outcome in more than one patient population. Recent randomised controlled trials (RCTs)4–7 and observational studies in adults with TB8–13 illustrate this lack of standardisation in the grading of radiological severity, with each of these studies utilising different non-validated investigator-generated systems to grade CXR severity.

The same problem of non-standardised radiological reporting has been recently articulated by Dawson et al relating to TB screening, who evaluated and recommended the Chest Radiographic Reading and Reporting System14 for TB screening in HIV-positive people.15 However, this and other screening tools16–18 seek to identify the presence of latent TB infection or active disease, and are not useful for researchers wishing to accurately document severity or response to treatment in active TB.

Problems in CXR reporting arise from the heterogeneous CXR manifestations of pulmonary TB (eg, in primary vs postprimary disease, adults vs children, immunocompetent vs immunocompromised)19–21 and to inaccuracies inherent in CXR performance and interpretation,7 including limited interobserver agreement on CXR findings.22–23 Despite these shortcomings, the utility of CXR is well established in TB diagnosis and clinical monitoring.

Associations between radiological extent and other measures such as forced expiratory volume in 1 s (FEV1), age or multidrug-resistant (MDR)-TB have previously been identified,2 8 24 but a standard, simple, numerical score, validated against TB outcome, in repeated data sets, is lacking. We therefore aimed to devise a simple CXR score for use in adults with smear-positive pulmonary TB, which predicts outcome and correlates with bacteriological and clinical severity markers, for the purpose of grading severity and monitoring treatment response in the context of TB clinical trials. We then determined the utility of the score in a separate, comparable patient population.

METHODS

Study setting

The study was conducted at a community-based TB clinic in Timika, Papua Province, Indonesia. Timika has population of ~200,000 and an estimated TB incidence of 511/100,000.25
Tuberculosis

Participants
Adults (>15 years) diagnosed with sputum smear-positive pulmonary TB who gave written informed consent were eligible for enrolment in the study. Study participants were recruited during two time periods: 2003–2004 (the derivation data set) and 2008–2009 (the validation data set). The demographic, clinical and microbiological findings and outcomes in the first data set have been reported previously.26 27

Chest radiography
Standard full-size posteroanterior CXR were performed at the time of TB diagnosis and 2 and 6 months thereafter, with reports provided by a clinician at the field site (first data set, PMK; second data set, APR) and, additionally for the first data set, by one of two radiologists (MJW or GD). During the first data collection period, the presence of small (1–2 mm) or large (>2 mm) nodules, patchy or confluent consolidation, cavitation, bronchial lesions or fibrosis was reported for each of three zones (upper, mid or lower zones) in each lung. The presence of effusion or lymphadenopathy was reported, the total percentage of each lung affected by any pathology was estimated, total cavity size in millimetres was recorded and the effusion volume (percentage of lung field) was estimated. To grade the percentage of affected lung, visual estimation of the extent of opacification, cavitation or other pathology as a percentage of visible lung was made; dense opacification of a zone was graded as 100% of that zone, while patchy opacification within a zone attracted scores <100% depending on the extent of opacification. Other remarks including presence of miliary disease were recorded. During the second data collection period, a simplified CXR report method was used (percentage lung affected, cavitation (0, <4 cm, ≥4 cm), effusion (0, <25%, ≥25% of hemithorax), presence of consolidation, fibrosis, nodules, miliary disease). Reporters were blinded to HIV status, bacteriological and clinical parameters and treatment outcome.

Sputum microscopy and clinical evaluations
Baseline sputum microscopy was performed at the onsite laboratory and repeated at the reference laboratory on samples collected at 0, 2 and 6 months, and the density of acid-fast bacilli (AFB) was graded as 1, 2 or 3+ according to standard protocols.26 27 Baseline and follow-up evaluations included: body mass index (BMI), FEV1 (spirometry performed using ML3535C, MicroLoop, MicroMedical, Chatham, UK), haemoglobin (Hb), measured using point-of-care HemoCue (Angelholm, Sweden) or iSTAT (Abbott Park, Illinois, USA) tests, 6 min walk test (distance walked in 6 min on a straight walking track), measured according to American Thoracic Society guidelines, and St George’s Respiratory Questionnaire (SGRQ) modified to reflect local conditions and translated into Indonesian.27 28 Standard definitions were used for nutritional category (normal, mild malnutrition, moderate malnutrition or severe malnutrition) according to BMI,29 and for TB treatment outcome at 6 months (cured, completed, transferred, defaulted, failed or died).30 Impairment in FEV1 as a percentage of predicted values was calculated using previously established local reference ranges.31

Outcome measure
The outcome measure used in this study is 2-month sputum AFB microscopy status. Two-month smear positivity has been previously shown to predict unfavourable outcomes including treatment failure and death,32–34 and determines the need for continued intensive-phase treatment versus switching to continuation-phase therapy.30 Although an imperfect predictor of outcome,35 in the absence of suitable alternatives, it remains a commonly used surrogate end point.

Data analysis
Statistical calculations were performed using Intercooled Stata 10.1 (StataCorp, College Station, Texas, USA); graphs were created in GraphPad Prism 5 (GraphPad, La Jolla, California, USA). Statistical tests were two sided, with a p value of <0.05 indicating statistical significance. Intergroup differences in means or medians were compared using two-sample t tests, Wilcoxon rank sum tests, analysis of variance or Kruskal–Wallis tests as appropriate.

Agreement between reporters in the derivation data set was tested using the concordance coefficients, p for continuous variables or the kappa statistic for categorical variables. Prevalence-adjusted, bias-adjusted kappa values were calculated according to the method described by Byrt et al.36 Kappa values were interpreted according to guidelines given by Landis and Koch.37 (kappa ≤0.00, poor; 0.00–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; 0.81–1.00, almost perfect).

The relationships between radiographic findings and clinical outcome were examined by multivariable regression analysis, using a forward stepwise approach in which any radiological variable found to be significant (p<0.05) in univariate analysis was included in the initial model. Goodness of fit of final models was assessed using the Hosmer–Lemeshow test and compared using the likelihood ratio test. The weighting for a numerical radiological score was derived from the regression coefficients. Its ability to predict outcome in the validation data set was determined using receiver operator characteristics (ROC; area under the curve (AUC)). The relationships between this score and demographic, biological and clinical variables were determined in data sets 1 and 2 using regression models using the same principles.

Ethics
Approval was granted by the ethics committees of the National Institute of Health Research and Development (Jakarta, Indonesia), Menzies School of Health Research (Darwin, Australia) and the Australian National University (Canberra, Australia). Written informed consent was obtained from participants in Indonesian or an appropriate Papuan language.

RESULTS
Characteristics of study participants in the two data collection phases are shown in table 1. All participants had smear-positive pulmonary TB (≥2 AFB smear-positive sputum samples); the result of an additional sample provided for microscopy and culture on the day of treatment commencement is reported here. This was negative in 5.7% and 7.2% of participants in the two data sets, respectively despite their previous samples being positive. Initial smear grade predicted the likelihood of smear conversion by 2 months. In the derivation data set, failure to convert to smear negative by 2 months was observed in 60.9% of patients with a baseline smear grade of 3 and in 38.7% of patients with a baseline smear grade of <3+ (p=0.051). In the validation data set, failure to convert to smear negative by 2 months was observed in 48.4% of patients with a baseline smear grade of 3 and in 11.8% of patients with a baseline smear grade of <3+ (p<0.001).

CXR reports were available at baseline, 2 and 6 months for 112, 76 and 76 study participants in the first data set, and 136, 93 and 76 study participants in the second data set (incomplete in the second data set as 30 of 139 had not yet completed

had worse lung function, with a mean percentage predicted FEV₁ of 59.0 (95% CI 54.4 to 63.6) in caviary disease versus 68.7 (95% CI 60.6 to 76.7) in non-caviary disease (p=0.03, two-sample t test). People with caviary disease had slightly lower BMI (18.5, 95% CI 18.0 to 19.1) compared with those with non-caviary disease (19.2 kg/m², 95% CI 18.5 to 20.0), but this difference was not statistically significantly (mean between-group difference 0.70 kg/m², 95% CI —0.15 to 1.59). No significant associations were identified between caviary disease and exercise tolerance (6 min walk distance), quality of life (SGRQ total or individual domain scores) or Hb.

The amount (%) of lung affected significantly predicted all clinical and laboratory variables. Specifically, greater proportions of affected lung were significantly associated with decreasing BMI category (p=0.002, Kruskal—Wallis test), lung function category

### Table 1 Study participant characteristics

<table>
<thead>
<tr>
<th>Demographic details</th>
<th>Derivation data set</th>
<th>Validation data set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>115</td>
<td>139</td>
</tr>
<tr>
<td>Age in years: median (range)</td>
<td>30 (17—69)</td>
<td>27 (15—65)</td>
</tr>
<tr>
<td>Female gender, n (%)</td>
<td>33 (28.7)</td>
<td>46 (34.5)</td>
</tr>
<tr>
<td>Papuan ethnicity, n (%)</td>
<td>57 (48.6)</td>
<td>66 (47.5)</td>
</tr>
<tr>
<td>Smokers, n (%)</td>
<td>38 (33.0)</td>
<td>41 (29.5)</td>
</tr>
<tr>
<td>HIV positive: no./no. tested</td>
<td>5/112 (4.5)</td>
<td>16/21 (13.2)</td>
</tr>
<tr>
<td>MDR-TB, n (%)</td>
<td>2 (1.7)</td>
<td>2 (1.4)</td>
</tr>
<tr>
<td>Baseline clinical findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI, median (range), kg/m²</td>
<td>18.6 (14.2—25.2)</td>
<td>19.0 (12.9—32.5)</td>
</tr>
<tr>
<td>Haemoglobin, median (range), g/dl</td>
<td>11.2 (6.8—18.0)</td>
<td>12.2 (7.1—16.0)</td>
</tr>
<tr>
<td>FEV₁: median (range), litres</td>
<td>1.76 (0.49—4.12)</td>
<td>1.70 (0.59—3.56)</td>
</tr>
<tr>
<td>SGRQ total score, median (range)</td>
<td>45.3 (2.5—83.5)</td>
<td>37.6 (5.2—91.9)</td>
</tr>
<tr>
<td>6 min walk distance, median (range)</td>
<td>405 (185—625)</td>
<td>410 (20—612)</td>
</tr>
<tr>
<td>Sputum AFB smear grade at diagnosis n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0*</td>
<td>6 (5.7)</td>
<td>10 (7.2)</td>
</tr>
<tr>
<td>Scanty or 1+</td>
<td>25 (23.6)</td>
<td>65 (45.7)</td>
</tr>
<tr>
<td>2+</td>
<td>28 (26.4)</td>
<td>35 (25.2)</td>
</tr>
<tr>
<td>3+</td>
<td>47 (44.3)</td>
<td>29 (20.9)</td>
</tr>
<tr>
<td>2-month smear status n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>25 (21.8)</td>
<td>31 (22.3)</td>
</tr>
<tr>
<td>Negative</td>
<td>81 (70.4)</td>
<td>95 (68.4)</td>
</tr>
<tr>
<td>No result available</td>
<td>9 (7.8)</td>
<td>13 (9.3)</td>
</tr>
<tr>
<td>6-month outcome n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cured/completed</td>
<td>88 (76.5)</td>
<td>92 (66.2)</td>
</tr>
<tr>
<td>Died</td>
<td>3 (2.6)</td>
<td>2 (1.4)</td>
</tr>
<tr>
<td>Failed</td>
<td>2 (1.7)</td>
<td>1 (0.7)</td>
</tr>
<tr>
<td>Default</td>
<td>13 (11.3)</td>
<td>8 (5.8)</td>
</tr>
<tr>
<td>Transferred</td>
<td>9 (7.8)</td>
<td>6 (4.3)</td>
</tr>
<tr>
<td>6 months not yet completed</td>
<td>0</td>
<td>30 (21.6)</td>
</tr>
</tbody>
</table>

*All study participants had at least two prior smear-positive sputum samples; some are reported as negative since this result pertains to the additional spot specimen provided at enrolment into the study (see the methods section). AFB, acid-fast bacilli; BMI, body mass index; FEV₁, forced expiratory volume in 1 s; MDR-TB, multidrug-resistant tuberculosis; SGRQ, St George’s Respiratory Questionnaire.

#### Table 2 Chest radiograph (CXR) results

<table>
<thead>
<tr>
<th>Number of CXR available for report</th>
<th>Derivation data set</th>
<th>Validation data set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>112</td>
<td>136</td>
</tr>
<tr>
<td>2 months</td>
<td>76</td>
<td>93</td>
</tr>
<tr>
<td>6 months</td>
<td>76</td>
<td>76*</td>
</tr>
<tr>
<td>Baseline radiological findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consolidation, n (%)</td>
<td>106 (94.6)</td>
<td>104 (76.5)</td>
</tr>
<tr>
<td>Cavitation, n (%)</td>
<td>71 (63.4)</td>
<td>77 (56.6)</td>
</tr>
<tr>
<td>Effusion, n (%)</td>
<td>37 (33.0)</td>
<td>24 (17.3)</td>
</tr>
<tr>
<td>Miliary, n (%)</td>
<td>2 (1.7)</td>
<td>5 (3.7)</td>
</tr>
<tr>
<td>Nodules, n (%)</td>
<td>83 (72.2)</td>
<td>29 (21.3)</td>
</tr>
<tr>
<td>Fibrosis, n (%)</td>
<td>62 (55.3)</td>
<td>10 (7.4)</td>
</tr>
<tr>
<td>Proportion of lung affected, median (range)</td>
<td>35 (0—100)</td>
<td>41.5 (0—100)</td>
</tr>
<tr>
<td><strong>CXR weighted score characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline score, median (range)</td>
<td>65 (5—130)</td>
<td>69 (0—140)</td>
</tr>
<tr>
<td>2 month score, median (range)</td>
<td>35 (0—130)</td>
<td>29 (0—140)</td>
</tr>
<tr>
<td>6 month score, median (range)</td>
<td>10 (0—105)</td>
<td>10 (0—115)</td>
</tr>
<tr>
<td>Median score decrease in 2 months, %</td>
<td>27.3</td>
<td>38.8</td>
</tr>
<tr>
<td>Median score decrease in 6 months, %</td>
<td>73.1</td>
<td>80.2</td>
</tr>
<tr>
<td>Baseline score by smear grade at diagnosis median (range)</td>
<td>0 (0.5—70)</td>
<td>10 (0—133)</td>
</tr>
<tr>
<td>Scanty or 1+</td>
<td>50 (0—110)</td>
<td>66 (0—140)</td>
</tr>
<tr>
<td>2+</td>
<td>70 (6.5—135)</td>
<td>76 (11—140)</td>
</tr>
<tr>
<td>3+</td>
<td>80 (10—140)</td>
<td>85 (4—140)</td>
</tr>
</tbody>
</table>

*30 patients still under follow up.

#### Table 3 Agreement on radiological findings, derivation data set

<table>
<thead>
<tr>
<th>Inter-rater agreement among dichotomous variables</th>
<th>Kappa</th>
<th>Prevalence-adjusted, bias-adjusted kappa</th>
<th>Interpretation of prevalence-adjusted, bias-adjusted kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of patchy consolidation</td>
<td>0.20</td>
<td>0.70</td>
<td>Substantial</td>
</tr>
<tr>
<td>Presence of confluent consolidation</td>
<td>0.33</td>
<td>0.31</td>
<td>Fair</td>
</tr>
<tr>
<td>Presence of any consolidation</td>
<td>0.24</td>
<td>0.81</td>
<td>Almost perfect</td>
</tr>
<tr>
<td>Presence of small nodules</td>
<td>0.12</td>
<td>0.06</td>
<td>Slight</td>
</tr>
<tr>
<td>Presence of large nodules</td>
<td>0.09</td>
<td>0.09</td>
<td>Slight</td>
</tr>
<tr>
<td>Presence of any nodules</td>
<td>0.19</td>
<td>0.19</td>
<td>Slight</td>
</tr>
<tr>
<td>Presence of fibrosis</td>
<td>0.08</td>
<td>0.06</td>
<td>Slight</td>
</tr>
<tr>
<td>Presence of cavitation</td>
<td>0.33</td>
<td>0.37</td>
<td>Fair</td>
</tr>
<tr>
<td>Presence of effusion</td>
<td>-0.35</td>
<td>-0.61</td>
<td>Poor</td>
</tr>
<tr>
<td>Presence of lymphadenopathy</td>
<td>0.01</td>
<td>-0.09</td>
<td>Poor</td>
</tr>
</tbody>
</table>

#### Table 3 Agreement on radiological findings, derivation data set

<table>
<thead>
<tr>
<th>Concordance among continuous variables</th>
<th>ρ₀</th>
<th>95% limits of agreement (Bland and Altman)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total amount (%) of lung affected</td>
<td>0.85</td>
<td>0.20 to 0.90</td>
</tr>
<tr>
<td>Extent of cavity size (mm)</td>
<td>0.89</td>
<td>0.20 to 0.90</td>
</tr>
</tbody>
</table>
(p<0.001, Kruskal–Wallis test), 6 min walk distance (0=0.001, linear regression) and Hb in males (p=0.003, linear regression), though not in females (p=0.4). A greater proportion of lung affected on the baseline CXR was also significantly associated with SGRQ total scores (p<0.001, linear regression) and with sputum smear grade at diagnosis (p<0.001, Kruskal–Wallis test).

To create a CXR score, cavitation and percentage of lung affected were included as independent variables in a logistic regression model for 2-month sputum smear status. The model containing both variables was significantly better than the model containing cavitation alone (likelihood ratio test p<0.001) or proportion of lung affected alone (likelihood ratio test p=0.016) at predicting 2-month sputum smear status. Regression coefficients were 0.03167 for proportion of lung affected and 1.26151 for presence/absence of cavitation, indicating a relative weighting of 40.27 for cavitation (1.26151/0.03167), thereby generating an equation for the weighted score as follows:

\[
\text{CXR score} = \text{proportion of total lung affected(%) + 40 if cavitation present}
\]

**CXR score results**

CXR score characteristics are shown in table 2 and figures 1–4. Scores did not significantly differ according to sex, ethnicity or smoking status (p>0.05, two-sample t tests), and were not significantly associated with age in univariate or multivariate analyses. Mean baseline CXR score in people with unfavourable (positive) 2-month outcomes was significantly higher (88.2; 95% CI 76.5 to 99.9) than in those with a favourable outcome (56.8; 95% CI 49.7 to 64.0), but the range of scores in each smear grade was wide (figure 1). Scores were also significantly associated with baseline microscopy grade (figure 1). CXR scores were inversely related to BMI, FEV₁, Hb and 6 min walk distance, were directly related to SGRQ total score (higher SGRQ scores indicate worse quality of life) and significantly decreased over time (figures 2–4).

**Performance of score using the validation data set**

The weighted score calculated for the new data set showed similar characteristics (table 2), including a median baseline score of 69, no significant relationship with demographic factors, significant positive association with baseline smear grade (p=0.009, Kruskal–Wallis test) and the same relationships as

<table>
<thead>
<tr>
<th>Radiological parameters</th>
<th>Odds ratio</th>
<th>95% CI</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of any consolidation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per 1% increments</td>
<td>1.30</td>
<td>0.14 to 12.18</td>
<td>0.001</td>
</tr>
<tr>
<td>Per 20% increments</td>
<td>1.90</td>
<td>1.30 to 2.68</td>
<td>0.001</td>
</tr>
<tr>
<td>Cavitary disease</td>
<td>3.26</td>
<td>1.11 to 9.56</td>
<td>0.032</td>
</tr>
<tr>
<td>Effusion</td>
<td>1.50</td>
<td>0.59 to 3.82</td>
<td>0.580</td>
</tr>
<tr>
<td>Nodules</td>
<td>1.17</td>
<td>0.41 to 3.32</td>
<td>0.773</td>
</tr>
<tr>
<td>Fibrosis</td>
<td>2.23</td>
<td>0.86 to 5.79</td>
<td>0.097</td>
</tr>
<tr>
<td>Lymphadenopathy</td>
<td>1.43</td>
<td>0.57 to 3.56</td>
<td>0.448</td>
</tr>
</tbody>
</table>

*OR was unable to be calculated for the two instances of military disease (both had negative sputum at 2 months).

**Table 4 Relationship between radiological and biological parameters and 2-month sputum acid-fast bacilli density in the derivation data set, showing results of univariate logistic regression analyses*"

*All p values calculated from one-way analyses of variance. Horizontal bars represent mean score.

**Figure 1** Chest x-ray (CXR) score at enrolment according to sputum acid-fast bacilli (AFB) grade at diagnosis and 2 months (derivation data set).

**Figure 2** Chest x-ray (CXR) score according to body mass index (BMI), percentage of predicted forced expiratory volume in 1 s (FEV₁) and haemoglobin (Hb) at diagnosis (derivation data set).

**Figure 3** Chest x-ray (CXR) score according to St George’s Respiratory Questionnaire (SGRQ) (total score and 6 minute walk distance at tuberculosis diagnosis (derivation data set).

*Both p values calculated using one-way analyses of variance. Horizontal bars represent mean score.
were found in the initial data set between CXR score and each of the clinical/laboratory measures (BMI, Hb, FEV1, SGRQ total score and 6 min walk distance; \( p < 0.05 \) in each case).

Comparing ROC scores to predict outcome, the weighted CXR score (AUC 0.75) was significantly better at predicting 2-month smear status than the percentage lung affected alone (ROC 0.69; \( p = 0.013 \), \( \chi^2 \) test; figure 5). The optimal cut-off point for weighted CXR score (value furthest from the diagonal) was 71, at which value the sensitivity for predicting a positive sputum smear status at 2 months was 80% (95% CI 61.4 to 92.3) and specificity 67.7% (95% CI 57.5 to 77.1). Comparative sensitivity and specificity values are shown in table 5.

**DISCUSSION**

The current need for a universal and standard system for reporting CXR in pulmonary TB is acknowledged.\(^{38}\) In order to grade CXR severity and assess radiological treatment response, we have derived a simple equation from radiographic parameters from adults with smear-positive pulmonary TB that predicts smear positivity at 2 months and provides a single numerical score for each CXR. The score shows good correlation with baseline bacteriological and clinical severity markers, and is sensitive to changes over time. The score performs better than its individual components: it was significantly better at predicting outcome than was the percentage of lung affected alone, and was significantly associated with a broader range of baseline severity measures (BMI, |Hb, exercise tolerance and quality of life) than presence of cavitation alone. Advantages of this method are that CXR assessment does not require aids, grids or rulers, and it is derived by fitting a statistical model to outcome data rather than by assigning points based on assumed relative importance of radiographic pathologies. It has been

![Figure 4](image-url) Chest x-ray (CXR) score at diagnosis, 8 weeks (end of intensive treatment phase) and 24 weeks (end of treatment).

![Figure 5](image-url) Receiver operator characteristics (ROC) for weighted chest x-ray (CXR) score (% lung affected + 40 if cavitation present) (solid line) compared with percentage of lung affected (dashed line) as predictors of 2-month smear status (positive or negative). Optimal cut-off point for score (circle) = 71.

![Table 5](image-url) Sensitivity and specificity of baseline radiological and bacteriological findings in predicting outcome (2-month sputum acid-fast bacilli (AFB) status)

<table>
<thead>
<tr>
<th>Radiological parameters at diagnosis</th>
<th>Derivation data set</th>
<th>Validation data set</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sensitivity</strong></td>
<td><strong>Specificity</strong></td>
<td></td>
</tr>
<tr>
<td>Total amount of lung affected</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>96</td>
<td>35.9</td>
</tr>
<tr>
<td>37.5% (optimal cut-off point)</td>
<td>84</td>
<td>68.0</td>
</tr>
<tr>
<td>80%</td>
<td>20</td>
<td>89.7</td>
</tr>
<tr>
<td>Cavitiation</td>
<td>80</td>
<td>44.9</td>
</tr>
<tr>
<td>Score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>18.0</td>
</tr>
<tr>
<td>72.5 (optimal cut-off point)</td>
<td>80</td>
<td>71.2</td>
</tr>
<tr>
<td>130</td>
<td>8</td>
<td>98.7</td>
</tr>
<tr>
<td><strong>Bacteriological parameters: baseline smear grade</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Derivation data set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFB density*</td>
<td>60.9</td>
<td>61.3</td>
</tr>
<tr>
<td>Validation data set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFB density*</td>
<td>48.4</td>
<td>88.2</td>
</tr>
</tbody>
</table>

*Baseline AFB density (smear grade) dichotomised as \( \leq 2 \) or 3.
validated in an independent data set, and offers a single, stand-
ardised solution where there are currently multiple unvalidated
methods in use.

The proportion of lung affected and/or cavitation feature as
the most important measures in many TB CXR grading meth-
ods. 1–5 7 Cavitation is well recognised to correlate with bacillary
load. 7, 39 We confirmed the association between cavitation and
bacteriological measures (baseline and 2-month sputum smear
status), and additionally showed cavitary disease to be predictive
of worse lung function. The proportion of lung affected was
associated with both bacteriological and a range of clinical
measures.

This score was derived in adult patients with TB with smear-
positive pulmonary disease, in a setting with relatively low rates
of HIV–TB co-infection and MDR-TB. The score requires
further evaluation in populations with high HIV prevalence, in
whom CXR findings characteristic of HIV–TB co-infection (sub-
ilinear or absent pathology, non-cavitary disease, lower lobe
infiltrates, hilar lymphadenopathy and pleural effusion) 20 40 may
mean that a differently weighted score is needed. Nevertheless,
the score remained valid and applicable in the newer data set in
which HIV–TB co-infection rates were higher (15%); the rise in
HIV prevalence may account for some of the differences
observed between the two data sets. The presence of MDR-TB
would not be expected to alter radiographic patterns, other than
being associated potentially with higher scores and smaller
incremental improvements over time.

Potential limitations of the study include the use of 2 month
smear status as an outcome measure (rather than a longer term
measure such as 6-month outcome or recurrence).

The absence of suitable biomarkers or other surrogate end
points in TB research is readily acknowledged, and recent
estimates derived from meta-analysis found a sensitivity of only
57% and specificity of 81% for 2-month smear status in
predicting treatment failure. 35 Nevertheless, until more suitable
measures become available, 2-month smear status remains a
suitable outcome measure. 32–34

Another limitation was the inherent problem of limited inter-rater agreement in CXR assessment. The low rates of
clinician–radiologist agreement between reporters on CXR
findings identified in the derivation data set are not unusual, with
only fair or poor agreement between radiologists and clinicians
also being reported elsewhere. 22 23 This emphasises the impor-
tance of using simple rather than complex scores and ensuring
individuals allocating CXR scores participate in continuing
education to maximise agreement. The score derived from radi-
ologist CXR evaluation in the first data set is simple. Moreover,
It was shown to be valid in the second data set when used by an
independent TB clinician, rather than a radiologist, confirming its
practical utility in a clinical and trial setting. Some systematic
differences in CXR results were noted between the two data sets;
while this may represent systematic difference in reporting
styles, the findings are in keeping with the possibility of less
severe disease in the validation data set, as indicated by their
lower bacillary burden (with slides read by the same senior
laboratory technician during both data collection periods).

CONCLUSION

In summary, we have derived and validated a simple method for
grading CXR severity in adults with smear-positive pulmonary
TB that predicts baseline clinical and microbiological severity
and response to treatment in two separate patient populations.
Although finer discriminatory accuracy might be achieved by
collecting more detailed CXR findings (such as cavity size), our
data did not indicate this. This method can be used where a
numerical score is required for the purpose of comparing
radiographic severity between adults with smear-positive
pulmonary TB, and to monitor an individual’s improvement
over time, such as in clinical trials of drug efficacy in TB.

Acknowledgements We thank the following for their support and assistance: Dr M Okosoyer, Pak Penias and Pak E Meokbun and the Timika District Health
Authority; Dr Dina Bisara Loloeng and Ibui Meryani Pirang and the National Institute of
Health Research and Development, Jakarta; Dr P Penfinten, Dr M Bangs and
Dr M Stone, Public Health & Malaria Control (PHMC) and International SDS; Pak
Istanto and PHMC laboratory staff; Pak J Lempy and Timika TB clinic staff,
Dr P. Sugianto and Mimika Community Hospital (RSMNM); Natalia Dwi Haryanti,
Sri Hasmunik, Sri Rahayu, G Bellatrix and clinical and laboratory staff, NIHRD-MSHR
Timika research programme; Mr R Lumb and Dr I Bastan at the Institute of Medical
and Veterinary Science; and Associate Professor R Price, MSHR.

Funding Australian Respiratory Council, the Royal Australasian College of Physicians
(Covance award), Australian National Health and Medical Research Council.

Competing interests None.

Ethics approval This study was conducted with the approval of the Human Research
Ethics Committees of the NT Department of Health & Families and Menzies School of
Health Research, Australia, the Australian National University, and the National
Institute for Health Research and Development, Indonesia.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Tuberculosis Chemotherapy Centre Madras. A concurrent comparison of
isoniazid plus PAS with three regimens of isoniazid alone in the domiciliary treat-
ment of pulmonary tuberculosis in South India. Bull World Health Or 1960;23:
355–95.
2. Simon G. Radiology in epidemiological studies and some therapeutic trials. BMJ
3. Falk A, O’Connor J, Pratt P. Classification of pulmonary tuberculosis. In: Falk A,
O’Connor JB, Pratt PC, et al, eds. Diagnostic standards and classification of
first 2 months of treatment for pulmonary tuberculosis. Am J Respir Crit Care Med
isoniazid adjuvant therapy for multidrug-resistant tuberculosis. Int J Tuberc Lung
a week versus rifampicin and isoniazid twice a week for treatment of drug-
susceptible pulmonary tuberculosis in HIV-negative patients: a randomised clinical
HIV-infected tuberculosis patients in Thailand. Southeast Asian J Trop Med Public
pulmonary tuberculosis with and without HIV infection in Nigeria. Scand J Infect Dis
11. Idd J, Westman A, Elise D, et al. Nitric oxide production in the exhaled air of
patients with pulmonary tuberculosis in relation to HIV co-infection. BMC Infect Dis
2008;8:146.
12. Thorson A, Long NH, Larsson LO. Chest X-ray findings in relation to gender
and symptoms: a study of patients with smear positive tuberculosis in Vietnam.
can chest radiograph reading and recording system for epidemiological surveys of
system: evaluation for tuberculosis screening in patients with advanced HIV.
16. Gatner EM, Bukhardt KR. Correlation of the results of X-ray and sputum culture in
17. Linh NN, Marks GB, Crawford AB. Radiographic predictors of subsequent
18. World Health Organization. Assessing tuberculosis prevalence through
population-based surveys. WHO, Western Pacific Region 2007 (October 2009).
A simple, valid, numerical score for grading chest x-ray severity in adult smear-positive pulmonary tuberculosis

Anna P Ralph, Muhamed Ardian, Andri Wiguna, Graeme P Maguire, Niels G Becker, Glen Drogumuller, Michael J Wilks, Govert Waramori, Emiliana Tjitra, Sandjaja, Enny Kenagalem, Gysje J Pontororing, Nicholas M Anstey and Paul M Kelly

Thorax 2010 65: 863-869
doi: 10.1136/thx.2010.136242

Updated information and services can be found at:
http://thorax.bmj.com/content/65/10/863

These include:
References
This article cites 36 articles, 4 of which you can access for free at:
http://thorax.bmj.com/content/65/10/863#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections
Radiology (diagnostics) (812)
TB and other respiratory infections (1273)
Tuberculosis (51)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/