Multidrug-resistant tuberculosis treatment failure detection depends on monitoring interval and microbiological method

Carole D. Mitnick1,2, Richard A. White3,4, Chunling Lu1,5, Carly A. Rodriguez1, Jaime Bayona6, Mercedes C. Becerra1,2,5, Marcos Burgos7, Rosella Centis8, Theodore Cohen9, Helen Cox10, Lia D’Ambrosio8,11, Manfred Danilovitz12, Dennis Falzon13, Irina Y. Gelmanova14, Maria T. Gler15, Jennifer A. Grinsdale16, Timothy H. Holtz17, Salmaan Keshavjee1,2,5, Vaira Leimane18, Dick Menzies19, Giovanni Battista Migliori8, Meredith B. Milstein1,20, Sergey P. Mishustin21, Marcello Pagano22, Maria I. Quelapio23, Karen Shean24, Sonya S. Shin1,2,5,6, Arielle W. Tolman1, Martha L. van der Walt25, Armand Van Deun26 and Piret Viiklepp27, on behalf of the Collaborative Group for Analysis of Bacteriology Data in MDR-TB Treatment28

ABSTRACT Debate persists about monitoring method (culture or smear) and interval (monthly or less frequently) during treatment for multidrug-resistant tuberculosis (MDR-TB). We analysed existing data and estimated the effect of monitoring strategies on timing of failure detection.

We identified studies reporting microbiological response to MDR-TB treatment and solicited individual patient data from authors. Frailty survival models were used to estimate pooled relative risk of failure detection in the last 12 months of treatment; hazard of failure using monthly culture was the reference.

Data were obtained for 5410 patients across 12 observational studies. During the last 12 months of treatment, failure detection occurred in a median of 3 months by monthly culture; failure detection was delayed by 2, 7, and 9 months relying on bimonthly culture, monthly smear and bimonthly smear, respectively. Risk (95% CI) of failure detection delay resulting from monthly smear relative to culture is 0.38 (0.34–0.42) for all patients and 0.33 (0.25–0.42) for HIV-co-infected patients.

Failure detection is delayed by reducing the sensitivity and frequency of the monitoring method. Monthly monitoring of sputum cultures from patients receiving MDR-TB treatment is recommended. Expanded laboratory capacity is needed for high-quality culture, and for smear microscopy and rapid molecular tests.

@ERSpublications

Monthly culture monitoring is crucial to earlier detection of treatment failure in MDR-TB patients http://ow.ly/w2MI301mK8M

This article has supplementary material available from erj.ersjournals.com

Support statement: Funding for this study was provided in part by Global Tuberculosis Programme of the World Health Organization, through a grant from the United States Agency for International Development. Funding for data gathering at participating centres was provided as follows: in Peru from Thomas J. White, the Bill and Melinda Gates Foundation, David Rockefeller Center for Latin American Studies at Harvard University and from the US National Institutes of Health (K01 A1065836 (C.D. Mitnick) and K01 HL080939 (M.C. Becerra)); in the State of California from the Centers for Disease Control and Prevention Cooperative Agreement Funds; and in South Africa from the South African Medical Research Council funding. D. Menzies was supported by salary awards from the Fonds de Recherche en Santé de Québec; and G.B. Migliori and R. Centis were funded by the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement FP7-223681. M.L. van der Walt was funded by the South African Medical Research Council. Funding information for this article has been deposited with the Open Funder Registry.

Conflict of interest: Disclosures can be found alongside this article at erj.ersjournals.com

This is one of a selection of articles published as ERJ Open papers, as part of an initiative agreed between the European Respiratory Society and the World Health Organization.

The content of this work is copyright the authors or their employers. Design and branding are copyright ©ERS 2016. This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0.
Affiliations: 1Dept of Global Health and Social Medicine, Harvard Medical School, Boston, MA, USA. 2Partners in Health, Boston, MA, USA. 3Dept of Infectious Disease Epidemiology, Norwegian Institute of Public Health, Oslo, Norway. 4Dept of Health Statistics, Norwegian Institute of Public Health, Oslo, Norway. 5Brigham and Women’s Hospital, Division of Global Health Equity, Boston, MA, USA. 6Socios en Salud Sucursala, Lima, Peru. 7Dept of Internal Medicine, University of New Mexico School of Medicine, Albuquerque, NM, USA. 8WHO Collaborating Centre for TB and Lung Diseases, Fondazione S. Maugeri, Care and Research Institute, Tradate, Italy. 9Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, USA. 10Médecins Sans Frontières, Cape Town, South Africa. 11Public Health Consulting Group, Lugano, Switzerland. 12Tartu University Hospital, Lung Clinic, Tartu, Estonia. 13Global TB Programme, World Health Organization, Geneva, Switzerland. 14Partners in Health Russia, Moscow, Russia. 15Makati Medical Center, Manila, Philippines. 16Tuberculosis Control, San Francisco Department of Public Health, San Francisco, CA, USA. 17US Centers for Disease Control and Prevention, Atlanta, GA, USA. 18State Agency for Tuberculosis and Lung Diseases (TPSVA), Riga, Latvia. 19Montreal Chest Institute, McGill University, Montreal, QC, Canada. 20Bouvé College of Health Sciences, Northeastern University, Boston, MA, USA. 21Tomsk TB and Pulmonary Medical Center, Tomsk Oblast, Russia. 22Dept of Biostatistics, Harvard School of Public Health, Boston, MA, USA. 23Tropical Disease Foundation, Philippine Institute of Tuberculosis, Makati City, Philippines. 24Lung Infection and Immunity Unit, Division of Pulmonology and UCT Lung Institute, Dept of Medicine, University of Cape Town, Cape Town, South Africa. 25South African Medical Research Council, Tuberculosis Research Platform, Pretoria, South Africa. 26Institute of Tropical Medicine, Antwerp, Belgium. 27National Institute for Heath Development, Tallinn, Estonia. 28Members of the Collaborative Group for Analysis of Bacteriology Data in MDR-TB Treatment are listed in the acknowledgement section.

Correspondence: Carole D. Mitnick, Dept of Global Health and Social Medicine, Harvard Medical School, 641 Huntington Avenue, 3rd Floor, Boston, MA 02115, USA. E-mail: carole_mitnick@hms.harvard.edu

Introduction
In 2013, 20% of the 480,000 multidrug-resistant tuberculosis (MDR-TB) cases estimated among notified pulmonary TB cases received appropriate second-line treatment [1]. Myriad challenges hamper attempts to provide universal access to MDR-TB treatment: lengthy, toxic treatment; inadequate supply of high-quality drugs; limited human resources; complex adverse event management; and a dearth of laboratory resources to diagnose MDR-TB and monitor treatment response [2]. Early in MDR-TB treatment, monthly culture and smear monitoring of treatment response are recommended; more frequent monitoring is recommended for patients with HIV disease and other comorbidities [3]. After sputum culture conversion [4], which occurs ~3 months after treatment initiation [5], recommendations are for less frequent culture with monthly smear examination. Historically, this has meant at least quarterly culture with monthly smear for the last 12 months of an 18–24-month regimen [3].

Costs of the different monitoring methods are variable and overlapping, ranging from $1.63 to $62.01 for culture and $0.26 to $10.50 for smear [6]. Key to selection of frequency and method of monitoring is information on their performance characteristics. A recent meta-analysis found that both smear and culture have low sensitivity and modest specificity for predicting relapse in drug-susceptible TB [7]. Additional reports establish a relationship between earlier culture conversion and successful treatment outcome [8, 9], and detection of initial conversion and reversion is delayed with less frequent monitoring during MDR-TB treatment [9].

The present report investigates the effect of monitoring interval (monthly versus bimonthly or quarterly) and method (smear versus culture) on timing of treatment failure detection during the final 12 months of treatment delivered under routine programme conditions. Preliminary results of this investigation, which used an individual patient-data meta-analysis, informed the 2011 update of the World Health Organization (WHO) Guidelines for the Programmatic Management of Drug-resistant Tuberculosis [10, 11] and are updated and published here in their entirety.

Methods
Study selection
The present study extends the work of the Collaborative Group for Meta-analysis of Individual Patient Data in MDR-TB [12–14]. Articles eligible for the present analysis were included in either of two published meta-analyses of MDR-TB treatment [15, 16] and in AHUJA et al.’s [12] 2012 meta-analysis of individual patient data. Additional sources were identified by WHO guidelines committee members. E-mail correspondence was used to request datasets, data dictionaries, definitions and clarification; in some cases, individual patient data on more or different patients from those included in the original analysis were provided. Summary information was used to assess initial eligibility. Exclusion criteria were duplicate datasets; nonprimary research; language other than English, French or Spanish; data comprising only extensivly drug-resistant TB patient populations; and no response after three or more requests to
participate. Included studies used second-line drugs in reported treatment; had a sample size >25 patients; and had monthly culture data.

Compilation of the parent dataset has been described previously [12, 13, 17]. The study was exempted from review by Harvard Medical School's committee on human subjects research. Data-use agreements executed with all participating projects stipulated that those who provide data will seek approval or exemption from local institutional review boards as appropriate.

**Analysis**

The following were evaluated among all patients in the included studies: 1) does timing of failure detection during the final 12 months of treatment differ across monitoring frequencies and methods? and 2) is this difference modified by patient characteristics?

All analyses were performed using Stata 12 (StataCorp, College Station, TX, USA).

The end-point was time to detection of bacteriological failure, classified as two or more positive cultures (or smears) in the last 12 months of treatment (failure month was the month of the second positive result) or at least one positive culture (or smear) in the last 3 months of treatment (failure month was the month of the first positive result), according to the consensus definitions in place when most primary data were collected [18]. For one included study [19], in which a large proportion of patients received a 9-month treatment regimen, failure was defined as a single positive culture among the last three cultures. Data handling conventions favoured failure: if multiple results were available in a single month, any positive result was used. In addition, missing observations were imputed by carrying the last value forward. Lastly, when there were missing observations among the last three expected observations, the analysis retrieved prior observations until there were three. All increased the probability of including positive cultures (or smears), and detecting failure. Consequently, failure percentages may be higher than those previously reported.

We simulated three monitoring scenarios against which we compared the timing of “observed failure” or failure detected by monthly culture in the final 12 months: one using only monthly smear; the second using results from samples collected every other month (to simulate bimonthly monitoring) and the third using results from samples collected every third month (to simulate quarterly monitoring). We estimated the timing of 1) failure detected by monthly smear only (primary analysis); 2) failure detected by bimonthly culture or smear; and 3) failure detected by quarterly culture or smear and compared each to the timing of observed failure. For each culture result, any positive smear within 14 days (either side) was elected as the corresponding smear result.

Unweighted pooled frequencies were calculated for descriptive statistics. Monthly concordance between smear and culture results during the whole treatment was estimated by calculating the mean proportion (with 95% confidence intervals) of concordant intra-patient results in each month.

Kaplan–Meier estimates of failure time in the last 12 months were calculated using standard techniques. Heterogeneity was evaluated using $I^2$. To evaluate the validity of pooled results in the presence of heterogeneity, we performed an additional analysis that included intercepts and interaction terms for each cohort, allowing us to obtain 12 realisations of the heterogeneous outcome (and corresponding error estimates). We then generated 1000 hypothetical sets of 12 realisations of the outcome, and calculated the mean and standard error for each of the 1000 sets. By estimating the means of these two statistics we obtained an estimate of the heterogeneity of the outcome across different cohorts. Alternative monitoring strategies were compared to monthly culture using frailty survival models (random intercept) to estimate the relative hazard of failure, pooled over all sites. Effect estimates were also calculated separately for each study site. Reported hazard ratios reflect the risk of failure detection: a hazard ratio of <1 indicates delayed detection, while a hazard ratio of >1 indicates earlier detection. We assessed heterogeneity in patient characteristics through stratification by study site, HIV serostatus, baseline smear result, radiography findings, low body mass index ($<$18.5 kg·m$^{-2}$). Multivariable (i.e. adjusted) regressions were considered and rejected as high rates of missing data resulted in unstable complete case analyses. We performed sensitivity analysis of the time to failure detection excluding patients with missing baseline bacteriology.

**Results**

**Study selection and population**

Search strategy and results are presented in figure 1. Review of the meta-analyses yielded 30 possible datasets. 21 additional potential data sources were identified by the guidelines committee and through reference lists. Five more were identified through the US Centers for Disease Control and Prevention (CDC)-sponsored case-based data collection efforts. Out of 56 summaries/abstracts evaluated for eligibility, nine were excluded. The remaining datasets and full articles were reviewed; 12 were eligible for inclusion and 35 were excluded for the following reasons: four studies did not use second-line drugs; one study...
reported on fewer than 25 patients; nine authors did not respond to requests for data; and smear and culture data were missing or insufficient in 21 studies (figure 1). In total, 12 datasets with 5730 unique patient treatment records were included; seven were from six separate published articles and five from the case-based data collection coordinated by the CDC, previously described [5, 9]. Table 1 illustrates the distribution of 12 study sites across 11 countries. Of note, four studies were conducted in Eastern Europe. At least one on-treatment culture result was available for 5410 (94.4%) patients with confirmed MDR-TB. In just over 20%, baseline smear (1117) and culture (1119) status was unrecorded, but subsequent monitoring data were available. These 5410 patients were included in analyses.

Approximately one-third (33.5%) of patients were female. Mean age ranged from 31.1 to 48.0 years (pooled estimate 37.0 years). The availability of data on and distribution of cavitary, bilateral disease varied. Among those with drug-susceptibility test results for the five first-line drugs, most had isolates resistant to four or five (91%). Second-line drug susceptibility test results were available in only 43.7% of observations and are not reported here. Data on HIV serostatus, body mass index and chest radiography findings were missing in 37.2%, 59.0% and 40.0% of observations, respectively (table 2).

Treatment outcomes, calculated according to consensus definitions [17] varied across studies: cure frequencies ranged from 16.3% at site 6 to 74.1% at site 12. Failure was observed in 16.9% (914) of cases (table 3).

**Concordance of monthly smear and culture results**

We evaluated the intra-observation agreement of smear and culture during the first 24 months of treatment, stratified by treatment outcome (died, cured/completed or failed). Summary results are presented in table 3 and figure 2. As smear and culture conversion occurred, concordance increased. Uncertainty around the estimates increased later in follow-up as fewer patients remained on treatment and contributed observations (online supplementary material). Mean agreement across months was greatest for those who were cured by or completed treatment (95.6%, 95% CI 85.7–100.0%). The type of discordance varied: among deaths and failures, smear-negative/culture-positive pairs were more common. Among cures, smear-positive/culture-negative pairs were more common. However, the latter occurred roughly as frequently among cures as among deaths and failures.
The timing of failure detection in the last 12 months of treatment varied within and across the monitoring strategies (figure 3). In the last 12 months of treatment monthly culture detected the first positive result that contributed to the failure definition in a median (interquartile range (IQR)) 3 (0.8–7.0) months; by bimonthly culture and quarterly culture, median detection occurred in 5 (2.0–10.0) months and 6 (2.0–12.0) months, respectively. Failure was detected by monthly smear in 10 (2.0–12.0) months. The rate of failure detection is lower for monthly smear relative to monthly culture (HR (95% CI) 0.38 (0.34–0.42)) and for quarterly culture compared to monthly culture (0.58 (0.53–0.64)).

Stratified analysis of failure detection in the last 12 months comparing monthly smear to monthly culture revealed further variability. First, detection was delayed by using smear alone at seven sites (1, 3, 6, 8–10 and 12), while no significant change in risk could be discerned elsewhere (figure 4a). The rate of failure detection by exclusive monthly smear monitoring was lower in patients with baseline negative sputum smears (HR (95% CI) 0.17 (0.13–0.23)) than in patients with baseline positive smears (0.46 (0.41–0.53)). No difference was observed in failure detection by HIV status; in both HIV-negative and HIV-positive patients, monthly smear delayed failure detection compared to monthly culture (HIV− 0.42 (0.36–0.50), HIV+ 0.33 (0.25–0.42); figure 4b).

There was significant heterogeneity ($I^2$ 83.5%, $p<0.001$) in the pooled analyses of the difference between smear and culture. Nevertheless, the sensitivity analysis that was stratified by site produced results similar to those observed in the survival frailty model. Excluding patients with missing baseline smear and culture results did not change the effect estimates. The high proportion of missing data for several covariates resulted in small numbers in complete case analyses; these are not presented.

### Discussion
This study combined individual patient data from multiple sources to investigate differences in time to failure detection in the last 12 months of MDR-TB treatment resulting from reducing the frequency or changing the method of bacteriological monitoring. Failure was detected most promptly with monthly culture. At any given time point during the 12-month evaluation period, quarterly culture was half as likely as monthly culture (and monthly smear was only 38% as likely as monthly culture) to detect failure. Earlier detection of treatment nonresponse may enhance the ability to adjust treatment regimens to avert poor outcomes, decrease the probability of death and reduce transmission due to uncontrolled disease [11]. The present results do not support changing the frequency or method of monitoring in the last
<table>
<thead>
<tr>
<th>Site/dataset</th>
<th>Site location (and dates when &gt;1 cohort/location)</th>
<th>Subjects n#</th>
<th>Females</th>
<th>Age years mean (n)</th>
<th>Known HIV*</th>
<th>BMI ≤18.5 kg·m⁻²</th>
<th>Cavitation, bilateral findings on radiography %</th>
<th>First-line resistance %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>San Francisco, CA, USA</td>
<td>55</td>
<td>45.3 (53)</td>
<td>45.7 (53)</td>
<td>15.2 (46)</td>
<td>0</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>Uzbekistan</td>
<td>77</td>
<td>39.0 (77)</td>
<td>36.8 (77)</td>
<td>0.0 (77)</td>
<td>71.4 (77)</td>
<td>28.6</td>
<td>21.1</td>
</tr>
<tr>
<td>3</td>
<td>South Africa (2000–2004)</td>
<td>2092</td>
<td>37.7 (2089)</td>
<td>36.5 (2070)</td>
<td>64.4 (1326)</td>
<td>59.8 (455) &amp; n &amp;</td>
<td>1425</td>
<td>27.4</td>
</tr>
<tr>
<td>4</td>
<td>Estonia (2004–2008)</td>
<td>274</td>
<td>27.7 (274)</td>
<td>48.0 (274)</td>
<td>4.1 (268)</td>
<td>70</td>
<td>24.3</td>
<td>77</td>
</tr>
<tr>
<td>5</td>
<td>Italy</td>
<td>71</td>
<td>39.4 (71)</td>
<td>37.8 (71)</td>
<td>7.2 (69)</td>
<td>158</td>
<td>48.1</td>
<td>274</td>
</tr>
<tr>
<td>6</td>
<td>South Africa (1990–1999)</td>
<td>319</td>
<td>37.3 (319)</td>
<td>33.9 (251)</td>
<td>4.6 (109)</td>
<td>51</td>
<td>23.5</td>
<td>32.5</td>
</tr>
<tr>
<td>7</td>
<td>Bangladesh</td>
<td>668</td>
<td>28.0 (668)</td>
<td>34.2 (668)</td>
<td>83.7 (325)</td>
<td>653</td>
<td>28.8</td>
<td>67.5</td>
</tr>
<tr>
<td>8</td>
<td>Estonia (2001–2005)</td>
<td>292</td>
<td>28.1 (292)</td>
<td>44.3 (292)</td>
<td>3.4 (262)</td>
<td>18.6 (161)</td>
<td>282</td>
<td>16.3</td>
</tr>
<tr>
<td>9</td>
<td>Latvia</td>
<td>446</td>
<td>22.6 (446)</td>
<td>42.4 (446)</td>
<td>3.3 (360)</td>
<td>20.4 (343)</td>
<td>282</td>
<td>16.3</td>
</tr>
<tr>
<td>10</td>
<td>Lima, Peru</td>
<td>712</td>
<td>40.3 (712)</td>
<td>31.1 (712)</td>
<td>1.3 (631)</td>
<td>84.4 (343)</td>
<td>579</td>
<td>16.3</td>
</tr>
<tr>
<td>11</td>
<td>Manila, Philippines</td>
<td>161</td>
<td>36.0 (161)</td>
<td>39.6 (161)</td>
<td>0.0 (9)</td>
<td>42.0 (81)</td>
<td>31</td>
<td>16.3</td>
</tr>
<tr>
<td>12</td>
<td>Tomsk, Russia</td>
<td>243</td>
<td>13.6 (243)</td>
<td>34.4 (243)</td>
<td>0.0 (242)</td>
<td>19.3 (243)</td>
<td>135</td>
<td>16.3</td>
</tr>
<tr>
<td>Pooled frequency</td>
<td>5410</td>
<td>33.5 (5405)</td>
<td>37.0 (5318)</td>
<td>26.8 (3399)</td>
<td>41.7 (2216)</td>
<td>3229</td>
<td>32.1</td>
<td>2369</td>
</tr>
</tbody>
</table>

Data are presented as % [n] or %, unless otherwise stated. BMI: body mass index; HR: isoniazid–rifampicin. #: refers to the number of patients in each cohort, with some culture information; otherwise, values of n indicate the number of patients for whom data were available on each covariate.
12 months of treatment. These main results helped to inform the WHO 2011 revised treatment guidelines for the programmatic management of MDR-TB [10].

These findings highlight the particular value of performing at-least monthly cultures throughout treatment among subgroups. In particular, in patients with smear-negative TB at baseline, culture was even more critical to detecting failure during the final 12 months, since positive cultures often were detected before positive smears. Although not surprising, these findings lend evidence to programmatic decision-making for this subgroup. HIV-infected patients are another subgroup in whom these results suggest continued use of monthly culture monitoring. Although there was no effect modification on delay of failure detection by HIV serostatus, sputum-smear microscopy is known to have significantly decreased sensitivity in HIV-co-infected TB patients [25–27]. In addition, delays in appropriate treatment can have more serious consequences in HIV-positive patients than among HIV-negative patients [28, 29]. Therefore, it would be reasonable to employ monitoring strategies that can detect and permit intervention in at-risk patients at the earliest possible juncture. This reaffirms previous recommendations to monitor these patients more aggressively throughout treatment [3].

Additional benefits of early failure detection must be considered in evaluation of monitoring costs. In absolute terms, work by Lu et al. [6] estimated that culture costs between 1.4 and 12.0 times more than smear, with significant variability. Monthly culture would increase these differences by three-fold compared to quarterly culture. However, patients affected by MDR-TB are often in the most productive age group (mean age 37.0 years), both as workers and care providers to children and parents. The cost of

### TABLE 3 Outcome distribution across reports

<table>
<thead>
<tr>
<th>Site/dataset</th>
<th>Site location (and dates when &gt;1 cohort/location)</th>
<th>Subjects n</th>
<th>Cure/completion</th>
<th>Failure</th>
<th>Death</th>
<th>Default/transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>San Francisco, CA, USA</td>
<td>55</td>
<td>67.3 (37)</td>
<td>12.7 (7)</td>
<td>9.1 (5)</td>
<td>10.9 (6)</td>
</tr>
<tr>
<td>2</td>
<td>Uzbekistan</td>
<td>77</td>
<td>66.2 (51)</td>
<td>13.0 (10)</td>
<td>13.0 (10)</td>
<td>7.8 (6)</td>
</tr>
<tr>
<td>3</td>
<td>South Africa (2000–2004)</td>
<td>2092</td>
<td>37.2 (778)</td>
<td>23.7 (496)</td>
<td>16.9 (354)</td>
<td>22.2 (464)</td>
</tr>
<tr>
<td>4</td>
<td>Estonia (2006–2008)</td>
<td>274</td>
<td>54.0 (148)</td>
<td>12.0 (33)</td>
<td>14.6 (40)</td>
<td>19.3 (53)</td>
</tr>
<tr>
<td>5</td>
<td>Italy</td>
<td>71</td>
<td>45.1 (32)</td>
<td>36.6 (26)</td>
<td>12.7 (9)</td>
<td>5.6 (4)</td>
</tr>
<tr>
<td>6</td>
<td>South Africa (1990–1999)</td>
<td>319</td>
<td>16.3 (52)</td>
<td>23.8 (76)</td>
<td>37.6 (120)</td>
<td>22.3 (71)</td>
</tr>
<tr>
<td>7</td>
<td>Bangladesh</td>
<td>668</td>
<td>84.3 (563)</td>
<td>5.5 (37)</td>
<td>4.8 (32)</td>
<td>5.4 (36)</td>
</tr>
<tr>
<td>8</td>
<td>Estonia (2001–2005)</td>
<td>292</td>
<td>45.2 (132)</td>
<td>19.5 (57)</td>
<td>13.0 (38)</td>
<td>22.3 (65)</td>
</tr>
<tr>
<td>9</td>
<td>Latvia</td>
<td>446</td>
<td>63.9 (285)</td>
<td>16.4 (73)</td>
<td>4.5 (20)</td>
<td>15.2 (68)</td>
</tr>
<tr>
<td>10</td>
<td>Lima, Peru</td>
<td>712</td>
<td>65.6 (467)</td>
<td>8.0 (57)</td>
<td>17.1 (122)</td>
<td>9.3 (66)</td>
</tr>
<tr>
<td>11</td>
<td>Manila, Philippines</td>
<td>161</td>
<td>60.9 (98)</td>
<td>11.2 (18)</td>
<td>11.2 (18)</td>
<td>16.8 (27)</td>
</tr>
<tr>
<td>12</td>
<td>Tomsk, Russia</td>
<td>243</td>
<td>74.1 (180)</td>
<td>9.9 (24)</td>
<td>4.5 (11)</td>
<td>11.5 (28)</td>
</tr>
<tr>
<td>Pooled frequency</td>
<td></td>
<td>5410</td>
<td>52.2 (2823)</td>
<td>16.9 (914)</td>
<td>14.4 (779)</td>
<td>16.5 (894)</td>
</tr>
</tbody>
</table>

Data are presented as % (n), unless otherwise stated. Cure/completion and failure were calculated retrospectively using all available culture data per consensus definitions; default, transfer and death were programme-assigned.

FIGURE 2. Intra-observation agreement of smear and culture during the first 24 months of treatment, stratified by treatment outcome.
additional monitoring should be evaluated in the context of the societal benefits of avoiding bad outcomes: mortality, acquisition of additional resistance and the creation of new cases of MDR-TB through transmission [30–36]. Future work could examine the relative impact and costs associated with different screening approaches, including in a dynamic model that addresses potential transmission. Such work could further estimate the individual and population impacts of monitoring policies, and associated costs and resource demands.

These findings raise the question of when and how to intervene to avert a predicted negative outcome. Guidelines recommend repeat drug susceptibility testing among patients whose cultures have not converted or have reverted after month 4. Although the present study demonstrates that failure can be predicted earlier with more frequent culture supplementing smear testing, additional clinical evaluation is required to determine which patients with positive culture results at month 4 or later require treatment adjustment to assure success and which will experience successful treatment outcome without any regimen adjustments.

FIGURE 3 The timing of failure detection in the last 12 months of treatment within and across the monitoring strategies. Data are presented as hazard ratio (95% CI). IQR: interquartile range.

FIGURE 4 Stratified analysis of failure detection in the last 12 months of treatment, comparing monthly smear to monthly culture. Panel a) presents individual (by study) and pooled hazard ratios (95% confidence intervals). Panel b) presents hazard ratios (95% confidence intervals) stratified by risk factors. BMI: body mass index. HR: isoniazid–rifampicin.
Benefits of reinforcing regimens with at least two new drugs, or considering adjunct interventions (e.g. surgery) [37], could accrue to patients receiving both individualised and standardised regimens [3].

These findings highlight the importance of further developing laboratory capacity for proper management of MDR-TB. Although dissemination of rapid tests for diagnosis of rifampicin-resistant TB has accelerated in recent years, this has not been accompanied by scaled-up culture services. Current minimum recommendations for improving diagnosis of smear-negative TB advise that high-burden countries have at least one culture laboratory per 5 million population. This low bar has been met in only 21 countries. Scaled-up culture capacity, to permit both regular monitoring throughout MDR-TB treatment and confirmatory drug susceptibility testing after molecular tests, is desperately needed to advance the goals of TB pre-elimination and elimination [35, 38, 39].

Limitations

There are several limitations to this study. First, large numbers of missing values for covariates precluded assessment of confounding; complete case analysis would have included only 78 of the 914 patients who failed. Second, there was substantial heterogeneity. However, stratified analyses revealed similar conclusions to the pooled analyses. Third, generalisability may be limited, as all sites included used the most frequent and sensitive monitoring method (monthly culture), which suggests a relatively strong health system. Fourth, this analysis was performed using the previous consensus definitions for treatment outcomes. The new definitions are still largely microbiologically based [40], so the general conclusions should not differ substantially with the new definitions. Last, variability in laboratory methods and quality, and in treatment, may have introduced unmeasured bias. Most sources of potential bias would result in underestimating the effect of substituting less sensitive monitoring for monthly culture. The one exception is smear quality, which was not directly assessed. If smear quality was suboptimal, the difference between smear and culture could be overestimated. At site 7, which invested substantially in smear quality, the estimated risk of detecting failure by smear versus culture was not statistically distinguishable from the null, highlighting this potential bias. However, all study sites participated in routine internal quality control and external quality assurance, so there is no reason to suspect poor quality of sputum smears. Conversely, the effect of substituting smear for culture may be underestimated. This is because only a small, unspecified proportion of the cultures were grown in liquid media, which is more sensitive than solid culture [41]. Since routine use of liquid culture is increasing, the true delay in failure detection incurred by substituting smear results for liquid culture results will probably be greater. Also creating the potential for bias toward the null is the use of individualised treatment in all sites except site 7. Individualised regimens are more likely than standardised regimens to be adjusted based on early indication of nonresponse. Although the simulated datasets do not use all available data, in practice, clinicians may have used results that were not counted in one of the less-frequent monitoring scenarios, to recommend a treatment change or adjunctive therapy, thereby avert failure. Consequently, the effect of less-sensitive monitoring on failure of individualised treatment may be underestimated.

Conclusions

Reduced frequency of culture monitoring or replacement of culture by sputum smear even in the final 12 months of treatment could threaten early detection of failure. Compromise on method could lead to greater delays to treatment adjustments or adjunctive therapies among important subpopulations treated for MDR-TB, e.g. those with HIV co-infection or baseline negative sputum smears. These findings reinforce the need for expanded global laboratory capacity for high-quality culture, in addition to substitute microscopy and rapid molecular tests. More sensitive indicators of nonresponse, with shorter turnaround time, would further facilitate early detection and avoid poor outcomes.

Acknowledgements

We express our deep gratitude to patients and providers in each of the 11 sites that provided data and to investigators from other sites who were willing to collaborate.

We would also like to thank the Collaborative Group for Analysis of Bacteriology Data in MDR-TB Treatment including Shama D. Ahuja (Bureau of Tuberculosis, New York, NY, USA), Yevgeny G. Andreev (Tomsk Penitentiary Services, Ministry of Justice, Tomsk, Russia), David Ashkin (A.G. Holley Hospital, Lantan, FL, USA), Monika Avendano (University of Toronto, Toronto, Canada), Rita Banerjee (Mayo Clinic, Rochester, MN, USA), Melissa Bauer (Montreal Chest Institute, McGill University, Montreal, Canada), Andrea Benedetti (Montreal Chest Institute, McGill University, Montreal, Canada), Jeannette Brand (Medical Research Council, Pretoria, South Africa), Edward D. Chan (Veterans Affairs Eastern Colorado Health Care System, Denver, CO, USA), Chen-Yuan Chiang and Kathy DeRiemer (UC Davis School of Medicine, Davis, CA, USA), Nguyen Huey Dung (National TB Control Program, Hanoi, Vietnam), Donald Enarson (International Union against Tuberculosis and Lung Disease, Paris, France), Katherine Flanagan (MRC Laboratories, Banjul, Gambia), Jennifer Flood (California Department of Public Health, Sacramento, CA, USA), Maria L. Garcia-Garcia (Instituto Nacional de Salud Publica, Mexico, Mexico), Neel Gandhi (US Centers for Disease Control and Prevention, Atlanta, GA, USA), Reuben M. Granich (US Centers for Disease Control and Prevention, Atlanta, GA, USA), Maria G. Holm-Delgado
(Montreal Chest Institute, McGill University, Montreal, Canada), Michael D. Iseman (National Jewish Health, Denver, CO, USA), Leah G. Jarlsberg (University of California, San Francisco, San Francisco, CA, USA), Hye-Ryun Kim (Korea Cancer Center Hospital, Seoul, Korea), Won-Jung Koh (Samsung Medical Center, Seoul, Korea), Joey Lancaster (Medical Research Council, Pretoria, South Africa), Christophe Lange (Medical Clinic, Tuberculosis Center Borstel, Borstel, Germany), Wi. C. M. de Lange (University Medical Center Groningen, Groningen, the Netherlands), Chi Chiu Leung (Tuberculosis and Chest Services, Hong Kong), Jiehui Li (New York City Health and Mental Hygiene, New York, NY, USA), Aung Kya Jai Maung (Damien Foundation, Dublin, Ireland), Masa Narita (University of Washington, Seattle, WA, USA), Rondel Odendaal (Medical Research Council, Pretoria, South Africa), Philly O’Riordan (City Road Medical Centre, London, UK), Madhukar Pai (Montreal Chest Institute, McGill University, Montreal, Canada), Domingo Palermo (Hospital F.J. Muñiz, Buenos Aires, Argentina), Seung-kyu (Park TB Center, Seoul, Korea), Geoffrey Pasvol (Imperial College London, London, UK), Jose Peña (Universidad Autonoma Madrid, Madrid, Spain), Carlos Pérez-Guzmán (Instituto de Salud del Estado de Aguascalientes, Aguascalientes, Mexico), Alfredo Ponce-de-Leon (Instituto Nacional de Ciencias Médicas y de Nutrición "Salvador Zubirán", Mexico, Mexico), Vija Riekstina (Clinic of Tuberculosis and Lung Diseases, Riga, Latvia-Latv.), Jerome Robert (Bacteriologic-Hygiene-UPMC, Paris, France), Sarah Royce (University of California, San Francisco, San Francisco, CA, USA), H. Simon Schaaf (Stellenbosch University, Stellenbosch, South Africa), Kwonju J. Seung (Brigham and Women’s Hospital, Boston, MA, USA), Lena Shah (Montreal Chest Institute, McGill University, Montreal, Canada), Tae Sun Shim (University of Ulsan College of Medicine, Seoul, Korea), Yuji Shiraishi (Fukuji Hospital, Tokyo, Japan), José Sifuentes-Osornio (Instituto Nacional de Ciencias Médicas y de Nutrición "Salvador Zubirán", Mexico, Mexico), Matthew J. Strand (National Jewish Health, Denver, CO, USA), Payam Tabarsi (Beheshti Medical University, Tehran, Iran), Thelma E. Tupasi (Tropical Disease Foundation, Makati City, Philippines), Robert van Altena (University Medical Center Groningen, Groningen, the Netherlands), Tjjp S. Van der Werf (University Medical Center Groningen, Groningen, the Netherlands), Mario H. Vargas (Instituto Nacional de Enfermedades Respiratorias, Mexico, Mexico), Janice Westenhousen (Center for Infectious Diseases-California Department of Public Health, Sacramento, CA, USA), Wing Wai Yew (Grantham Hospital, Hong Kong), Jae-Joon Yim (Seoul National University College of Medicine, Seoul, Korea).

The findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of the US Centers for Disease Control and Prevention. D. Falzon is a staff member of the World Health Organization (WHO). He alone is responsible for the views expressed in this publication and they do not necessarily represent the decisions or policies of WHO. The designations used and the presentation of this material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area, or of its authorities, nor concerning the delimitation of its frontiers or boundaries.

References

DOI: 10.1183/13993003.00462-2016