Comparison of two versus three smears in identifying culture-positive tuberculosis patients in a rural African setting with high HIV prevalence

A. C. Crampin,*† S. Floyd,† F. Mwaungulu,* G. Black,**† R. Ndhlovu,* E. Mwaiyeghele,* J. R. Glynn,† D. K. Warndorff,‡ P. E. M. Fine†

* Karonga Prevention Study, Chilumba, Malawi; † Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK; ‡ MRC Laboratories, Fajara, The Gambia

SUMMARY

SETTING: Karonga district, northern Malawi.
OBJECTIVE: To compare the sensitivity and specificity of two versus three smears for the diagnosis of pulmonary tuberculosis in a setting with high HIV prevalence.
DESIGN: A total of 1992 pulmonary tuberculosis suspects with three sputum smears taken over a 2–7 day period and at least one culture result were studied. Smears were auramine stained and examined using fluorescence microscopy, and positives were confirmed with Ziehl-Neelsen staining and light microscopy. Cultures were set up on Löwenstein-Jensen media. True negative and positive status was defined on the basis of culture. The sensitivity, specificity, and positive and negative predictive values of two and three smears were compared.
RESULTS: Compared to culture, the sensitivity, specificity, and positive and negative predictive values of three smears were 70%, 98%, 92%, and 92%, respectively. Restriction to the first two smears gave similar results. Of those detected as smear-positive using three smears, at least 97% would have been detected by two. Among those with HIV serology results available, the sensitivity of two smears for detecting culture-positive tuberculosis was identical to that using three.
CONCLUSION: In this setting, using fluorescence and light microscopy, collecting two smears rather than three would only marginally reduce sensitivity and would slightly improve the specificity of diagnosis of tuberculosis; this is unaffected by HIV status. The potential for improving specificity is important because of the costs of misdiagnosis. In practice, both sensitivity and specificity may be increased due to the time saved by examining two rather than three smears.
KEY WORDS: tuberculosis; diagnosis; smears; HIV; Africa

IN MALAWI, as in many countries in sub-Saharan Africa, tuberculosis has increased dramatically over the last 10 years, mainly due to the human immunodeficiency virus (HIV) epidemic. The number of people examined as tuberculosis ‘suspects’ has also risen. The impact on laboratory services is large, and is exacerbated by low morale and high mortality in government health staff, and continual problems with supplies and laboratory security.

The current World Health Organization and International Union Against Tuberculosis and Lung Disease (IUATLD) recommendations for diagnosis of pulmonary tuberculosis are based on examination of three sputum specimens collected from each suspect over a 2-day period.1 It is likely that, in practice, many of these three specimens are not examined well (a minimum of 100 high power fields should be examined before declaring a smear negative). Worldwide, tens if not hundreds of thousands of sputum samples and smears are processed each day. Were the recommended number of sputum samples to be reduced, to two smears or even one, there would be significant savings on sputum containers, slides and reagents, and time.2,3 Staff may have more incentive to examine smaller numbers, and thorough examination of two slides might even provide a more sensitive and specific diagnostic routine than rushed examination of three slides. It could therefore be helpful to change this policy in over-stretched systems.4

In one district of Malawi a change in policy from three to two smears did not affect the proportion of suspects found to be smear-positive, the ratio between smear-positive and -negative cases registered, or other programme outcome indicators, in comparison with an earlier time period.2 Analyses of the proportions of smear-positive cases who would have been diagnosed...
if only one or two rather than three smears were examined have been performed in neighbouring countries. In a large study in Tanzania it was estimated that 83% of smear-positive cases would have been found on the first smear and 96% on the first two smears. Identical figures were found in a small study in Malawi, and in a Zambian study the proportions were 77% and 93%, respectively. Similar figures have been found in other settings and in older studies, suggesting that these proportions may not have been greatly affected by HIV, despite reports that HIV-positive pulmonary tuberculosis patients are less likely than those who are HIV-negative to be smear-positive and to have high grade smears.

Where sets of sputa from individual patients are identifiable in the laboratory (as is the case, for example, if the IUATLD forms are used) there is a risk that slides are not interpreted independently of each other. There is a temptation to read thoroughly only the first one or two of each set and fill in the other results accordingly. In this situation a retrospective review of recorded results is likely to be misleading, and will overestimate the sensitivity and specificity of using just two smears if the results from the first two smears are compared with the results overall. It is not clear from the published studies if slides were routinely read blind to the other results from that patient.

As part of the Karonga Prevention Study (KPS) in Northern Malawi we have compared the sensitivity and specificity of two versus three sputum smears for the diagnosis of tuberculosis. The KPS is responsible for the diagnosis of tuberculosis in Karonga District, a rural area of Malawi covering a population of approximately 200 000 people. Due to its organisation as a research project, the KPS is able to provide a quality laboratory service with rigorous examination and audit methods within a rural African context, where the prevalence of tuberculosis amongst suspects and the nature of clinical disease are similar to those found elsewhere in the region.

The HIV prevalence amongst patients with laboratory-confirmed tuberculosis in Karonga from 1996–1999 was 63%, and among women seen in antenatal clinics in 1999 the HIV prevalence was 10%. The HIV prevalence amongst tuberculosis ‘suspects’ is likely to be high, even if they do not have tuberculosis, as tuberculosis-like symptoms are common among those with HIV infection, whether or not they have tuberculosis.

**METHODS**

KPS staff are responsible for the collection of sputum specimens from tuberculosis suspects, in particular those with a cough of at least 3 weeks' duration. Suspects may either self refer or else be found actively by KPS staff at health centres, among hospital inpatients admitted for other reasons, or in the field in the course of other studies. According to national protocol, three sputum specimens are collected from each suspect, one on-the-spot specimen, a second early morning specimen and a third on-the-spot specimen collected at the time of delivery of the second specimen.

Smears are prepared in the KPS laboratory, stained with auramine and examined under a fluorescence microscope at ×40 magnification. All smears are examined independently in random order, and the microscopists are not aware which combination of sputa make up an individual patient’s ‘set’. Positives are destained and restained with Ziehl-Neelsen (ZN) stain and examined with a light microscope at ×100 magnification. Only those confirmed by ZN are recorded as positive.

For culture, 2 ml of sputum is added to 2 ml of sterile 4% NaOH solution in a 20 ml plastic Universal bottle (sputum with a volume of less than 2 ml is made up to 2 ml with sterile distilled water). The specimens are then vortexed briefly before incubation at room temperature for 10 minutes. They are then vortexed briefly for a second time and incubated at room temperature for a further 10 minutes. Next, 16 ml of sterile distilled water is added to the specimens, and they are then centrifuged at 2000 rpm for 20 minutes. The supernatant is then discarded, the deposit is mixed gently, and two drops of the deposit are dispensed directly onto solid Löwenstein-Jensen (LJ) media containing pyruvate. The cultures are incubated horizontally at 37°C overnight to allow the inoculum to disperse evenly over the surface of the medium, after which they are incubated vertically at 37°C for a further 10 weeks (during which time they are reviewed every 2 weeks). Cultures are examined macroscopically, and any with growth of one or more colonies that are consistent with *Mycobacterium tuberculosis* are identified to species level at the UK Mycobacterium Reference Laboratory in Dulwich.

It has not been possible to set up a local external quality control system, but rigorous quality control and audit procedures are in place in the Chilumba laboratory. Since 1998 each slide has been examined independently by two examiners and conflicting results re-examined by both examiners until a consensus is reached. Three supplementary controls are set up with each batch of cultures. One is a positive control that had 10–99 acid-fast bacilli (AFB) per 100 fields on microscopy which has been demonstrated previously to be culture-positive and has been stored at −20°C. There are also two negative controls: LJ medium alone and LJ medium inoculated with the distilled autoclaved water used during the culture session. This water sample is treated in the same way as the sputum samples and thus provides a check for potential contamination of all reagents as well as for sterile technique. Also, if the first culture from a patient who is smear-positive gives a negative result, then it is recultured.
Suspects who are to be registered as TB patients are counselled for HIV testing; those who consent are bled for HIV serology, employing both particle agglutination and ELISA methods.

The results reported here were initially recorded into field and laboratory registers and all data were double entered into Foxpro databases. The data set covers a period from the last week of November 1996 to the end of December 1999. Data were cleaned and analysed in Excel 5.0 and Stata 6.0. A tuberculosis suspect was included if three specimens were collected within 7 days of each other and if the three specimens were not all collected on the same day. Only the first eligible set of sputa for each patient was included, so each individual contributes only one set.

For the purposes of this analysis, a ‘true negative’ case is defined as someone with at least one negative culture and no positive cultures for M. tuberculosis complex from sputa collected within one month of the first of a set of three ‘suspect’ sputa (a month was chosen to maximise the sensitivity of culture, as this provided the reference standard for the assessment of smears). A ‘true positive’ case is someone with one or more cultures positive for M. tuberculosis or M. bovis within the same time period. Smears with any degree of positivity (even ‘scanty’, defined as 1–10 AFB in 100 fields) were counted as positive in defining a positive set. Although in a clinical setting a single scanty smear might not be enough to warrant starting on treatment, it would initiate further investigation and is thus sufficient for a screening test.

We compared the sensitivity and specificity of diagnosis based on two or three smears against that determined by culture. We also compared the sensitivity of two smears for detecting those classified as smear-positive on the basis of three smears. We have assumed that the order in which sputa were collected was important, as early morning specimens have been shown to have a higher yield than do on-the-spot specimens.7,13 The order of specimens was therefore taken into account in the analysis.

RESULTS

Over the study period, 2459 people were seen as TB suspects. Of these, 198 were excluded because they produced fewer than three sputum specimens, 101 were excluded because the three sputa were not collected within seven days of each other, 67 were excluded because there was no information on culture results (for 15 this was due to contamination of all three cultures, giving a loss due to contamination of 15/2160, or 0.7%), and 63 were excluded because all three smears were collected on the same day, and their order could not be determined. A further 38 were excluded because, although they had at least one culture containing AFB, no species identification was available. All reported analyses are based on the remaining 1992 sets. The proportions smear- and culture-positive are shown in Table 1. The proportion culture-positive was correlated strongly with the number of positive smears in a set.

Of the 1992 sets of smears, 1524 sets (76.5%) were collected over a 2-day period as recommended (with the first an on-the-spot specimen, the second an early-morning specimen and the third taken at the time of delivery of the early-morning specimen). For 148 sets (7.4%) the first two smears were collected on the first day, with the third smear 1–4 days later. For the remainder, the second and third smears were collected on a different day to the first smear: for 273 individuals (13.7%) the three smears were obtained over a 3–4 day period, and for 47 individuals (2.4%) over a 5–7 day period. Of the 445 classified as culture-positive, 443 grew M. tuberculosis and two grew M. bovis. Among those classified as culture-negative were 88 patients who had at least one specimen confirmed as an environmental mycobacterium.

The sensitivity, specificity and positive predictive value of sputum smears compared to culture are shown in Table 2. In eight of the sets it was not possible to determine whether the first two smears should be included as positive or negative, as the first smear was negative and the other two specimens had the same date and only one was positive. The results have thus been calculated in two ways: first, assuming that the two negative smears would have been seen first, so that if only two smears were examined the patient would have been classified as smear-negative (‘worst case’); and secondly, assuming that the positive smear would have been among the first two, so that the patient would have been classified as smear-positive (‘best case’). Of these eight sets, five were culture-positive (M. tuberculosis) and three negative (of which one grew an environmental mycobacterium). As Table 2 shows, the sensitivity and specificity of two smears for detecting culture-positive tuberculosis were very similar to those using three smears, using either scenario.

In a programme setting, there may be more interest in identifying smear-positive cases than culture-positive smear-negative cases, because the former constitute a greater transmission hazard. Table 2 also shows the sensitivity of two smears for detecting those who were positive on at least one of three smears. (By def-

---

### Table 1: Positive smears and cultures among 1992 patients with suspected tuberculosis

<table>
<thead>
<tr>
<th>Number of positive smears per set</th>
<th>Total sets</th>
<th>Percentage</th>
<th>Culture-positive n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>289</td>
<td>14.5</td>
<td>273 (94.5)</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>1.8</td>
<td>30 (85.7)</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>0.7</td>
<td>7 (50.0)</td>
</tr>
<tr>
<td>0</td>
<td>1654</td>
<td>83.0</td>
<td>135 (8.2)</td>
</tr>
<tr>
<td>Total</td>
<td>1992</td>
<td>100.0</td>
<td>445 (22.3)</td>
</tr>
</tbody>
</table>
DISCUSSION

In this study, of 338 patients detected as smear-positive on the basis of three specimens, at least 329 (97%) would have been detected if only two specimens had been examined. Using two smears rather than three, the sensitivity for detecting culture-positive patients was only very marginally reduced, and the specificity was marginally improved. The high yield of two smears is consistent with other studies.3,5,7,8 As the slides were read blind in this study, these estimates are not artefactually inflated.

The high sensitivity in this study may be attributed in part to the use of fluorescence microscopy, which is known to be more sensitive than ZN microscopy.9 In addition, the duplicate reading of smears by two technicians, which began in 1998, might have improved the quality of smear reading. However, the sensitivity and specificity showed little change over this period, having been 66% and 98%, respectively, in 1997, compared to 71% and 99% in 1999 (calculated with the worst case assumption).

The zero false-positive rate, and the fact that in all culture batches at least one patient with a smear of 10–99 AFB in 100 fields had a positive culture, demonstrates that the culture results used as the ‘gold standard’ are reliable. The relatively high proportion of positive cultures for which growth was not due to M. tuberculosis (88/533) is consistent with the known high level of exposure to environmental mycobacteria in this population.14

It is important to consider the potential for improving specificity when evaluating two versus three smears. The more smears that are examined the higher is the chance that a positive will be found, but the proportion of false positives also increases.15,16 In this study the proportion of patients with culture-confirmed tuberculosis was much lower in those with only one positive smear than in those with two or three (Table 1). This proportion will be even lower if the single smear has only scanty bacilli.17

Although it is important to diagnose as many tuberculosis cases as possible, particularly smear-positive cases, the costs of misdiagnosis can be high. These include financial and time costs both to the health service and patients, and costs in terms of poor treatment outcome in those misdiagnosed as having tuberculosis. In this population we have shown that mortality was four times as high in smear-positive tuberculosis cases as possible, particularly smear-positive bacillae.17

In this population, the time saving may also improve sensitivity in overloaded laboratories. Also, if smear-negative, symptomatic tuberculosis suspects are investigated further with X-rays, trials of antibiotic therapy, and additional sputum specimens if there is no clinical

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Diagnostic performance of two or three smears against gold standards based upon either culture or summary smear status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
</tr>
<tr>
<td>Gold standard based on culture status</td>
<td>Smear-positive on ≥1 of</td>
</tr>
<tr>
<td>Three smears</td>
<td>69.7% (310/445)</td>
</tr>
<tr>
<td>Two smears (worst case)</td>
<td>68.5% (305/445)</td>
</tr>
<tr>
<td>Two smears (best case)</td>
<td>69.7% (310/445)</td>
</tr>
<tr>
<td>Gold standard based upon three smears</td>
<td>Smear-positive on ≥1 of</td>
</tr>
<tr>
<td>Two smears (worst case)</td>
<td>97.3% (329/338)</td>
</tr>
<tr>
<td>Two smears (best case)</td>
<td>99.7% (337/338)</td>
</tr>
</tbody>
</table>
improvement, then the proportion of tuberculosis patients missed should be minimised.1

As well as contemplating a reduction in the number of smears collected from three to two, policy makers should also consider the collection of two smears on the same day, rather than one on-the-spot followed by an early morning specimen (although it has been shown that early morning specimens have a higher yield). Additional time savings might be made, and a higher proportion of suspects would contribute a full ‘set’, as with the current recommendation it is inevitable that some fail to attend for the second contact. This study cannot be used to evaluate such a policy (almost all patients in this study gave an on-the-spot specimen, followed by two further specimens the following day, as recommended), but an evaluation should be carried out.

Several studies have suggested that smears may be less sensitive for detecting pulmonary tuberculosis in HIV-positive than in HIV-negative patients,10–12 but other studies have found similar proportions of HIV-positive and HIV-negative pulmonary tuberculosis patients to be smear-positive.19,20 Cavities are less common in conjunction with HIV, but HIV-positive patients with no or minor X-ray changes can still be smear-positive.19 In our study, although HIV-positive patients with culture-proven pulmonary tuberculosis were less likely than HIV-negative patients to be smear-positive, the proportion of smear-positive patients detected on the first two smears was similar in HIV-positive and -negative patients. This is reassuring, as it is in communities with a high prevalence of HIV that the need to improve the efficiency of tuberculosis diagnosis is greatest.

Acknowledgements

We thank M Yates (Mycobacterium Reference Unit, Public Health Laboratory Service, Dulwich) for species confirmation of all isolates, the Government of the Republic of Malawi for their interest in and support of the Project, the Malawi National Tuberculosis Control Programme for their cooperation, and the Malawi Health Sciences Research Committee for permission to publish this paper.

The Karonga Prevention Study is funded primarily by the Wellcome Trust, with contributions by the British Leprosy Relief Association (Lepra). JRG is partially supported by the British Department for International Development. 

References


CADRE : District de Karonga, Malawi du Nord.
OBJECTIF : Comparer la sensibilité et la spécificité de deux versus trois frottis pour le diagnostic de la tuberculose pulmonaire dans un contexte à haute prévalence du VIH.
SCHEMA : On a étudié 1.992 suspects de tuberculose pulmonaire où trois frottis d’expectoration avaient été prélevés pendant une période de 2 à 7 jours et disposant d’au moins un résultat de culture. Les frottis ont été colorés à l’auramine et examinés par microscopie à fluorescence ; les cas positifs ont été confirmés par coloration de Ziehl-Neelsen et microscopie optique. Les cultures ont été ensemencées sur milieu de Löwenstein-Jensen. On a défini le statut de vrai négatif et de vrai positif sur la base de la culture. On a comparé la sensibilité, la spécificité et les valeurs prédictives positive et négative des deux versus trois frottis.
RÉSULTATS : Par comparaison avec la culture, la sensibilité, la spécificité et les valeurs positive prédictive et négative de trois frottis étaient respectivement de 70%, 98%, 92% et 92%. Si l’on se limite aux deux premiers frottis, les résultats sont similaires. Parmi les cas à bacilloscopie positive détectés au moyen de trois frottis, au moins 97% l’auraient été par deux. Parmi ceux dont les résultats de la sérologie VIH sont disponibles, la sensibilité des deux frottis pour la détection d’une tuberculose à culture positive est identique à celle de trois frottis.
CONCLUSION : Dans ce contexte, par l’utilisation de la microscopie à fluorescence et à lumière du jour, le recueil de deux frottis plutôt que trois ne réduirait que de façon marginale la sensibilité et améliorerais légèrement la spécificité du diagnostic de tuberculose et ceci quel que soit le statut VIH. La potentialité d’amélioration de la spécificité est importante en raison des coûts d’un diagnostic erroné. En pratique, tant la sensibilité que la spécificité peuvent être augmentées grâce au temps épargné par l’examen de deux plutôt que de trois frottis.

RESUMEN

MARCO DE REFERENCIA : Distrito de Karonga en el Norte de Malawi.
OBJETIVO : Comparar la sensibilidad y la especificidad de dos versus tres bacilloscopias para el diagnóstico de la tuberculosis, en un contexto de alta prevalencia del virus del inmunodeficiencia humana (VIH).
MÉTODO : Se estudiaron 1992 sujetos sospechosos de tuberculosis que habían tenido tres frotis de expectoración en un periodo de 2 a 7 días y que disponían del resultado de por lo menos un cultivado de expectoración. Los frotis fueron teñidos con auramina y examinados con microscopía de fluorescencia ; los casos positivos fueron confirmados con el método de Ziehl-Neelsen. Para los cultivos se utilizó el medio de Löwenstein-Jensen. El criterio de verdadero positivo y verdadero negativo se estableció en base al cultivo. Se comparó la sensibilidad y especificidad y los valores predictivos positivo y negativo de dos versus tres frotis.
RESULTADOS : En comparación con la cultura, la sensibilidad, la especificidad y los valores predictivos positivo y negativo de tres frotis eran 70%, 98%, 92% y 92%, respectivamente. Si se toman en consideración solamente los dos primeros frotis, los resultados son similares. Entre los casos con bacilloscopía positiva detectados por medio de tres frotis, por lo menos el 97% de ellos habrían sido detectados por dos. Entre los pacientes con resultados de serología VIH disponibles, la sensibilidad de dos frotis para detectar una tuberculosis con cultivo positivo fue idéntica a la de tres frotis.
CONCLUSIÓN : En este contexto, utilizando la microscopía de fluorescencia y el método de Ziehl Neelsen, el examen de dos frotis en lugar de tres, reduciría sólo marginalmente la sensibilidad y mejoraría ligeramente la especificidad del diagnóstico de la tuberculosis, independientemente de la situación con respecto al VIH. El potencial de mejoramiento de la especificidad es importante, debido a los costos de los errores de diagnóstico. En la práctica, tanto la sensibilidad como la especificidad pueden aumentar gracias a la economía de tiempo al examinar dos más bien que tres frotis.