Reducing the number of sputum samples examined and thresholds for positivity: an opportunity to optimise smear microscopy

M. Bonnet,* A. Ramsay,† L. Gagnidze,* W. Githui,‡ P. J. Guerin,* F. Varaine§

* Epicentre, Paris, France; † Liverpool School of Tropical Medicine, Liverpool, UK; ‡ Centre for Respiratory Diseases Research, Kenya Medical Research Institute, Nairobi Kenya; § Médecins Sans Frontières, Paris, France

SUMMARY

OBJECTIVE: To evaluate the impact on tuberculosis (TB) case detection and laboratory workload of reducing the number of sputum smears examined and thresholds for diagnosing positive smears and positive cases.

DESIGN: In this prospective study, three Ziehl-Neelsen stained sputum smears from consecutive pulmonary TB suspects were examined blind. The standard approach (A), ≥2 positive smears out of 3, using a cut-off of 10 acid-fast bacilli (AFB)/100 high-power fields (HPF), was compared with approaches B, ≥2 positive smears (≥4 AFB/100 HPF) out of 3, one of which is ≥10 AFB/100 HPF; C, ≥2 positive smears (≥4 AFB/100 HPF) out of 3; D, ≥1 positive smear (≥10 AFB/100 HPF) out of 2; and E, ≥1 positive smear (≥4 AFB/100 HPF) out of 2. The microscopy gold standard was detection of at least one positive smear (≥4 AFB/100 HPF) out of 3.

RESULTS: Among 644 TB suspects, the alternative approaches detected from 114 (17.7%) (approach B) to 123 cases (19.1%) (approach E) compared to 105 cases (16.3%) for approach A (P < 0.005). Sensitivity ranged between 82.0% (105/128) for A and 96.1% (123/128) for E. The single positive smear approaches reduced the number of smears by 36% compared to approach A.

CONCLUSION: Reducing the number of specimens and the positivity threshold to define a positive case increased the sensitivity of microscopy and reduced laboratory workload.

KEY WORDS: tuberculosis; microscopy; diagnosis

* Tuberculosis (TB) care that includes effective detection and treatment of patients is central to the global TB control strategy. Currently available diagnostics are either insensitive, time-consuming or require laboratory infrastructures that are considerably more sophisticated than those commonly found in the developing countries where TB burdens are highest. There is an urgent need for a new TB diagnostic test that is simple, rapid, sensitive and specific and can be made widely available. The test development pipeline is unlikely to deliver a test with such a profile in the short term. A recently developed model to evaluate the potential role of better diagnostics in improving TB control in developing countries has predicted that improving the performance of sputum smear microscopy and reducing the loss of patients during the diagnostic process would be associated with considerable public health impact.

Pulmonary TB patients with more than 10^5 Mycobacterium tuberculosis organisms per ml of sputum (approximately half of all pulmonary TB patients worldwide) can be diagnosed using direct sputum smear microscopy. International guidelines recommend the microscopic examination of three serial sputum specimens for acid-fast bacilli (AFB) in the investigation of pulmonary TB suspects, and define a positive case as a case with at least two smear-positive results. The first specimen is a ‘spot’ specimen collected when the patient presents at the health facility. The second is an ‘early morning’ specimen collected at home, and the third a further ‘spot’ specimen collected when the patient returns to the health facility to deliver the early morning specimen. At least one further visit to the health facility is usually required for the patient to collect the smear results.

Smear microscopy as currently recommended is associated with two major problems: variable sensitivity in detecting smear-positive cases, and a high rate of patient drop-out during the diagnostic process. A systematic review has demonstrated that the average incremental yield and/or increase in sensitivity of examining a third sputum specimen ranged from 2% to 5%, and concluded that reducing the number of specimens examined from three to two (and particularly...
to two specimens collected on the same day) could potentially increase case detection.\textsuperscript{4} It was suggested that this could be achieved by reducing the workload in over-burdened laboratories and the number of patient visits required. Limitations of the review, as identified by the authors, were that few studies were blinded or reported microscopy quality assurance, thresholds for defining a positive smear or the nature of the sputum specimens (i.e., whether a specimen was a first spot, morning or second spot specimen). Recent studies have shown the high specificity of single positive smear results and/or scanty smear results in high TB prevalence settings.\textsuperscript{9,10} These studies have not as yet influenced policy change with regard to international guidelines on the definitions of a smear-positive case.

We prospectively evaluated different approaches to sputum smear microscopy based on the examination of two and three specimens, with differing definitions of a ‘smear-positive’ case. In so doing we sought to avoid the methodological deficiencies of many previous studies that were identified by the authors of the afore-mentioned systematic review.\textsuperscript{3} We considered the effect of the different approaches on case detection, laboratory workload and diagnostic delay.

\section*{STUDY POPULATION AND METHODS}

The present study was conducted in an urban human immunodeficiency virus (HIV)-TB health clinic supported by Médecins Sans Frontières in Mathare, a slum area of Nairobi, Kenya. The clinic’s laboratory routinely performs sputum smear microscopy. All consecutive patients aged $\geq$ 15 years with cough $\geq$ 2 weeks were eligible for the study. After providing written informed consent, patients submitted three sputum specimens over 2 consecutive days. The first specimen was collected on the spot at the initial consultation, the second at early morning and the third on the spot when the patient delivered the morning specimen. The hot Ziehl-Neelsen method (carbol-fuchsin 1\%) was used. Slides were examined using bright-field microscopy (magnification $\times$ 1000) by two independent laboratory technicians blinded to the result of the previous smear. The exact number of AFB observed in 100 high power fields (HPF) for each smear result was recorded on a separate laboratory form. Results were then recorded independently in the programme laboratory register by the study supervisor. Several approaches to sputum smear microscopy for the detection of smear-positive TB cases were compared (Table 1). The approaches varied in the AFB cut-off used to define a positive smear, the number of positive smears required to define a smear-positive case and whether two or three specimens were examined.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A\textsuperscript{4}</td>
<td>At least 2 smear-positive results out of 3, using a cut-off of 10 AFB/100 HPF</td>
</tr>
<tr>
<td>B\textsuperscript{4}</td>
<td>At least 2 smear-positive results ($&lt;$ 4 AFB/100 HPF) out of 3, one of which being $\geq$ 10 AFB/100 HPF</td>
</tr>
<tr>
<td>C\textsuperscript{1}</td>
<td>At least 2 smear-positive results out of 3, using a cut-off of 4 AFB/100 HPF</td>
</tr>
<tr>
<td>D</td>
<td>At least 1 smear-positive result out of the 2 first specimens collected, using a cut-off of 10 AFB/100 HPF</td>
</tr>
<tr>
<td>E</td>
<td>At least 1 smear-positive result out of the 2 first specimens collected, using a cut-off of 4 AFB/100 HPF</td>
</tr>
</tbody>
</table>

AFB = acid-fast bacilli; HPF = high power fields.

50\% to 100\% of positive and 10\% to 20\% of randomly selected negative slides. An external quality assessment of 100 randomly selected slides at the end of the study was performed blind at the TB laboratory of the Centre for Respiratory Diseases Research of the Kenyan Medical Research Institute (KEMRI), Nairobi, Kenya. Inter-reader and intra-reader reliability was assessed on a sample of 200 smears randomly selected to be reread blind by the second reader or by the same reader at 1-day intervals.

Data were double-entered using EpiData 3.1 (EpiData, Odense, Denmark, www.epidata.org). Intention to treat analysis was performed using SPSS\textsuperscript{®} 11.0 for Windows (SPSS Inc., Chicago, IL, USA).

The smear-positive case detection rate was calculated for each approach and compared between the alternative approaches (B to E) and the standard approach (A), using McNemar’s test to compare matched data. The performance of the different approaches was measured using the detection of at least one smear with $\geq$ 4 AFB/100 HPF out of three smears as the microscopy ‘gold standard’, based on the very good correlation with culture results reported in previous studies.\textsuperscript{10,12}

The detection yield of the first smear and the incremental yield of the second smear compared to the first smear and of the third smear compared to the first two smears were calculated as the ratio of the number of cases with first smear positive, first smear negative and second smear positive, and two first smears negative and third smear positive, divided by the total number of cases with at least one positive smear out of three examined. The smear-positive detection rate of the second, early morning, specimen was compared with that of the third, on-the-spot, specimen, which was prepared together with the second specimen.

Inter-reader and intra-reader reliability was assessed by the calculation of the Kappa coefficient, which measures the extent to which the results of both tests vary when read by two independent readers or at a 1-day interval by the same reader. A $\kappa$ value between 0.80 and 1 signifies almost perfect agreement.

The mean time between initial consultation and final smear results and the total number of slides read per approach were calculated.
The study was approved by the Ethical Review Committees of KEMRI (Nairobi, Kenya) and the 'Comité de Protection des Personnes', Saint Germain-en-Laye, France.

## RESULTS

Between February and December 2005, 644 patients were included in the study (Figure), resulting in the examination of 1879 specimens. The mean age was 32 years (standard deviation [SD] 10.3) and the male/female sex ratio was 0.8 (287/356). One hundred and twenty one patients (18.8%) had a past history of TB and 37 (5.7%) had received broad-spectrum antibiotics in the 2 weeks before sputum investigation. In addition to cough, 593 patients had a history of fever (92.3%), 596 (92.5%) chest pain, 109 (16.9%) haemoptysis, 245 (38.0%) reported weight loss and 484 (95.3%) were able to produce three sputum specimens, 23 (88.5%) had scanty results (1–9 AFB/100 HPF) and 8 salivary (0.4%). A final total of 1855 (98.7%) good quality smears were defined microscopically by the presence of blue cellular elements without debris or artifacts.

Standard approach A detected 105 smear-positive cases (16.3%). Smear-positive detection rates of the different approaches compared to the standard WHO approach (A) are presented in Table 2. All detected significantly more smear-positive cases than approach A. The microscopy gold standard detected a total of 128 smear-positive cases. The performance of the different approaches is presented in Table 3. D and E were the most sensitive approaches.

Using a cut-off of 10 and 4 AFB/100 HPF, respectively 120 and 128 patients were detected with at least one smear-positive result out of three examined. The incremental detection yield of the first, second and third specimen is shown in Table 4. Regardless of the cut-off, the incremental yield of the second smear compared to the first smear was significant (P < 0.001). This was not the case for the incremental yield of the third smear compared to the two first smears (P = 0.25 for 10 AFB/100 HPF and P = 0.06 for 4 AFB/100 HPF).

The smear-positive detection yield of the morning specimen (using 10 AFB/100 HPF cut-off) was 18.4% (114/621) compared to 14.2% (88/621) for the third specimen (P < 0.001). Of the 26 cases detected as positive only after examination of the morning specimen, 23 (88.5%) had scanty results (1–9 AFB/100 HPF) after examination of the third (on-the-spot) specimen. Using 4 AFB/100 HPF as the cut-off, the detection yield of the morning specimen was 19.0% (118/621) compared to 17.6% (109/621) for the third specimen; the difference was not significant (P = 0.06). Of nine cases detected as positive only after examination of the second specimen, 6 (66.7%) had 1–3 AFB/100 HPF after examination of the third specimen.

## Table 3 Performance of the different approaches for detecting smear-positive TB cases using at least one positive smear (>4 AFB/100 HPF) result out of three as the microscopy gold standard

<table>
<thead>
<tr>
<th>Approach</th>
<th>Detected n</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>105 (16.3)</td>
<td>82.0</td>
<td>100</td>
<td>100</td>
<td>95.7</td>
</tr>
<tr>
<td>B</td>
<td>114 (17.7)</td>
<td>89.1</td>
<td>100</td>
<td>100</td>
<td>97.4</td>
</tr>
<tr>
<td>C</td>
<td>115 (17.9)</td>
<td>89.8</td>
<td>100</td>
<td>100</td>
<td>97.5</td>
</tr>
<tr>
<td>D</td>
<td>117 (18.2)</td>
<td>91.4</td>
<td>100</td>
<td>100</td>
<td>97.9</td>
</tr>
<tr>
<td>E</td>
<td>123 (19.1)</td>
<td>96.1</td>
<td>100</td>
<td>100</td>
<td>99.0</td>
</tr>
</tbody>
</table>

TB = tuberculosis; AFB = acid-fast bacilli; HPF = high power fields; PPV = positive predictive value; NPV = negative predictive value.

## Table 4 Smear microscopy incremental yield of the first, second and third specimen using one positive smear to define a positive case and two different AFB cut-offs to define a positive smear

<table>
<thead>
<tr>
<th>AFB cut-off</th>
<th>Detected cases on first specimen n (%)</th>
<th>Newly detected cases on second specimen n (%)</th>
<th>Newly detected cases on third specimen n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 AFB/100 HPF</td>
<td>120 (73.3)</td>
<td>29 (24.7)</td>
<td>3 (2.5)</td>
</tr>
<tr>
<td>4 AFB/100 HPF</td>
<td>128 (85.2)</td>
<td>14 (10.9)</td>
<td>5 (3.9)</td>
</tr>
</tbody>
</table>

AFB = acid-fast bacilli; HPF = high power fields.

WHO = World Health Organization.
The mean time between the initial consultation and the final smear results was 1.6 days (SD 2.9) using approaches D and E based on collection of two specimens, compared to 2.0 days (SD 3.4) with approaches A–C based on collection of three specimens (P < 0.001).

Inter-reader and intra-reader reliability were very good, with a κ coefficient of 0.83 (95% confidence interval [CI] 0.76–0.86) and 0.91 (95% CI 0.88–0.94), respectively. The sensitivity of the monthly quality control ranged between 92% and 100% and specificity between 95% and 100%. There was no positive/negative discordance and only one scanty result controlled as negative in the external quality control.

The three-specimen based approaches (A, B, C) resulted in the reading of 1879 slides (644 + 621 + 614). The two-specimen based approaches D and E resulted in the reading of 1200 (644 + [644 – 88]) and 1179 slides (644 + [644 – 109]), respectively, which reduced by 36.1% (679/1879) and 37.2% (700/1932) the number of slides to be read compared to the three-specimen based approaches.

**DISCUSSION**

Adopting an approach based on two specimens and defining a smear-positive patient as having at least one positive smear result, regardless of AFB cut-off, would increase smear-positive case detection significantly compared to the use of the standard approach based on two smear-positive results out of three specimens. A national survey in Indonesia recently reported an increase in prevalence of smear-positive cases from 104 to 120 per 100 000 population using one positive smear instead of two to define a positive case.13

The addition of the third sputum smear did not significantly increase the detection of smear-positive cases: the incremental yield of the third compared to the two first sputum samples was in the range of 2% to 5%, as reported in a recent systematic review.8 A recent retrospective laboratory registers-based study in Mongolia and Zimbabwe reported 0.7–4.5% suspects detected with a positive result only on the third smear examination.14 Better smear-positive detection was reported on the early morning specimen compared to the on-the-spot specimen, as has already been shown elsewhere, but was statistically non-significant when a smear-positive result was defined by the detection of ≥4 AFB/100 HPF.15 In our study, the increase in detection was mainly the consequence of a better concentration of AFB in the morning specimen and was therefore more significant when using a higher AFB cut-off. With the use of 4 AFB/HPF cut-off, our results are consistent with the results of a recent study that assessed the diagnostic yield of the examination of two sputum specimens collected on a single day at the initial consultation against the standard 2-day approach, and did not report a significantly different smear-positive case detection rate.6

The use of a 4 AFB/100 HPF cut-off to define a positive smear result rather than a 10 AFB cut-off increased smear-positive case detection. Due to the absence of a culture gold standard in our study, we used 4 AFB/100 HPF as the cut-off rather than 1 AFB/100 HPF because previous studies in similar settings reported a good correlation between culture results and smear results with ≥4 AFB when good, quality-assured microscopy was in place, even in high HIV-prevalence settings.9,10,16

This was the case in our study, in which the high inter-reader reliability and good results of the quality controls support the high quality of the smear microscopy. Our choice of AFB cut-off was strengthened by the low rate of false-positive smear results (54/3830, 1.4%) reported by the very large study conducted in 1980 in four Eastern African laboratories assessing the true specificity of smear microscopy by examination of spiked artificial specimens.17 Nevertheless, the use of a more sensitive cut-off (4 AFB/100 HPF) compared to the standard 10 AFB cut-off may not result in more patients started on TB treatment under routine programme conditions, where patients with only scanty results are treated anyway, even if they are reported as being smear-negative.

In ideal conditions, when smears can be read on the day of specimen collection, an approach using a case definition based on one positive smear would result in two thirds of smear-positive patients starting treatment on the first visit. This may be valuable, given that some national TB programmes report patient drop-out rates during the diagnostic phase of as high as 37%.18 The time between initial consultation and final smear result was reduced using approaches based on two smear examinations compared to those with three smears. However, the translation of these shorter times to smear results into earlier treatment start was not addressed in this study, and requires further investigation.

In the present study, approaches based on two sputum specimens resulted in a significant reduction in the number of smears compared to those based on the collection of three specimens. This could alleviate the overwhelming workload of laboratories, particularly in countries with high HIV prevalence, which have experienced both a dramatic increase in TB cases and a human resource crisis. Reducing the number of sputum smears that need to be examined by one third would allow more time for examination of the remaining smears, may impact positively on the quality of smear microscopy and would finally increase the number of smear-positive TB cases detected. In addition, dropping the third smear will lead to a significant reduction in costs.19,20

In conclusion, the approach using one positive smear result after examination of two specimens in a high HIV prevalence setting increases the sensitivity of the method and reduces both the laboratory workload.
and diagnostic delays. Lowering the AFB threshold to define a smear-positive case would increase the number of smear-positive cases detected, and could be considered in settings with quality-assured smear microscopy.

**Acknowledgements**

The authors would like to thank Purity Muthoni Ndwigia, Ali Abdullahi and Mwangi Chege for their daily participation in the research project, and Laurence Bonte for providing precious technical support. They also thank the Blue House Clinic team in Mathare for their cooperation and Médecins Sans Frontières for sponsoring this research.

**References**


**CONTEXTE :** Centre de santé urbain, Nairobi.

**OBJECTIVE :** Evaluer l’impact de la réduction du nombre de crachats examinés et du seuil de positivité pour définir un frottis positif sur le taux de détection et la charge de travail au laboratoire.

**DESIGN :** Étude prospective. Examen microscopique (Ziehl-Neelsen) à l’aveugle de trois crachats de patients consécutifs suspects de tuberculose. La méthode standard (A), ≥2 frottis positifs sur 3, en utilisant un seuil de 10 bacilles acido-résistants (BAAR)/100 champs microscopiques (CM) a été comparée avec B, ≥2 frottis positifs (≥4 BAAR/100 CM) sur 3, un étant ≥10 BAAR/100 CM ; C, ≥2 frottis positifs (≥4 BAAR/100 CM) sur 3 ; D, ≥1 frottis positif (≥10 BAAR/100 CM) sur 2 ; et E, ≥1 frottis positif (≥4 BAAR/100 CM) sur 2. La détection d’au moins 1 frottis positif (≥4 BAAR/100 CM) sur 3 était l’examen microscopique de référence.

**RÉSULTATS :** Sur 644 suspects, les approches alternatives B et E ont détecté entre 114 (17,7%) et 123 cas (19,1%) par rapport à 105 cas (16,3%) pour A (P < 0,005). La sensibilité variait entre 82,0% (105/128) pour A et 96,1% (123/128) pour E. Les approches basées sur un frottis positif réduisaient de 36% le nombre de frottis par rapport à l’approche A.

**CONCLUSION :** La réduction du nombre de crachats et du seuil de positivité a augmenté la sensibilité de la microscopie et diminué la charge de travail.
RESUMEN

MARCO DE REFERENCIA: Dispensario urbano en Nairobi.

OBJETIVO: Evaluación de la repercusión de una reducción de la cantidad de muestras de esputo examinadas y de los valores discriminatorios que definen las baciloscopias positivas y los casos de tuberculosis (TB) sobre la detección de casos y el volumen de trabajo de los laboratorios.

MÉTODO: Estudio prospectivo con anonimato en el cual se examinaron, mediante coloración de Ziehl-Neelsen, tres muestras de esputo de cada paciente consecutivo con presunción clínica de TB. La estrategia (A), \( \geq 1 \) baciloscopias positivas de un total de 3, con un valor discriminatorio de 10 bacilos acidorresistentes (BAAR) por 100 campos con objetivo de gran aumento (CGA), se comparó con B, \( \geq 2 \) baciloscopias positivas de 3, con \( \geq 4 \) BAAR/100 CGA, una de las cuales con \( \geq 10 \) BAAR/100 CGA; C, 2 o 3 baciloscopias positivas de 3, con \( \geq 4 \) BAAR/100 CGA; D, \( \geq 1 \) baciloscopias positivas de 2, con \( \geq 10 \) BAAR/100 CGA; y con E, \( \geq 1 \) baciloscopias positivas de 2, con \( \geq 4 \) BAAR/100 CGA. El rendimiento diagnóstico se comparó con la estrategia de referencia: por lo menos 1 baciloscopia positiva de un total de 3 (\( \geq 4 \) BAAR/100 CGA).

RESULTADOS: En los 644 casos con presunción diagnóstica de TB, se detectaron entre 114 casos (17,7%) con la estrategia B y 123 casos (19,1%) con la estrategia E, en comparación con 105 casos (16,3%) detectados con la estrategia A (\( P < 0,005 \)). La sensibilidad osciló entre 82,0% (105/128) con la estrategia A y 96,1% (123/128) con la estrategia E. La supresión de una muestra de esputo disminuyó de 36% la cantidad de baciloscopias examinadas, en comparación con la estrategia de base A.

CONCLUSIÓN: La reducción de la cantidad de muestras y del umbral de positividad con el fin de definir un caso de tuberculosis aumenta la sensibilidad y reduce el volumen de trabajo de los laboratorios.