

# Quality control of sputum microscopic examinations for acid fast bacilli in southern Ethiopia

Estifanos Biru Shargie<sup>1,2</sup>, Mohammed Ahmed Yassin<sup>2,3</sup>, Bernt Lindtjørn<sup>1</sup>

## Abstract

**Background:** There has been an ongoing effort to strengthen the acid-fast bacilli (AFB) diagnostic laboratory network in the Southern Region of Ethiopia, and quality control has been initiated as part of this effort.

**Objective:** To evaluate the level of agreement in the readings of sputum smears for AFB between the peripheral diagnostic centers and the reference laboratories.

**Methods:** Blinded rechecking of routine slides collected from the peripheral diagnostic laboratories was undertaken in nine zones of the southern Ethiopian region between October 2000 and June 2002.

**Results:** Two thousand two hundred and nine slides, 54% of which were positive and 46% of which were negative slides, were collected from the peripheral diagnostic centres. Overall, false reading was 3.2% and the overall agreement was 96.8%. Among the 95 discordant slides that were re-examined at the central reference laboratory, 74% were found to be in agreement with the regional laboratory readings while 26% were in agreement with the peripheral readings. Nine zones participated in the quality control activity in 2000, six in 2001 and three in 2002.

**Conclusion:** Good agreement was recorded in the readings of AFB among the peripheral and reference laboratories. However, the false positive reading of 3.2% exceeds the recommended cut-off point of 2%. The number of zones participating decreased from time to time, and there seems to be an urgent need to re-vitalise the quality control initiative. Regular on the job training of the peripheral laboratory staff together with supportive supervision would help improve their performance. [*Ethiop.J.Health Dev.*2005;19(2):104-108]

## Introduction

Case finding through sputum microscopy is one of the five elements of the directly observed therapy, short course (DOTS) strategy- the WHO strategy for the management of tuberculosis (TB) (1). Thus, TB control requires a functional laboratory set-up with quality diagnostic services and a trained diagnostician and a microscopist (2). However, the performance of such laboratories depends on continuous monitoring and quality improvement mechanisms put in place (3). The principal objective of external quality assessment (EQA) is to standardise sputum microscopy for the detection of infectious TB cases across the peripheral diagnostic centers and to validate the reported acid-fast bacilli (AFB) microscopy results from these centers. The National Tuberculosis and Leprosy Control Programme (TLCP) manual for laboratory technicians recommends the rechecking of a sample of routine slides as a method of ensuring the quality of diagnosis and classification of TB patients within the health services (4).

The aim of this study was to evaluate the performance of sputum microscopic examinations in the peripheral laboratories of the SNNPRS.

## Methods

In the Southern Nations, Nationalities and Peoples' Regional State (SNNPRS), 167 health facilities are engaged in the diagnosis of TB (5). Over the past seven years, efforts have been made to strengthen the AFB laboratory network in the region. In addition to the

provision of microscopes and laboratory supplies to the diagnostic laboratories, a quality control unit was established within the regional reference laboratory in Awassa, the regional capital. As a result, there has been an ongoing quality control of Ziehl-Neelsen (ZN) stained slides collected on a quarterly basis from the peripheral laboratories by the zonal and woreda (district) TLCP co-ordinators.

## Patients

The main source of data for this cross-sectional study was the quality control register at the regional TLCP office. The ZN stained slides that were collected for quality control from the peripheral diagnostic centers within the region from October 2000 through June 2002 formed the study units. Variables such as sputum smear results in the peripheral diagnostic centers versus the regional and national reference laboratories, the proportion of positive slides, and the consistency of the quality control activities were investigated.

## Methods

The procedures employed for the quality control were in line with the recommendations of the National TLCP (2). In brief, the peripheral laboratories retained all positive slides and an equal number of negative slides were kept in dark places until the Zonal TLCP co-ordinators collected the required number of slides. The number of slides to be collected was determined based on the number of smear positive slides seen in the same quarter. Zonal TLCP co-ordinators submitted the collected slides

<sup>1</sup>Center for International Health, University of Bergen, Armauer Hansen Building, N-5021 Bergen, Norway; <sup>2</sup>Southern Nations, Nationalities and Peoples' Regional State Health Bureau, P. O. Box 149, Awassa, Ethiopia; <sup>3</sup>Liverpool School of Tropical Medicine, Pembroke place, L3 5QA, Liverpool, UK

to the regional programme co-ordinator with the results filled in a specially designed form. The regional TLCP co-ordinator filled out slide numbers in a new set of forms without the results, and submitted the slides and the forms to the regional reference laboratory.

The laboratory technicians in the reference laboratory examined the slides and reported the results to the regional TB program unit. The microscopists in the regional laboratory did not observe the initial results of the smears at the peripheral diagnostic centers. The regional TLCP coordinator compared the results of the peripheral laboratories against the regional reference laboratory and sent out all discordant slides to the national reference laboratory for re-reading (4,6).

The regional TLCP co-ordinator maintained the results of the peripheral as well as the regional and central laboratory readings. Feedback on the smearing and staining quality together with the final AFB result at the reference laboratory was sent regularly to the peripheral diagnostic laboratories.

In case of agreement between the peripheral and the regional laboratory readings, the regional laboratory results were taken as final results. Slides with discordant readings were re-read at the central reference laboratory (6), and the results from the central reference laboratory were considered as final results. The sensitivity, specificity, as well as the positive (PPV) and negative predictive values (NPV) of the peripheral readings were calculated against the final readings of the reference laboratories as a gold standard.

#### Statistics and data analysis

Data were analysed using the Statistical Package for Social Sciences (SPSS) 11.0 (7) software. In this study, discordant slides were defined as positive slides read as negative or vice versa, as the positive slides were not quantified (graded) either at the peripheral or regional laboratories. Sensitivity, specificity, predictive values, and proportions of false readings were compared using the  $\chi^2$  test and 95% confidence intervals for proportions. The level of agreement among various diagnostic levels was measured using Kappa. The level of significance was set at 5%.

Table 2: Agreement in readings of slides between peripheral diagnostic centers and the Regional reference laboratory, and the final rereading

	Regional Reference Laboratory			Final Rereading <sup>§</sup>		
	Positive slides	Negative slides	Total slides	Positive slides	Negative slides	Total slides
Peripheral laboratory:						
Positive slides	1136	46	1182	1147	32	1179
Negative slides	53	957	1010	38	971	1009
<b>Total slides</b>	<b>1189</b>	<b>1003</b>	<b>2192</b>	<b>1185</b>	<b>1003</b>	<b>2188</b>

§ In case of agreement between the peripheral and regional readings, the regional reading was taken as the final reading; in case of discordance, the central reference lab reading was taken as the final rereading.

#### Results

Two thousand and two hundred nine slides were collected for checking quality and from the peripheral diagnostic centers in nine zones of the southern region. A median of 257 (ranging from 17 to 475) slides were collected from each zone. Fifty four percent (1,184) of the collected slides were positive at the peripheral diagnostic centers and 46% (1,025) were negative. Table 1 shows the number of slides collected each year and the proportion of positive slides. The number of zones participating in EQA decreased from nine in 2000 to 6 in 2001 and to three in 2002.

Table 1: Zones that participated in the External Quality Assessment and the number of slides collected each year, southern Ethiopia 2000-2002

Zone	Slides collected, n (% positive)		
	2000	2001	2002*
Guraghe	151 (52.3)	-	-
Hadiya	157 (58.0)	118 (55.1)	97 (55.7)
Kambata-Alaba-Tambaro	171 (48.0)	167 (45.5)	37 (51.4)
Sidama	105 (54.3)	152 (57.9)	-
Wolaita	142 (69.0)	149 (52.3)	150 (50.0)
Gamo-Goffa	51 (51.0)	41 (58.5)	-
Gedeo	256 (51.6)	219 (52.5)	-
Bench-Maji	17 (58.8)	-	-
Sheka	29 (51.7)	-	-
<b>Total</b>	<b>1079 (54.7)</b>	<b>846 (52.7)</b>	<b>284 (52.1)</b>

\* Data collected for the first two quarters of the fiscal year.

Seventeen (1%) of the slides had poor smearing and/or staining quality. Thus, 2192 slides were re-read by the regional laboratory team, and 99 (4.5%) of them were found to be discordant. The discordant slides were sent to the central reference laboratory where four of the slides could not be read due to fading or poor staining. Ninety-five slides were re-read by the microscopists at the central reference laboratory, and the readings of 70 (74%) of the slides were found to be concordant with that of the regional laboratory. The readings of the remaining 25 (26%) of the slides were also found to be concordant with that of the peripheral diagnostic centers (Table 2).

Considering the final reading result as the gold standard for the true smear result, both the sensitivity and specificity of the peripheral diagnostic laboratories was 96.8%, and the PPV and NPV were 97.3% and 96.2 % respectively. While the false readings at the peripheral laboratories in the region were 3.2% (range 0-5.3), there was a high degree of agreement (96.8%) between the readings at the peripheral diagnostic laboratories and the final reading (K = 0.936, SE=0.008).

The proportion of false-positive slides declined from 4.4% (21/475) in the year 2000 to 2.3% (9/394) in 2001

and to 1.5% (2/134) in mid-2002 ( $\chi^2_{\text{trend}}=4.33$ ;  $p=0.038$ ). Meanwhile, there was a reduction of false-negatives from 3.9% in the year 2000 to 2.4 and 2.6 in the years 2001 and 2002 respectively ( $\chi^2_{\text{trend}}=1.39$ ;  $p=0.238$ ) (Figure 1).

The sensitivity and specificity of sputum microscopy at the peripheral diagnostic centers showed little variations across the zones (Table 3). The diagnostic centers in Wolaita, Kambata-Alaba-Tambaro and Hadiya zone exhibited the highest sensitivity; whereas those in Sidama, Kambata-Alaba-Tambaro and Hadiya exhibited the highest specificity.

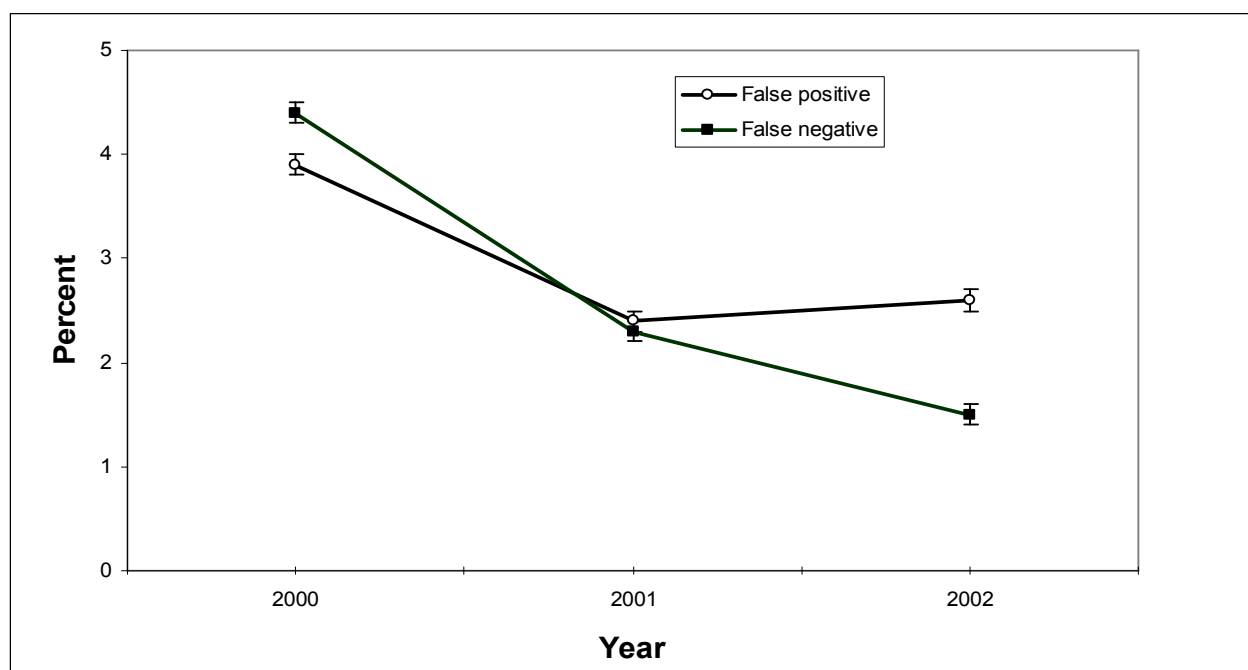


Figure 1: Trends in false readings of sputum smears at the peripheral diagnostic centers, southern Ethiopia (the bars indicate 95%CI for the proportions)

Table 3: Sensitivity, specificity and predictive values of sputum microscopy at the peripheral diagnostic centers of Southern region, 2000-2002

Zone	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Measure of agreement Kappa (SE)
Guraghe (n=149)	95.0 (90.2-99.8)	95.7 (90.8-100)	96.2 (90.0-99.0)	94.3 (86.8-98.2)	0.906 (0.035)
Hadiya (n=371)	97.6 (95.6-99.7)	96.9 (94.2-99.6)	97.6 (94.8-99.1)	96.9 (93.3-98.9)	0.945 (0.017)
Kambata-Alaba-Tambaro (n=370)	98.9 (97.3-100)	97.9 (96.0-99.9)	97.7 (94.6-99.3)	99.0 (96.6-99.8)	0.967 (0.013)
Sidama (n=254)	96.0 (92.9-99.1)	99.0 (97.2-100)	99.3 (96.6-100)	94.5 (88.9-97.7)	0.943 (0.021)
Wolaita (n=438)	98.8 (97.4-100)	95.9 (93.1-98.7)	96.8 (94.0-98.5)	98.4 (95.8-99.6)	0.949 (0.015)
Gamo-Goffa (n=92)	96.0 (90.6-100)	95.2 (88.8-100)	96.0 (87.4-99.34)	95.2 (85.1-99.2)	0.912 (0.043)
Gedeo (n=469)	93.7 (90.7-96.7)	95.8 (93.2-98.5)	96.3 (93.4-98.2)	92.8 (88.8-95.7)	0.893 (0.021)
Bench-Maji (n=16)	100 (71.7-100)	100 (65.2-100)	100 (71.7-100)	(65.2-100)	1.000
Sheka (n=29)	100 (81.9-100)	100 (80.7-100)	100 (81.9-100)	100 (80.7-100)	1.000
<b>Total (n=2188)</b>	<b>96.8 (95.8-97.8)</b>	<b>96.8 (95.7-97.9)</b>	<b>97.3 (96.2-98.1)</b>	<b>96.2 (94.9-97.3)</b>	<b>0.936 (0.008)</b>

CI= confidence interval; PPV = Positive predictive value; NPV = Negative predictive value

## Discussion

This study shows the existence of a good agreement between slides coming from all zones, and the proportions of false readings were seen to decline from year to year. However, despite efforts to strengthen and expand quality assurance schemes in the region, only nine of the 13 zones participated in the EQA program. Further more only three of nine zones sent slides for quality check in 2002 and the number of slides collected for EQA decreased by 25% in 2001 and by 75% in 2002. This indicates that there is an urgent need to re-vitalise the quality assurance activities within the region.

Rechecking of routine slides could be used interchangeably or in conjunction with proficiency testing. Although the former is said to be more laborious, it represents the routine performance (8). Laboratories that send slides for rechecking are more likely to perform better (9). Sometimes, the rechecking of routine slides is not blinded and this might bias the rereading result (9). The rechecking in this study setting was performed under blinded conditions, and the probability of such bias is also minimal. However, sampling bias in the selection of slides for rechecking is potentially unavoidable; technicians at peripheral laboratories may tend to retain slides with good quality stains and readings.

Furthermore, many peripheral diagnostic centers were not involved in the EQA. Some laboratories may not submit slides for rechecking simply due to lack of confidence on their performance or because of uncertainty about the possible implications in case their performance is not satisfactory. To this end, the findings in this study might not reflect the overall performance in the region.

The present study could not allow the study team to measure quantification errors. As a result, it was not possible to estimate the exact degree of discordance between the peripheral and reference laboratories. However, quantification does not have implications for case management.

Despite these limitations, this study shows the reality under routine conditions.

The overall agreement between the readings of the peripheral diagnostic centers and that of the reference laboratory is comparable to findings from other studies (9,10). However, these studies used a different mechanism of EQA, namely proficiency testing- by sending a centrally prepared panel of stained smears to peripheral laboratories. Some studies have reported false-positive results of as high as 6.9% and a false-negativity of 3.4% after re-staining from the routine slides before rechecking (11).

Considering patient management, the finding of a 97% agreement between the peripheral and final readings is encouraging. Nevertheless, the 3.2% false-positive reading is larger than the critical cut-off point of 2% set by the NTLCP. The proportion of false negatives is below the national threshold for remedial action (3.2% vs. 5%) (4).

The documentation of more than 25% discordance between the regional and central reference laboratory readings is also point of concern. Slides sent to the central laboratory because of discordance may have few bacilli in the smear. As the chance of making errors is higher with the low-positive slides, this might partly explain the unacceptably high discordance existing between the two centers. Such slides might be read, for example, by the peripheral reader as positive; but the regional microscopist may read another 100 fields of the same slide that do not contain any bacilli. The central reader, being conscious that these slides are discordant ones, may search more vigorously and find some bacilli. However, we cannot justify this assertion, as quantification was not done as part of the quality control. It may thus be appropriate to revise the national guideline to include quantification as part of the routine AFB microscopy and the quality control exercise.

Furthermore, taking the reference laboratory result as a gold standard has its own limitations. First, there is a chance of fading of stains by the time they reach the central laboratory as was seen in some African settings (11). Slight fading might affect the readings at the central level, especially for the low-positives and result in false-negative readings. Secondly, there is still a chance of making error with low-positive slides even at the central level. Systematic restaining of peripheral smears as was done in Burundi (11) could help in identifying some false-negative slides that would have otherwise been left unidentified due to fading of stains. In addition, well-planned periodic proficiency testing might help in assessing the performance of microscopists at different levels of the laboratory network.

The expansion of the quality control scheme to all peripheral laboratories with regular supervisory visits, the provision of timely feedback and regular refresher training programmes are thus highly recommended.

## References

1. World Health Organization. Towards a TB-free future: (pre-publication issue). Geneva, Switzerland. World Health Organization, 2001.
2. Manual of National Tuberculosis and Leprosy Control Programme. Addis Ababa, Eth. Ministry of Health of Ethiopia, 2002.

3. World Health Organization. Framework for effective tuberculosis control. WHO/TB/94-197. WHO Geneva, Switzerland. 1994.
4. National Tuberculosis and Leprosy Control Programme. Manual for Laboratory Technicians. NTLCP, Addis Ababa, Eth. 1999.
5. Southern Nations Nationalities and Peoples Regional State Health Bureau. Annual TB and Leprosy Control Programme Report. SNNPRS Health Bureau, Awassa, Ethiopia, 2002.
6. Tesema D. National Quality Assurance Manual for Ethiopia. Ethiopian Health and Nutrition Research Institute, Addis Ababa, Eth, 2003.
7. SPSS for window. Standard version. Release 11.0.0. SPSS inc., 2001.
8. Van Deun A. External quality assessment of sputum smear microscopy: A matter of careful technique and organisation (editorial). *Int J tuberc Lung Dis* 2003;7:507-508.
9. Martinez-Guarneros A, Balandrano-Campos S, Solano-Ceh M A, et al. Implementation of proficiency testing in conjunction with a rechecking system for external quality assurance in tuberculosis laboratories in Mexico. *Int J Tuberc Lung Dis* 2003;7:516-521.
10. Paramasivan CN, Venkatarama P, Vasanthan JS, Rahman F, Narayanan PR. Quality assurance studies in eight state tuberculosis laboratories in India. *Int J Tuberc Lung Dis* 2003;7:522-527.
11. Buzingo T, Sanders M, Masabo JP, Nyandwi S, van Deun A. Systematic re-staining of sputum smears for quality control is useful in Burundi. *Int J Tuberc Lung Dis* 2003;7:439-444.

