

Comparison of Macroscopic and Microscopic Assessment of Specimens Collected for the Diagnosis of Tuberculosis

Sayera Banu¹, Shahed Hossain¹, Mohammad Khaja Mafij Uddin¹, Md. Toufiq Rahman¹, Razia Khatun¹, K Zaman¹, M A Quaiyum¹ and Frank van Leth^{*,2,3}

¹International Centre for Diarrhoeal Disease Research, Bangladesh

²KNCV Tuberculosis Foundation, The Hague, The Netherlands

³Department of Global Health, Academic Medical Centre, University of Amsterdam, Amsterdam Institute for Global Health and Development, The Netherlands

Abstract: Diagnosis of tuberculosis in a field setting depends on the quality of specimens submitted for smear-microscopy. Macroscopic assessment (sputum or saliva) of the specimen for suitability for further examination is common practice in routine care. We examined whether macroscopic assessment could correctly identify sputum specimens based on four published algorithms using microscopic features in the setting of active case finding in a community survey.

The study included 901 randomly selected adults who reported cough for 3 weeks or more in the national tuberculosis prevalence survey in Bangladesh. A single specimen of each was assessed with microscopy and microscopy (Gram-stain) to classify it as either sputum or saliva. The primary outcome was the agreement between the two assessment methods (Kappa statistic)

From 901 specimens, 561 (62%) were macroscopically classified as saliva and 340 (38%) as sputum. From these, 888 Gram-stained slides could be examined for microscopic features. The agreement between the macroscopic assessment with any of the four microscopy algorithms for sputum was very poor (all Kappa's below 0.1).

While macroscopic assessment of submitted specimens might be of value in routine care, it is not warranted in a setting of active case finding in a community survey. Submitting a specimen in the first place should be the primary goal in this setting.

Keywords: Macroscopy, microscopy, sputum quality, survey, tuberculosis.

INTRODUCTION

Tuberculosis (TB) is a disease of global importance. One-third of the world's population is estimated to have been infected with *Mycobacterium tuberculosis* and eight million new cases of tuberculosis arise each year [1]. TB diagnosis can only be made reliably by demonstrating the presence of tubercle bacilli in the sputum by means of microscopy and/or culture in the laboratory. The gold standard for diagnosing pulmonary tuberculosis is culture of a sputum specimen. However, due to lack of access to culture facilities and the long turn-around times involved with sputum culture, the cornerstone of the diagnosis of tuberculosis is still the microscopic examination of sputum specimens for acid-fast bacilli (AFB) in many developing countries including Bangladesh.

Under this condition it is very important that a high quality sputum specimen is available for microscopic examination in order to diagnose TB accurately. Patients frequently provide pure saliva or very small amounts of sputum in saliva instead of an appropriate, purulent, sputum

specimen. Saliva smears are in general more likely to be negative or have an AFB density below the threshold of microscopy detection than sputum smears [2]. A negative result issued on examination of such a specimen is misleading, since it implies that a correct sputum specimen has been examined. This holds particularly true in settings where AFB microscopy is used in the context of large-scale TB prevalence surveys, where participants are less likely to be able to produce an adequate specimen.

These surveys are currently being prepared or implemented in a large number in countries in an attempt to provide data for the assessment of the progress on goal 6 of the Millennium Development Goals [3].

Macroscopically, a good sputum specimen consists of recently-discharged material from the bronchial tree, with minimum amounts of oral or nasal material. Satisfactory quality implies the presence of mucoid or mucopurulent material and is of greater significance than volume. Poor quality specimens are thin and watery or composed largely of bubbles. It is common practice in many TB diagnostic centres that sputum provided for AFB examination is first assessed macroscopically by laboratory staff. If the quality of the sputum is deemed inadequate, the specimen is rejected and the TB suspect asked to provide a new specimen. Providing sequential sputum specimens in a short period of

Address correspondence to this author at the Amsterdam Institute for Global Health and Development, Pietersbergweg 17, 1105BM Amsterdam, The Netherlands; Tel: +31 20 566 1593; E-mail: f.vanleth@aighd.org

time is considered difficult by many TB suspects. The risk exists that the TB suspect will not return at a later stage to provide a sputum specimen leading to a missed opportunity for TB case finding.

A macroscopic assessment of specimen quality must be done by skilled laboratory staff involved but can be subjective.

Microscopic specimen assessment, in which smears are examined and graded according to an algorithm, has been proposed as a means to ensure sufficient specimen quality. Further specimens can be then requested if quality is inadequate [4]. However, this is much more time consuming for the laboratory. There are several published algorithms for microscopic specimen assessment. Among the defined criteria, Murray & Washington [5] and Van Scoy [6] consider the specimens as saliva or sputum based on only the number of white blood cell (WBC) or epithelial cells (EPI) per low powered field (LPF). However, variations in the thickness of material in different areas of the slide may require extensive examination to obtain an overall average for each slide. To minimize the variability, the other two criteria (Barry [7] and Gal-Oz [8]) involve assessment of the ratio of WBC to EPI in several areas of the slide. It has been shown earlier that in the context of diagnosing respiratory infections the different microscopic criteria performed in a comparable manner [4].

These algorithms were developed to assess if it was worthwhile to culture a respiratory specimen. Using these algorithms for assessing the possibility for adequate AFB microscopy has been much less examined. A previous study on the usefulness of macroscopic assessment of respiratory specimens for AFB diagnosis was carried out in a routine TB diagnostic setting [9]. This implies a passive case finding strategy in which individuals report themselves to a health facility because of symptoms. The current study is part of a TB prevalence survey which, by definition, entails an active case finding strategy in which individuals are approached and asked for symptoms. Investigators do not have the possibility to return to the individual at a later stage in such a setting. It is therefore pertinent that any decision on rejecting a specimen for investigation is made on solid arguments in this brief encounter. The objective of this study was to assess if these arguments should be based on macroscopic or microscopic assessment of respiratory specimens. Several countries will be embarking on similar large-scale TB prevalence surveys in the near future. The result of the present study will be able to guide field activities in these surveys in relation to the handling of submitted specimens.

METHODS

Setting and Population

The nation-wide TB prevalence survey was implemented during 2007-09 in which a representative sample size of 52,089 adults ≥ 15 years were divided over 40 (20 rural and 20 urban) randomly selected clusters [10]. Each participating adult (15 years or older) provided two sputum specimens (one spot and one morning) regardless of symptoms during a household survey. The specimens were kept cool until transport to the field laboratory. Specimens were collected from several locations at the same time during the survey

and arrived in a random fashion in the field laboratory. In this sub-study all survey participants who reported cough for 3 weeks or more were eligible for inclusion. The study was restricted to the last 21 clusters of the total 40 clusters for logistical reasons.

Field Activities

Each collected specimen was scored macroscopically as saliva or sputum by the trained field laboratory technicians at the time of arrival of the specimen in the field laboratory. Based on National Tuberculosis Control Programme guidelines, a specimen was considered saliva on visual assessment if it had a clear, watery appearance and contained no purulent material [11]. Two smears from each of the fresh specimens (uncentrifuged) were prepared at the field laboratories, one for AFB and another for Gram staining. The slides prepared for Gram staining were heat fixed at the field laboratories and transported to the laboratory of the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR, B). Specimens were refrigerated at the field sites and transported to the central laboratory for culture using conventional method within 48 hours of specimen collection. No preservatives were added.

Laboratory Activities

The prepared slide of the morning specimen of each enrolled individual in the sub-study was Gram stained. If there was no morning specimen available, the prepared slide of the spot specimen was used for Gram staining. Slides were stained with crystal violet and then treated with an I₂-KI mixture (mordant) to fix the stain, washed briefly with 95% alcohol (destained) and finally counterstained with carbol fuchsin. The stained slides were examined under low-power (X100) magnification by the microscopist on duty. After examining 10 fields, two parameters were recorded: the average number of polymorphonuclear neutrophils (PMNs) and squamous epithelial cells (SEC) per LPF. Four different algorithms (Table 1) were used for classifying a specimen as sputum or saliva based on microscopic features; (1) Murray & Washington: based on the average number of SECs; (2) Van Scoy: based on polymorphic cells; (3) Barry: based on epithelial and polymorphic cells; (4) Gal-Oz: Sputum quality was assessed by the presence of PMNs and SECs per LPF.

Statistical Analysis

The data were analyzed using SPSS v 17.0. The agreement between the scoring algorithms and macroscopic assessment was done by using the κ (kappa) statistic, where the κ value of > 0.4 was considered to be adequate concordance.

RESULTS

Of the 901 specimens available in the study, we were able to obtain complete macroscopic and microscopic assessment data from 888 (98.56%). The quality of the remaining 13 slides was inadequate for proper reading.

Among the 561 (63%) macroscopically saliva specimens, 282 (51.4%) and 299 (54.5%) were classified as sputum by Murray & Washington and Gal-Oz respectively. Among 340 (37%) specimens classified macroscopically as sputum, 187 (55.2%) and 198 (58.4%) were also sputum by both of the

Table 1. Summary of Four Published Criteria for Judging Acceptability of Sputum Specimens

Author	Method	Criteria for Acceptability
Murray & Washington	Average no. of EPI/LPF	<10 EPI/LPF
Van Scoy	Average no. of WBC/LPF	>25 WBC/LPF
Barry	Assign + and – values, 3+ if > 150 WBC/LPF; 2+ if 76-150 WBC/LPF; 1+ if 1-75 WBC/LPF; -3 if >25 EPI/LPF; -2 if 16-25 EPI/LPF; -1 if 5-15 EPI/LPF	Any positive score (sum of + and – values)
Gal-oz	Informative: <10 SEC/LPF & >25 PMNs/LPF Semi-informative: <10 SEC/LPF or >25 PMNs/LPF Uninformative: >10 SEC/LPF & <25 PMNs/LPF	(Semi) informative considered to be Sputum

criteria. There was much poorer agreement between macroscopic assessment and the algorithms of Van Scoy and Barry (Table 2). All kappa statistics were below 0.1 indicating a very poor agreement between macroscopic and microscopic assessment of the specimens.

Table 2. Agreement Between Macroscopic and Microscopic Assessment of Sputum Specimens

Sputum According to	Saliva (n=561)	Sputum (n=340)	Kappa
Murray & Washington	282 (51.4%)	187 (55.2%)	0.04
Van Scoy	2 (0.4%)	7 (2.1%)	0.02
Barry	126 (23.0%)	98 (28.9%)	0.06
Gal-Oz	299 (54.5%)	198 (58.4%)	0.04

All specimens in this sub-study were microscopically negative for AFB, but five of them were found to be positive on conventional culture. If culture of the specimens was conditional of the screening criterion used, then macroscopic assessment of the specimen would have missed one culture positive for TB, the microscopic criterion of Murray & Washington, Barry, and Gal-Oz two, and the criterion of Scoy all five.

DISCUSSION

The use of Gram-stained smears to assess the quality of sputum specimens has received considerable attention as a means for improving the reliability of sputum microscopy as well as sputum culture for tuberculosis diagnosis [12]. The evaluation of the quality of sputum using cytological parameters is very important but requires experience and qualification and is also time and resource-consuming.

This study showed that the agreement between macroscopic and microscopic assessment of sputum specimens was poor in the setting of an active case-finding strategy of a TB prevalence survey. Assuming the microscopic classification as a gold standard, this indicates that macroscopic assessment of provided specimens could not properly distinguish between saliva and sputum. In the national TB prevalence survey a total of 33 new smear-positive TB cases were detected among 52,089 population [10]. Out of these 33 cases the morning sputum specimens collected from 19 (57.6%) were macroscopically classified

as sputum whereas the remaining 14 (42.4%) were saliva. These two observations make that rejecting specimens based on the macroscopic assessment in this setting is not recommended since it will underestimate the prevalence of smear-positive specimens.

This study also showed that microscopic sputum assessment as a procedure did not contribute to the identification of specimens that were culture positive. Apparently it is more important that an actual specimen is submitted for sputum examination than that the submitted specimen is meeting specific macroscopic or microscopic qualities. From this it follows that survey personnel should be trained in obtaining specimens for subsequent microscopy, especially in patients who have difficulties in coughing up material. Recent studies showed that instruction given for a good quality sputum specimen increased both the quality and quantity of the sputum specimen with an increased positivity rate [13].

The poor performance of macroscopic assessment found in our study differs from the observation by Kahn *et al.*, who stated that macroscopic assessment was a valid approach for identification of smear-positive respiratory specimens [9]. This conclusion was based on the fact that macroscopic assessment rejected the least number of specimens in this study. As mentioned earlier, the main difference between the two studies is the setting, which indicates that operating procedures should be tested for validity in the setting they are being used rather than taken as face-value. Since the physicians are primarily responsible for submitting proper specimens to a laboratory, the use of the gram stain might be of some value in isolated instances where one needs to know whether specimens from patients suspected of having mycobacterial disease, but who have consistently negative smears and cultures, are of lower respiratory origin. In all other situations, including that of active case finding in a TB prevalence survey, efforts should be directed to obtain a specimen for processing.

These findings should be considered in the design of large-scale national TB prevalence surveys in other countries. Survey protocols tend to incorporate procedures derived from routine patient care. Without careful considering the different setting of the survey, this might lead to biased measurements in the field and the reporting of invalid prevalence estimates. This observation does not hold true only for TB prevalence surveys but can be considered

for ongoing community surveillance systems in the field of respiratory disease in general.

ACKNOWLEDGEMENTS

This work was supported by grants from the KNCV Tuberculosis Foundation, The Netherlands, World Health Organization, and the International Centre for Diarrhoeal Disease and Research, Bangladesh (ICDDR, B).

CONFLICT OF INTEREST

None of the authors reports a conflict of interest.

ABBREVIATIONS

AFB	=	Acid-fast Bacilli
EPI	=	Epithelial cell
LPF	=	Low power field
PMN	=	Polymorphonuclear neutrophil
SEC	=	Squamous epithelial cell
TB	=	Tuberculosis

REFERENCES

- [1] Global Tuberculosis Control 2010. Geneva, Switzerland: World Health Organization
- [2] Enarson D, Rieder H, Arnadottir T, Trebuch A. Technical guide: sputum examination for tuberculosis by direct microscopy in low income countries. Paris: International Union Against Tuberculosis and Lung Disease, 2007.
- [3] WHO Global Task Force on TB Impact Measurement [Internet]. WHO. Available at: http://www.who.int/tb/advisory_bodies/impact_measurement_taskforce/en/ [cited 2012 Jan 4].
- [4] Wong LK, Barry AL, Horgan SM. Comparison of six different criteria for judging the acceptability of sputum specimens. *J Clin Microbiol* 1982; 16: 627-31.
- [5] Murray PR, Washington JA. Microscopic and bacteriologic analysis of expectorated sputum. *Mayo Clin Proc* 1975; 50: 339-44.
- [6] Van Scoy RE. Bacterial sputum cultures. A clinician's viewpoint. *Mayo Clin Proc* 1977; 52: 39-41.
- [7] Barry AL. Clinical specimens for microbiologic examination. In: Hoepfich PD (ed). *Infectious diseases*. New York: Harper & Row Publishers 1978; 92-6.
- [8] Gal-Oz A, Kassis I, Shprecher H, Beck R, Bentur L. Correlation between rapid strip test and the quality of sputum. *Chest* 2004; 126(5): 1667-71.
- [9] Khan MS, Dar O, Tahseen S, Godfrey-Faussett P. Judging respiratory specimen acceptability for AFB microscopy: visual vs microscopic screening. *Trop Med Int Health* 2009; 14: 571-5.
- [10] Zaman K, Hossain S, Banu S, et al. Prevalence of smear-positive tuberculosis in persons aged ≥15 years in Bangladesh: results from a national survey, 2007-2009. *Epidemiol Infect* 2011; [Epub ahead of print].
- [11] National Tuberculosis Control Program. Laboratory Manual on Smear Microscopy for Tuberculosis and its Quality Control in the NTP of Bangladesh. Dhaka: National Tuberculosis Program 2003.
- [12] Curione CJ Jr, Kaneko GS, Voss JL, Hesse F, Smith RF. Gram stain evaluation of the quality of sputum specimens for mycobacterial culture. *J Clin Microbiol* 1977; 5: 381-2.
- [13] Khan MS, Dar O, Sismanidis C, Shah K, Godfrey-Faussett P. Improvement of tuberculosis case detection and reduction of discrepancies between men and women by simple sputum-submission instructions: a pragmatic randomised controlled trial. *Lancet* 2007; 369: 1955-60.

Received: December 12, 2011

Revised: January 6, 2012

Accepted: January 6, 2012

© Banu et al.; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.