

# **Performance of a novel algorithm using automated digital microscopy for diagnosing tuberculosis**

Nazir A Ismail<sup>1,2</sup>, Shaheed V Omar<sup>1</sup>, James J Lewis<sup>3</sup>, David W Dowdy<sup>4</sup>, Andries W Dreyer<sup>1,5</sup>, Hermina van der Meulen<sup>6</sup>, George Nconjana<sup>1</sup>, David A Clark<sup>6</sup>, Gavin J Churchyard<sup>3,6,7</sup>

<sup>1</sup>National Institute for Communicable Diseases, South Africa

<sup>2</sup>University of Pretoria, South Africa

<sup>3</sup>London School of Hygiene and Tropical Medicine, UK

<sup>4</sup>Bloomberg School of Public Health, Johns Hopkins University, USA

<sup>5</sup>University of Witwatersrand, South Africa

<sup>6</sup>Aurum Institute, South Africa

<sup>7</sup>School of Public Health, University of Witwatersrand, South Africa

## **Corresponding author:**

Dr Nazir Ismail, Centre for Tuberculosis, National Institute for Communicable Diseases,

1 Modderfontein Road, Sandringham 2031, South Africa

## **Abstract:**

Rationale: TBDx automated microscopy is a novel technology that processes digital microscope images to identify acid-fast bacilli (AFB). TBDx as part of a diagnostic algorithm could improve the diagnosis of tuberculosis (TB), but its performance characteristics have not been formally tested.

Objective: To evaluate the performance of the TBDx automated microscopy system in algorithms for diagnosis of TB.

Methods: Prospective samples from patients with presumed TB were processed in parallel with conventional smear microscopy, TBDx microscopy, and liquid culture. All TBDx positive specimens were also tested with Xpert MTB/RIF (GXP). We evaluated the sensitivity and specificity of two algorithms: 1) TBDx-GXP (TBDx, with positive specimens tested by Xpert MTB/RIF) and 2) TBDx alone, against the gold standard liquid-media culture.

Results: Of 1210 samples, 1009 were eligible for evaluation, of which 109 were culture positive for *Mycobacterium tuberculosis*. The TBDx system resulted 70 specimens (68 culture-positive) as having  $\geq 10$  putative AFB (high positive) and 207 (19 culture-positive) as having 1-9 putative AFB (low positive). An algorithm in which “low positive” results on TBDx were confirmed by GXP had 78% sensitivity (85/109) and 99.8% specificity (889/900), requiring 21% (207/1009) specimens to be processed by GXP. As a stand-alone test, a “high positive” result on TBDx had 62% sensitivity and 99.7% specificity.

Conclusions: TBDx used in diagnostic algorithms with GXP provided reasonable sensitivity and high specificity for active TB while dramatically reducing the number GXP tests performed. As a stand-alone microscopy system, performance was equivalent to a highly experienced TB microscopist.

**Keywords:**

Tuberculosis, Diagnosis, Microscopy, Automated, TBDx, Xpert, MTB/RIF, Triage

## **INTRODUCTION**

An estimated 1 billion people have died of tuberculosis (TB) in the past 200 years, making it the leading infectious killer in known history(1) . Despite this shameful history, the diagnosis of TB has relied, until recently, on sputum smear microscopy. This is a technique that is nearly 100 years old and misses half of all people with active TB with operator-dependent sensitivities ranging between 30% and 80% in different epidemiological settings(2, 3).

In the past decade, important advances have been made in TB diagnosis, notably the introduction of the Xpert MTB/RIF assay(GXP, Cepheid, Sunnyvale, USA),which has a sensitivity for pulmonary TB of 89% (approaching 100% among those with smear-positive TB but reduced to 68% in smear negative individuals) and a specificity of 99%(4).

The cumulative worldwide volume of Xpert MTB/RIF (GXP) tests conducted has reached over 6.2 million since its wide implementation in 2011 (5). However, this is far short of the estimated 77.6 million microscopy tests performed annually in the public sector of just the 22 high TB burden countries (6). One of the primary reasons for incomplete uptake of GXP is its cost (\$9.98 per cartridge under concessional pricing)(7), which in most settings is several fold higher than that of microscopy (6) on a per-test basis and well beyond the existing TB budgets of most high-burden countries(8). Diagnostic “triage” algorithms that can reduce the number of GXP tests conducted without substantially lowering diagnostic performance are now an important priority (9).

Automated microscopy systems for TB diagnosis have the potential to fill this important diagnostic niche. Use of microscopy as a platform has the advantage of very durable,

low-cost equipment that is already available for TB diagnosis throughout the world (6). Use of automated systems (rather than human eyes) to read microscopy slides can reduce subjectivity in results and potentially improve performance. The TBDx automated microscopy system (Signature Mapping Medical Sciences, Virginia, USA) is a computer-aided detection system that automatically recognises and counts images of putative acid-fast bacilli (AFB) in digitized fields of view.

A previous proof of concept publication(10) evaluating an earlier version of the TBDx system showed a sensitivity of 49% and specificity of 98.9%, when it was combined in a test and confirm algorithm with human microscopy confirming slides reported with 1-9 AFB on the TBDx system. Use of TBDx in this triage approach reduced the workload of microscopists by 47%. However, this study noted that the use of GXP as a confirmatory test, rather than manually read microscopy, may improve the sensitivity of the TBDx based algorithm without compromising specificity.

Based on these early findings, improvements to the system have been made by the inclusion of a stepwise classification (SWC) algorithm which performs post-processing categorization of false positive (FP) objects. These FP objects are automatically removed based on a binary Boolean decision tree, with the consequence of improving specificity of the algorithm. Further details of this process have been presented elsewhere(11).

This study aimed to examine such a screening-confirmation algorithm using a combination of TBDx and GXP. The primary objective was to evaluate the performance of algorithms using the latest version of the TBDx system as a screening test, confirmed by GXP, in an operational laboratory setting. A secondary objective was to evaluate the

latest version of the TBDx system as a stand-alone tool to replace conventional smear microscopy in resource constrained settings. Some of the results of these studies have been previously reported in the form of an abstract(12).

## METHODS

**Setting:** The study was conducted at the National Tuberculosis Reference Laboratory within the Centre for Tuberculosis at the National Institute for Communicable Diseases, Johannesburg, South Africa. Approval for the study was received from the Faculty of Health Sciences Research Ethics Committee at the University of Pretoria.

**Samples and study procedure:** This study was nested within an on-going TB surveillance program. Consecutive sputum samples from adult patients ( $\geq 18$  years) with presumed TB and not on treatment for TB were received over a two-week period. These sputum samples (1 per patient) were digested and decontaminated as previously described(13) and sediments were then processed for auramine smear microscopy, read both manually and with the TBDx automated system, as well as for liquid-medium culture for mycobacteria (Mycobacteria Growth Indicator Tube, MGIT 960, Becton Dickinson, Sparks, USA). The remnant sediments of processed samples were then stored at 2-8°C. If the TBDx system identified one or more putative AFB in images of 300 fields of view (FoV), the remnant sediment was re-suspended in the GXP sample reagent at a 2:1 ratio and analyzed using the Xpert MTB/RIF assay(14).

**Laboratory testing:** Auramine smear microscopy (reading 100 FoV) and TB culture were performed as previously described(15) and smear grading was done following the

International Union Against Tuberculosis and Lung Disease (IUATLD) guidelines(16). All cultures recorded as positive by the MGIT 960 system were confirmed as *Mycobacterium tuberculosis* by demonstration of AFB on Ziehl-Neelsen staining and a positive MPT-64 antigen result.

To ensure that the quality of microscopy was of a high standard, smears were read by a TB microscopist (GN) with more than 40 years' experience in AFB smear microscopy. All positive smears and at least 10% of negative smears from each batch of tests read by the microscopist, were confirmed by a second highly experienced microscopist with similar experience. Any discordance was reviewed by the two microscopists again (blinded from the TBDx and GXP results) and a final result decided by consensus. Sediments of specimens from which smears were assessed to be negative on review described above were retrieved and tested by the GXP assay as an additional quality assurance measure. These negatives were in addition to the TBDx positive sediments that were tested by GXP which formed part of the algorithm where only TBDx positives would be confirmed by GXP.

Once smears were examined by the microscopists, the slides containing the smears were transferred to a TBDx instrument cartridge by an independent technician for automated microscopy batch processing. Captured images from each of the smears on the loaded slides were processed with the TBDx software, assessing 100 FoV for each of the following areas on the smears: 1) along the perimeter of a rectangle with a length of 40 and width of 10 FoV (100 FoV) in the centre of the slide; 2) four rows of 25 FoV (100 FoV) elsewhere on the slide; and 3) two rows of 50 FoV (100 FoV) elsewhere on the slide. Results were then reported as putative AFB positive or negative for each set of 100 FoV and the aggregated set of 300 FoV. Each digital FoV represents approximately

33% of an optical FoV seen by a microscopist. Acquiring 300 digital FoV in this study ensures an equivalent sampling area by the camera and the microscopist(10). For putatively positive smears, grading was also performed using the –International Union Against Tuberculosis and Lung Disease (IUATLD) criteria (using 300 automatically-read FoV as equivalent to 100 manually-read FoV). The operator of the TBDx was independent and blinded from the routine results, even though the TBDx results were computer-generated.

**Statistical methods:** TBDx positive was defined as at least one acid fast bacilli seen in 300 fields of view. For the analysis we defined, using a-priori criteria, “high positive” by TBDx as those slides with 10 or more putative AFB per 300 FoV, and “low positive” as those with 1-9 putative AFB per 300 FoV.

We conducted three separate analyses, each evaluating a different potential algorithm. The first two analyses addressed the primary objective: In the first analysis, we assumed that all specimens ruled as high positive or low positive by TBDx would be confirmed by GXP. In the second analysis, we assumed GXP confirmation only of low positive specimens. In this analysis, there would be no drug susceptibility results for TBDx high positive specimens.

In the third analysis to address the secondary objective, TBDx was used as a stand-alone test with low positive results treated as negative. Drug susceptibility results, if available, were ignored. We further stratified these analyses with and without the inclusion of specimens having only one putative AFB detected by the TBDx system.

The culture results were used as the reference for all assessments of diagnostic performance. For sensitivity calculations the denominator was the number of specimens for which the culture was positive for *Mycobacterium tuberculosis*, and the denominator for specificity calculations was the number of specimens for which the culture was negative. Likelihood ratios were also calculated. Contaminated cultures or cultures positive for nontuberculous mycobacteria were excluded. The binomial exact method was used for calculation of confidence intervals.

## **RESULTS**

A GXP assay was done for all 277 results that were TBDx positive and on 127 that were negative as part of the quality assessment . GXP was repeated for 27 of the 28 with an error or invalid result (one had insufficient specimen volume for a repeat). The GXP was positive in 5% (6/127) of the TBDx negatives; 2 were low positive and 3 were very low positive. Among the TBDx positives, the GXP was positive in 35% (96/277) of cases.

### **Performance characteristics of routine smear:**

Of 1210 sputum specimens available for the study 1,009 (83.5%) were eligible for analysis (Figure 1). Of these, 109 were culture positive for *M.tuberculosis* (10.8%). The microscopists graded 75 of these 109 as smear positive, resulting in a sensitivity of 68.8% (95% confidence interval [CI]: 59.2%, 77.3%; Table 1). The remaining 900 were culture negative (89.2%), of which the microscopists graded 7 as smear positive, giving a specificity of 99.2% (95% CI: 98.4%, 99.7%).

### **Number of AFB detected on the TBDx system compared with culture**

The proportion of TBDx positive specimens that were also *M tuberculosis* culture positive ranged from 7 of 158 smears (4.4%) where only one AFB was detected by

TBDx, to 68 of 70 smears (97.1%) with  $\geq 10$  AFB detected by TBDx (Table 2). Of the two specimens that had  $\geq 10$  AFB on TBDx that were culture negative, one was positive on GXP.

### **Performance characteristics of TBDx-GXP algorithm**

The performance characteristics of the different combinations and number of GXP tests required for confirmation are shown in Table 3. An algorithm in which all TBDx high positive and low positive results were confirmed with GXP had a sensitivity of 77% (84/109, 95% CI: 68%-85%) and specificity of 98.9% (890/900, 95% CI: 98.0%-99.5%), requiring 277 GXP assays to be run on a total sample size of 1009 people with presumed TB. If TBDx cases with only 1 putative AFB were regarded as negative, the number of GXP assays to be run was lowered to 119, with a slightly increased specificity (99.4%, 895/900, 95% CI: 98.7%-99.8%) but sensitivity dropped by 5% (72%, 78/109, 95%CI: 62-80) .The five culture-negative specimens that would be classified as positive by this algorithm included four that were smear-positive and five that were positive on GXP (three classified as “medium” and two as “high”), suggesting true specificity could be even higher.

If TBDx high positive results were taken as positive (without GXP confirmation), and only low positive results confirmed by GXP, sensitivity was 78% (85/109, 95%CI: 69%-85%) and specificity was 98.8% (889/900, 95%CI: 97.8%-99.4%) with only 207 GXP assays required. Thus, relative to the “confirm all positives” algorithm, confirming only low positive with GXP led to one additional true-positive and one additional false-positive diagnosis, whereas confirming only those with 2-9 putative AFB per 300 FoV led to five fewer true-positive and four fewer false-positive diagnoses.

## **Performance characteristics of TBDx as a stand-alone tool**

The TBDx used as a standalone tool showed an overall sensitivity 80% (87/109; 95% CI: 71%-87%) and a specificity of 78.9% (710/900; 95% CI: 76.1%-81.5%) when including 1 putative AFB as positive. The highest specificity was observed when only TBDx high positive slides were assumed positive (and those with 0-9 AFB assumed negative) at 99.8% (898/900, 95% CI: 99.2%-100.0%) but sensitivity was 62% (68/109; 95% CI: 53%-71%). The positive likelihood ratio for the latter algorithm was 310 and the negative likelihood ratio was 0.38 compared with 3.8 and 0.25 respectively for scenario where any putative AFB on TBDx is regarded as positive. In the algorithm where all cases with two or more AFB detected by TBDx were assumed positive (and those with zero or one AFB assumed negative), the sensitivity was 11% higher at 73% (79/109; 95% CI: 64-81) and specificity at 95.7% (861/900; 95% CI: 94.1-96.9). This is the same sensitivity as the algorithm with GXP confirmation of specimens with 2-9 AFB detected by TBDx, however in this algorithm the specificity is 99.3% (894/900; 95% CI: 98.6-99.8%) but requiring only 49 GXP in 1009 (5%) patients.

## **DISCUSSION**

Automated microscopy has the potential to improve the diagnosis of TB, either by reducing the volume of expensive confirmatory tests required (where such tests are performed), or by reducing the need for trained microscopists (where more expensive tests are unavailable). This evaluation demonstrated that the improved TBDx system achieved both of these outcomes with good overall performance. Specifically, an algorithm in which GXP is used to confirm low positive results by TBDx can achieve a sensitivity for active pulmonary TB of 78%, specificity of 98.8%, and with only one in five specimens requiring GXP testing. Without GXP confirmation, TBDx alone achieved similar sensitivity and specificity as two highly experienced microscopists working in

tandem, while processing a large number of slides and without the need for trained personnel. These results, if confirmed in other epidemiological settings, suggest that TBDx can substantially improve diagnostic algorithms for TB by reducing resource requirements without substantively compromising diagnostic accuracy.

The World Health Organization (WHO) has released target product profiles for new diagnostics for TB, one of which is for a triage tool(17). The minimal specification is a sensitivity of >90% compared with the confirmatory test (with an ideal specification of 95%) and specificity of >70% (ideal >80%). Using GXP as the confirmatory test, TBDx had a sensitivity near 90%, and a specificity of near 80% in this study. We have demonstrated the value of this approach with targeted use of confirmatory test in a higher prevalence sub-group resulting in excellent performance with reduced risk of false positives in with the second diagnostic.

Importantly, the vast majority of culture negative samples (79%) were eliminated as negative by TBDx allowing effective utilization of a more accurate and expensive diagnostic as a reflex on cases with a higher likelihood of TB. A study from Pakistan and the another from South Africa (18, 19), have demonstrated that GXP performs very well relative to microscopy in these programmatic scenarios; however, the overall yields at a population level were low with 13% of GXP tests being positive in Pakistan, 8% in South Africa. The use of a triage test such as TBDx in these algorithms might therefore have dramatically reduced the number of GXP tests performed, without substantial loss of sensitivity. Furthermore, a very high LR+ was observed when TBDx was confirmed by Xpert, with a modest LR-.

The algorithm where only cases with 2-9 putative AFB detected by TBDx were confirmed by GXP had sensitivity of 73% and specificity of 99.3% and required only 49 GXP tests in 1009 patients. This provides a valuable middle ground solution in improving diagnostic yields above conventional microscopy to a level comparable to other molecular tests(20) for resource constrained settings. In this scenario, with a positive likelihood ratio of 104.3, if the pretest probability for TB disease is only 20% for an individual patient, the post-test probability of TB disease of the positive result will be above 95% while a negative result would have a 5% post-test probability of TB disease for a negative result.

A limitation of this algorithm is the lack of rifampicin resistance results in all microscopy positive cases with >10AFB per 300FoV. Conducting a GXP test on these cases for the sake of a rifampicin resistance result would commensurately come at a higher cost, which countries could justify or refuse based on existing budgets. Thus in the African and South East Asian regions where the prevalence of MDR is below 2.5% in new cases of TB this limitation may be less of a concern, and reflex testing in this sub-group could be reserved for cases with risk factors for drug resistant TB or those failing initial first-line therapy.

A recent diagnostics landscape report(21) has shown a large array of new technologies nearing market entry. These technologies face two fundamental challenges, namely affordability and performance variability across tested populations, which may limit the generalisation of results. By contrast, as a triage test, TBDx can lower potential costs and reduce subjectivity compared with human assessments or interpretations. Novel studies applying such algorithms could fast track these new technologies to fill an important global diagnostic gap.

TB disproportionately affects resource limited settings, and the paradigm of new high performance (and high cost) diagnostics may not fit with the existing reality and sustainability concerns in resource poor settings(22), suggesting that the practicalities of the real world present challenges for a wholesale move to molecular diagnostics. This has also been noted in a cost and affordability analysis stating that it would not be financially viable for low income countries to adopt GXP as a primary test for all presumptive TB cases(8).

TBDx does offer a solution for addressing the human resource shortages in high burden settings as a stand-alone system with excellent performance against a high microscopy standard consisting of two microscopists with a combined experience exceeding 80 years. The TBDx as a stand-alone diagnostic tool in the current version has shown improvement from the previous proof of concept version with sensitivity increasing from 49% to 62% and specificity 98.9% to 99.7% but importantly without the need for manual review. Unlike human reading at risk of “change blindness”(23) computer aided systems such as the TBDx do not have these risks and is the likely reason for the high consistency observed in each set of the 100 FoV of the 300 read by TBDx (data not shown). Further evaluation of the TBDx to improve the WHO quality assurance of routine microscopy through its re-checking program (8) should also be considered.

For TBDx with only 1 or 2 putative bacilli observed the likelihood of the culture being positive is less than 5% (Table 2), highlighting the importance of confirming such results. As with manual microscopy, automated microscopy does not distinguish bacilli on the basis of viability. Thus, it is possible that false positives by automated microscopy represent non-viable organisms in cases of patients recently diagnosed and are on treatment and did not report this or these organisms are in a non-replicating persister

state (24). However, only five of these false-positives tested positive on GXP (which likewise detects DNA without assessment of viability), making it more likely that these specimens represent fluorescent artifacts. Importantly, apparent artifacts on automated microscopy can be visually adjudicated by human experts and may have clinical relevance.

The findings of this study, though positive, have several limitations; the study was conducted in a high HIV prevalence setting where microscopy is expected to perform poorly and used at a single reference laboratory where the overall microscopy standard was high. We were also unable to stratify performance by HIV status as this data was not available. The exclusion of pediatric cases limits the generalizability of our data to this population group, however performance of microbiological tests including smear microscopy and GXP are known to be poor relative to a clinical reference standard(25) and it is unlikely that the proposed algorithms would provide any further performance advantage for these cases.

In addition, all tests were performed on digested and decontaminated sputum samples, which is not generally the case in programmatic settings. In this study we excluded cultures that were contaminated ( $n=128/1210$ ) from the analysis; however, the cases identified by TBDx and confirmed by GXP in this subset are likely to be true cases and would potentially increase the true yield from the algorithms using a strict culture only standard (excluding contaminated results). Lastly, the current study used fluorescence microscopy rather than the new light emitting diode (LED) based microscopy systems which may further improve microscopy detection rates(26). All of these limitations can be addressed in further studies.

## **CONCLUSION**

The TBDx system demonstrates potential to be a useful triage tool, with GXP used to confirm intermediate results. This novel system could also potentially serve as a triage test for other emerging technologies, ensuring that these tools are efficiently used while maintaining good performance. In our study, the technology as a stand-alone system proved comparable to a highly experienced microscopist and therefore offers a diagnostic solution that could provide quality-assured microscopy services in settings where trained microscopists are difficult to find. In either algorithm, the use of TBDx automated microscopy has the potential to improve upon the existing diagnostic standard of care.

### **Authors' contributions:**

Conception and design: NAI, SVO, DC, GC; Analysis and interpretation: AWD, JL, DD, HM, GN; Drafting the manuscript for important intellectual content: NAI, SVO, DD, JL, AWD, DC, GC.

### **Sources of support:**

No external sources of funding were used. However the manufacturer of the TBDx system supplied the instrument for the period of the evaluation.

### **At a glance commentary:**

The microbiological diagnosis of tuberculosis (TB) has improved dramatically with the introduction of the Xpert MTB/RIF (GXP) assay, an automated molecular test. However, the use of GXP is limited in high-burden countries because of its cost, meaning that TB diagnosis still relies on conventional smear microscopy, which may miss half of all cases. The TBDx automated microscopy system is a novel TB test that relies on digital

evaluation of high-throughput microscopy images and with no requirement for a skilled microscopist. Using TBDx to screen specimens prior to using GXP could detect 90% of patients with GXP-positive TB while reducing the number of GXP tests required by 73%. Using the TBDx system as a stand-alone tool can deliver performance equivalent to two highly skilled microscopists without the need to hire such personnel.

## REFERENCES

1. Paulson T. Epidemiology: A mortal foe. *Nature* 2013; 502: S2-S3.
2. Steingart KR, Henry M, Ng V, Hopewell PC, Ramsay A, Cunningham J, Urbanczik R, Perkins M, Aziz MA, Pai M. Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. *The Lancet Infectious Diseases* 2006; 6: 570-581.
3. Steingart KR, Ng V, Henry M, Hopewell PC, Ramsay A, Cunningham J, Urbanczik R, Perkins MD, Aziz MA, Pai M. Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. *The Lancet Infectious Diseases* 2006; 6: 664-674.
4. Steingart KR SI, Horne DJ, Pai M, Boehme CC, Dendukuri N. Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database of Systematic Reviews* 2014: Art. No.: CD009593.
5. WHO. Monitoring of Xpert MTB/RIF roll-out. 2014 [4 August 2014]. Available from: <http://who.int/tb/laboratory/mtbrifrollout/en/>.
6. Kik SV, Denkinger CM, Chedore P, Pai M. Replacing smear microscopy for the diagnosis of tuberculosis: what is the market potential? *European Respiratory Journal* 2014; 43: 1793-1796.
7. UNITAID. UNITAID approves US\$ 30 million for innovative project to roll out ground-breaking tuberculosis test at reduced cost. 13 July 2014]. Available from: [www.unitaid.eu/en/resources/press-centre/news/986-public-private-partnership-announces-immediate-40-percent-cost-reduction-for-rapid-tb-test](http://www.unitaid.eu/en/resources/press-centre/news/986-public-private-partnership-announces-immediate-40-percent-cost-reduction-for-rapid-tb-test).
8. Pantoja A, C F, A V, K W, Floyd K. Xpert MTB/RIF for diagnosis of tuberculosis and drug-resistant tuberculosis: a cost and affordability analysis. *Eur Respir J* 2013; 42(3): 708-720.
9. van't Hoog AH, Cobelens F, Vassall A, van Kampen S, Dorman SE, Alland D, Ellner J. Optimal triage test characteristics to improve the cost-effectiveness of the Xpert MTB/RIF assay for TB diagnosis: a decision analysis. *PloS one* 2013; 8: e82786.

10. Lewis JJ, Chihota VN, van der Meulen M, Fourie PB, Fielding KL, Grant AD, Dorman SE, Churchyard GJ. "Proof-Of-Concept" Evaluation of an Automated Sputum Smear Microscopy System for Tuberculosis Diagnosis. *PLoS one* 2012; 7: e50173.
11. Divekar A, Pangilinan C, Coetzee G, Sondh T, Lure FYM, Kennedy S. Automated detection of tuberculosis on sputum smeared slides using stepwise classification. 2012. p. 83153E-83153E-83159.
12. Ismail N OS, Dowdy D, Clark D, v/d Meulen H, Kennedy S, Churchyard G A novel TB diagnostic algorithm using automated microscopy achieves high sensitivity while reducing the volume of Xpert MTB/RIF testing. 44th Union World Conference on Lung Health. Paris; 2013.
13. Kubica GP, Dye WE, Cohn ML, Middlebrook G. Sputum digestion and decontamination with N-acetyl-L-cysteine-sodium hydroxide for culture of mycobacteria. *Am Rev Respir Dis* 1963; 87: 775-779.
14. Cepheid. Xpert MTB/RIF assay. 2014. Available from:  
<http://www.cepheid.com/en/cepheid-solutions-uk/clinical-ivd-tests/critical-infectious-diseases/xpert-mtb-rif>.
15. Salman Siddiqi SR-G, Heather Alexander, Francis Drobniowski, Knut Feldman. MGIT Procedure Manual for Bactec MGIT 960 TB system. 2006 3 January 2015]. Available from:  
[http://www.finddiagnostics.org/export/sites/default/resource-centre/find\\_documentation/pdfs/mgit\\_manual\\_nov\\_2007.pdf](http://www.finddiagnostics.org/export/sites/default/resource-centre/find_documentation/pdfs/mgit_manual_nov_2007.pdf).
16. GLI. Laboratory diagnosis of TB by smear microscopy: The handbook. Global Laboratory Initiative; 2013.
17. WHO. Meeting Report: High-priority target product profiles for new tuberculosis diagnostics. [serial online] 2014 [cited 3 January 2015]. Available from:  
[http://apps.who.int/iris/bitstream/10665/135617/1/WHO\\_HTM\\_TB\\_2014\\_18\\_eng.pdf?ua=1](http://apps.who.int/iris/bitstream/10665/135617/1/WHO_HTM_TB_2014_18_eng.pdf?ua=1)
18. Shah SK, Kumar AMV, Dogar OF, Khan MA, Qadeer E, Tahseen S, Masood F, Chandio AK, Edginton ME. Xpert® MTB/RIF under routine conditions in diagnosing pulmonary tuberculosis: a study in two hospitals in Pakistan [Short communication]. *Public Health Action* 2013; 3: 20-22.
19. Hanrahan CF, Selibas K, Deery CB, Dansey H, Clouse K, Bassett J, Scott L, Stevens W, Sanne I, Van Rie A. Time to treatment and patient outcomes among TB suspects screened by a single point-of-care xpert MTB/RIF at a primary care clinic in Johannesburg, South Africa. *PLoS one* 2013; 8: e65421.
20. Scott LE, McCarthy K, Gous N, Nduna M, Van Rie A, Sanne I, Venter WF, Duse A, Stevens W. Comparison of Xpert MTB/RIF with Other Nucleic Acid Technologies for Diagnosing Pulmonary Tuberculosis in a High HIV Prevalence Setting: A Prospective Study. *PLoS Med* 2011; 8: e1001061.
21. UNITAID. Tuberculosis Diagnostics Technology and Market Landscape. UNITAID; 2014.
22. Lumb R. For better or worse, microscopy is here to stay . . . [Editorial]. *Public Health Action* 2013; 3: 3-3.
23. Fischer J, Whitney D. Serial dependence in visual perception. *Nat Neurosci* 2014; 17: 738-743.

24. Garton NJ, Waddell SJ, Sherratt AL, Lee S-M, Smith RJ, Senner C, Hinds J, Rajakumar K, Adegbola RA, Besra GS, Butcher PD, Barer MR. Cytological and Transcript Analyses Reveal Fat and Lazy Persister-Like Bacilli in Tuberculous Sputum. *PLoS Med* 2008; 5: e75.
25. WHO. Xpert MTB/RIF assay for the diagnosis of pulmonary and extrapulmonary TB in adults and children Policy update. World Health Organization; 2013.
26. WHO. Fluorescent light-emitting diode (LED) microscopy for diagnosis of tuberculosis Worl Health Organization; 2011.

Table 1: Frequency and percentage distributions of microscopist's smear status (based on reading 100 fields of view) by culture results in the 1,009 specimens.

| Microscopist reading* | Culture positive for <i>M. tuberculosis</i><br>(n=109) | Culture negative<br>(n=900) |
|-----------------------|--|-----------------------------|
| Smear negative        | 34 (31.2%)   | 893 (99.2%)                 |
| Scanty positive       | 3 (2.8%)   | 1 (0.1%)                    |
| 1+ positive           | 10 (9.2%)  | 3 (0.3%)                    |
| 2+ positive           | 18 (16.5%)   | 2 (0.2%)                    |
| 3+ positive           | 44 (40.4%)   | 1 (0.1%)                    |

\*Scanty: 1-9 AFB/100 FoV, 1+: 10-99 AFB/ 100 FoV , 2+: 1-9 AFB/FoV, 3+:

>10AFB/FoV

Table 2: The distribution of the number of putative acid fast bacilli (AFB) detected by TBDx on 300 fields of view and relationship to culture positivity for *M. tuberculosis*.

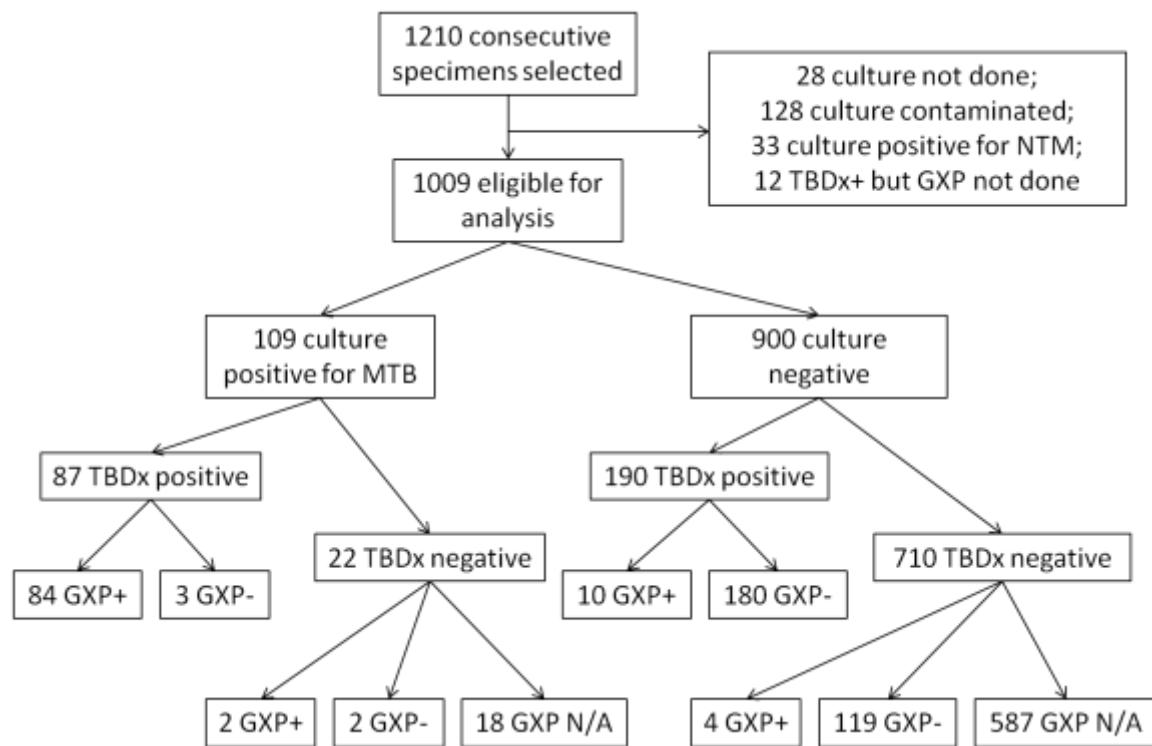
| Number of putative<br>AFB detected by<br>TBDx | Culture positive for |                        |
|---|----------------------|------------------------|
|   | Overall              | <i>M. tuberculosis</i> |
|   | N (column %)         | N (row %)              |
| 0   | 732 (72.5%)          | 22 (3.0%)              |
| 1   | 158 (15.7%)          | 7 (4.4%)               |
| 2   | 26 (2.6%)            | 5 (19.2%)              |
| 3   | 7 (0.7%)             | 2 (28.6%)              |
| 4   | 8 (0.8%)             | 3 (37.5%)              |
| 5-9   | 8 (0.8%)             | 2 (25.0%)              |
| 10-99   | 14 (1.4%)            | 14 (100.0%)            |
| 100-999                                       | 26 (2.6%)            | 25 (96.2%)             |
| 1000+   | 30 (3.0%)            | 29 (96.7%)             |
| Total   | 1,009 (100.0%)       | 109 (10.8%)            |

Table 3: Performance characteristics of TBDx and TBDx/GXP alone algorithms, by number of putative acid fast bacilli detected on 300 fields of view, compared to gold standard of culture for 1,009 smears.

| TBDx without GXP confirmation                                       |                  | Sensitivity<br>(95% CI) | Specificity (95% CI)   | LR +  | LR - | GXP tests required |
|---|------------------|-------------------------|------------------------|-------|------|--------------------|
| Low & high positive   | ≥1 putative AFB  | 80%<br>(71-87%)         | 78.9%<br>(76.1-81.5%)  | 3.8   | 0.25 | N/A                |
|   | >1 putative AFB  | 73%<br>(64-81%)         | 95.7%<br>(94.1-96.9%)  | 17.0  | 0.28 | N/A                |
| High positive only  | ≥10 putative AFB | 62%<br>(53-71%)         | 99.8%<br>(99.2-100.0%) | 310.0 | 0.38 | N/A                |
| TBDx with GXP confirmation of only low positive TBDx results        |                  | Sensitivity<br>(95% CI) | Specificity (95% CI)   | LR +  | LR - | GXP tests required |
| Low & high positive   | ≥1 putative AFB  | 78%<br>(69-85%)         | 98.8%<br>(97.8-99.4%)  | 65.0  | 0.22 | 207                |
|   | >1 putative AFB  | 73%<br>(63-81%)         | 99.3%<br>(98.6-99.8%)  | 104.3 | 0.27 | 49                 |
| TBDx with GXP confirmation of any high or low positive TBDx results |                  | Sensitivity<br>(95% CI) | Specificity (95% CI)   | LR +  | LR - | GXP tests required |
| Low & high positive   | ≥1 putative AFB  | 77%<br>(68-85%)         | 98.9%<br>(98.0-99.5%)  | 70.0  | 0.23 | 277                |
|   | >1 putative AFB  | 72%<br>(62-80%)         | 99.4%<br>(98.7-99.8%)  | 120.0 | 0.28 | 119                |
| High positive only  | ≥10 putative AFB | 62%<br>(52-71%)         | 99.9%<br>(99.4-100.0%) | 620.0 | 0.38 | 71                 |

GXP = Xpert MTB/RIF; AFB = acid fast bacilli. High positive: AFB ≥10 putative AFB. Low positive: 1-9 putative AFB. LR + = Positive likelihood ratio. LR - = Negative likelihood ratio

Figure 1: Flowchart showing eligibility for analysis and results of culture, TBDx and the Xpert MTB/RIF assay.



MTB = Mycobacterium tuberculosis; NTM = non-tuberculous mycobacteria; GXP+ = positive on Xpert MTB/RIF; GXP- = negative on Xpert MTB/RIF; GXP N/A = Xpert MTB/RIF not done.