TITLE PAGE

Comparing accuracy of lipoarabinomannan urine tests for diagnosis of pulmonary tuberculosis in children from four African countries: a cross-sectional study

Esin Nkereuwem[1] MBBCh, Toyin Togun[1,2] PhD, Marie P. Gomez[1] BSc, Rita Székely[3] PhD, Aurélien Macé[3] PhD, Dawda Jobe[1] BSc, Samuel G. Schumacher[3] PhD, Beate Kampmann[1,4] PhD*, Claudia M. Denkinger[3,5] MD* and Reach4KidsAfrica (R4KA) Consortium

[1] Vaccines and Immunity Theme, Medical Research Council Unit The Gambia at London School of Hygiene and Tropical Medicine.

[2] Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

[3] FIND (Foundation for Innovative New Diagnostics), Geneva, Switzerland

[4] The Vaccine Centre, London School of Hygiene and Tropical Medicine, London, United Kingdom

[5] Division of Tropical Medicine, Centre of Infectious Diseases, University Hospital Heidelberg, Heidelberg, Germany

* These authors contributed equally

Corresponding author:

Prof Beate Kampmann, The Vaccine Centre, London School of Hygiene and Tropical Medicine, London, United Kingdom; Theme Leader Vaccines and Immunity, MRC Unit The Gambia @ LSHTM
Tel: +44 20 7 612 7925, Email: Beate.Kampmann@lshtm.ac.uk

Keywords:
Urine lipoarabinomannan (LAM); childhood tuberculosis; diagnostics; sensitivity; specificity.

Summary

Background

A sensitive and specific non-sputum-based test would represent a game-changer for the diagnosis of childhood tuberculosis (TB). We assessed the diagnostic accuracy of the Fujifilm SILVAMP TB LAM (FujiLAM) and Alere Determine TB LAM Ag (AlereLAM) assays side by side for the detection of childhood TB.

Methods

In this cross-sectional study, we tested urine samples from children aged <15 years with presumed pulmonary TB. Children were consecutively recruited at the study sites in The Gambia, Mali, Nigeria and Tanzania from July 2017 to December 2018. Bio-banked urine samples were thawed and tested using FujiLAM and AlereLAM assays. We measured diagnostic performance against a microbiological reference standard (confirmed TB) and a composite reference standard (confirmed and unconfirmed TB). Sensitivity and specificity were estimated by performing bivariate random effects meta-analyses.

Findings

We included urine samples from 415 children; 63 (15%) had confirmed tuberculosis, 113 (27%) had unconfirmed tuberculosis, and 239 (58%) were unlikely tuberculosis. The HIV prevalence was 14·7% (61 of 415). Using the microbiological reference standard, the sensitivity of FujiLAM was 64·9% (95% CI 43·7-85·2; 40 of 63) compared with 30·7% (8·6-61·6; 19 of 63) for AlereLAM. The specificity of FujiLAM was 83·8% (76·5-89·4; 297 of 352) and 87·8% (79·0-93·7; 312 of 352) for AlereLAM. Against the composite reference standard, both assays had lower sensitivity, 32·9% (24·6-41·9; 58 of 176) for FujiLAM *vs* 20·2% (12·3-29·4; 36 of 176) for AlereLAM. Specificity of FujiLAM (83·3% [71·8-91·7]; 202 of 239) was comparable to AlereLAM (90·0% [81·6-95·6]; 216 of 239).

Interpretation

In comparison to AlereLAM, FujiLAM demonstrated a higher sensitivity with comparable specificity, and could potentially add value to the rapid diagnosis of TB in children.

Funding

German Federal Ministry of Education and Research through KfW, Global Health Innovative Technology Fund and UKRI-GCRF and MRC program grant award.

Research in context

Evidence before this study

We searched PubMed Central for all studies or reports of lipoarabinomannan for the diagnosis of pulmonary tuberculosis in children. We used search terms “pulmonary tuberculosis” OR “ptb” AND “child” OR “infant” OR “adolescent” AND “lipoarabinomannan” OR “lam” AND “urine”, up to Mar 9, 2020. No language restrictions were applied. Our search returned five relevant publications that included children below 15 years undergoing LAM testing of urine for diagnosis of pulmonary tuberculous using AlereLAM. None of the studies used FujiLAM. The sensitivity of LAM, compared with confirmed TB, ranged from 48·3% to 73·2% for all children. Even though the WHO has recommended the use of AlereLAM to aid in the diagnosis of TB in children and adolescents with HIV, the test still shows sub-optimal sensitivity in most studies.
When we searched using the search terms “fuji silvamp tb” OR “fujilam” OR “fuji silvamp tb lam”, we found two relevant studies, both of which were performed in adults with HIV and found that FujiLAM had a notably higher sensitivity than AlereLAM. No comparison of the two tests in children has been done to date, and the role of FujiLAM in childhood PTB diagnosis remains unknown.

Added value of this study

To our knowledge, this study is the first side-by-side comparison of FujiLAM and AlereLAM on samples of children from four countries. In our study, we compared the diagnostic performance of both tests. We showed the sensitivity of FujiLAM to be substantially higher than that of AlereLAM in all children, and in all sub-groups, relative to both microbiological and composite reference standards for tuberculosis. The high sensitivity of FujiLAM, however, was mainly observed in children with confirmed TB, and remains suboptimal. FujiLAM and AlereLAM had similar specificity.

Implications of the available evidence

FujiLAM showed substantially higher sensitivity compared with AlereLAM for the diagnosis of pulmonary tuberculosis in children, while maintaining comparable specificity. FujiLAM's sensitivity (64·9%) remains too low to be used as a rule-out test for childhood pulmonary tuberculosis but could add value to diagnostic algorithms, given that urine is an easily obtainable sample.

Introduction

Childhood tuberculosis (TB) is estimated to account for about 11% of the 10 million new cases of TB worldwide, and 15% of associated total mortality.[1,2] However, these figures are likely to underestimate the true burden of the disease due to difficulty in obtaining good quality sputum samples from young children and the sub-optimal sensitivity of currently available routine diagnostic tools in children.[3–5] Even when the sputum is successfully collected, traditional diagnostic methods such as culture or smear microscopy have low sensitivity in children due to the paucibacillary nature of the disease.[6] Therefore, confirming the diagnosis of tuberculosis in children remains a challenge, with a majority of the diagnoses being made based on clinical criteria.[4]

In recognition of these challenges, the World Health Organisation (WHO) has prioritised the need for rapid, point-of-care diagnostic solutions for detection of childhood TB.[7] This new tool should ideally be non-sputum based, and be able to identify *Mycobacterium tuberculosis* (*Mtb*) with high sensitivity and specificity regardless of age, nutritional status or HIV status.[8,9] The WHO has recommended the use of the urine-based point-of-care test Alere Determine LAM ('AlereLAM'; Abbott, Palatine, IL, USA), to assist in the diagnosis of TB in children and adolescents with HIV.[10] However, AlereLAM has demonstrated sub-optimal sensitivity in studies in children and therefore warrants the need for a more sensitive point-of-care test.[11,12]

A novel test, Fujifilm SILVAMP TB LAM assay ('FujiLAM'; Fujifilm, Tokyo, Japan), which has emerged is, similar to AlereLAM, based on the detection of mycobacterial lipoarabinomannan (LAM) antigen in urine.[10] The FujiLAM is a new urine-based test for TB, that is based on the use of Silver Amplification Immunochromatography on a lateral flow strip which enables a 30-fold lower analytical sensitivity compared to the only commercially available LAM test, the AlereLAM. Broger and colleagues reported a significantly higher diagnostic sensitivity when FujiLAM was compared with AlereLAM among in-patient adults with HIV.[13] Given the significantly higher clinical sensitivity in adults, FujiLAM might also perform well in children, but to date, there are no published results.

In this study, we set out to assess the diagnostic accuracy of FujiLAM for the detection of active pulmonary tuberculosis (PTB) compared side-by-side with AlereLAM in a centralised laboratory environment using previously collected, frozen urine samples from a cohort of children under the age of 15 years.

Methods

Study participants, setting, study design

Children aged below 15 years with presumed PTB were consecutively enrolled at dedicated out-patient childhood TB clinics at four sites in The Gambia, Mali, Nigeria, and, Tanzania from July 2017 to December 2018 (detail on cohorts available in the Appendix, page 2). Children were eligible for enrolment if they had symptoms suggestive of PTB characterised by persistent or unremitting cough for more than two weeks with any of weight loss/failure to thrive, and persistent unexplained fever.[14]

No child was included in the study based on the presence of mediastinal lymphadenopathy alone. At enrolment, medical history and demographic information were obtained for each child. The anthropometric measurements, which included weight-for-age, length/height-for-age, and body mass index-for-age *Z*-scores were calculated using WHO 2007 reference standards.[15] Clinical investigation included chest radiography and HIV testing in children whose HIV status was not known (HIV rapid test followed by a confirmatory PCR for children below 18 months, or HIV ELISA for children aged 18 months or older). Each child then provided a urine specimen for LAM testing on a future date.

This study is reported in accordance with the Standards for Reporting of Diagnostic Accuracy Studies guidelines.[16] All samples were obtained using a well-defined protocol that received approval from the Gambia Government/MRC Joint Ethics Committee and from the IRB/IEC at the respective institutions where the samples were obtained. Written informed consent was obtained from the parent or legal guardian of each patient, and assent was obtained from the older children. Legal representatives of the participants were informed that all specimens collected would be used for TB diagnostic studies, and those who were willing to have specimens stored were asked to provide written informed consent specifically for these activities.

Procedures

Sputum samples were obtained from participants spontaneously or by sputum induction using nebulised hypertonic saline. For reference standard testing, the specimens were processed using standardised protocols at each study site. Reference standard testing was done on all available sputum specimens and included Xpert MTB/RIF Ultra assay ('Xpert Ultra'; Cepheid, Sunnyvale, USA), mycobacteria growth indicator tube liquid culture (Becton Dickinson, Franklin Lakes, NJ, USA) and solid culture on Löwenstein-Jensen medium. The presence of *Mtb* in positive cultures was confirmed with acid-fast staining and MPT64 antigen detection (Abbott, Palatine, IL, USA) or MTBDRplus line probe assays (Hain Lifesciences, Nehren, Germany). For all sites except Nigeria, the microbiological reference standard used for the diagnosis of tuberculosis was a mycobacterial confirmation by Xpert Ultra or culture from at least one respiratory specimen. For Nigeria, Xpert Ultra alone was used as culture facilities were not available. The patients were categorised as having Confirmed TB, Unconfirmed TB or Unlikely TB by experienced clinicians at each study site using a combination of clinical and laboratory findings based on the revised classification (Table 1).[14] Details of the reference standard outcomes among the diagnostic categories is available in the Appendix, page 3.

At enrolment, spontaneously voided urine specimens were collected into a sterile container for the older children, and into a urine bag for the younger children. All urine specimens were stored immediately after collection until used for analysis. The duration of biobanking of the urine samples was between two and 19 months at minus 80^{o}C in temperature-monitored freezers before the urine LAM assays were performed. For both urine LAM assays, frozen aliquots of unprocessed urine were thawed to ambient temperature and mixed manually. Samples that were not immediately used for testing were stored at 4°C for a maximum of four hours. FujiLAM tests were performed and graded

by trained study staff according to manufacturer recommendations. The detailed protocol and procedure for the FujiLAM test have been previously described.[13] In brief, around 200 µl of urine was added to a reagent tube, mixed, and incubated for 40 minutes at room temperature. Following this, two drops of the urine were added to the sample port of the cartridge labelled one, and button two was pressed immediately. After observing a colour change to orange in the *Go Next* port, button three was pressed, and the result was read after the black control band appeared. The cartridge was visually inspected and interpreted by two independent staff who were blinded to other results from reference standard tests, comparator tests and to the index test results from each other. In case of discordant results, the two readers agreed upon a consensus result, and the final call was used for the analysis of primary endpoints. In the case of assay failure, the test was repeated once, and the repeat result used in the final analysis if successful.

AlereLAM tests were performed according to manufacturer recommendations.[17] The steps followed involved adding 60 µl of urine to the sample pad and reading off after 25 minutes using the manufacturer's reference scale card.[13] Identical to the FujiLAM reading, the AlereLAM was also interpreted by two independent and blinded staff. In the case of discordant AlereLAM results, the two readers agreed upon a consensus result, and the final call was used for the analysis of primary endpoints.

The FujiLAM and AlereLAM testing were performed in parallel in batches of ten urine samples. After testing, laboratory staff took turns to record their independent interpretation of the AlereLAM for the entire batch followed by FujiLAM. Both LAM assays were interpreted by the same laboratory staff who recorded their interpretation of each test on a separate report form blinded to their own comparator test results for the same sample. We excluded urine samples from further analysis where at least one valid microbiological test for *Mtb* (either Xpert or culture) from the subject was not available. All available urine samples from the study cohort were tested with both tests.

Statistical analysis

Due to the secondary nature of the study, there was no formal sample size calculation. A convenience sample of all consecutively enrolled children was included in the primary and secondary analyses. In the primary analysis, using a microbiological reference standard, participants categorised as Confirmed TB were considered reference standard positive and participants with Unconfirmed TB and Unlikely TB as negative. In a composite reference standard, participants with Unconfirmed TB were reclassified as positive. We determined point estimates for sensitivity and specificity with 95% confidence interval (CI) of AlereLAM and FujiLAM against both a microbiological and composite reference standard for the entire cohort. In the secondary analysis, sensitivity and specificity of AlereLAM and FujiLAM were determined for predefined subgroups as follows: age below five years old *vs* five years and above; children living with HIV *vs* children that are HIV-negative; study country (Gambia, Mali, Nigeria or Tanzania); normal height *vs* stunted; and normal weight *vs* underweight.. Samples from all enrolled subjects were included in the secondary analysis. We estimated the pooled

sensitivity, specificity, and 95% CI using the Bayesian bivariate random-effects meta-analysis to account for the possible effects of heterogeneity across the subgroups. For the per country analysis, we estimated the 95% CI using the exact method. The results of the urine LAM assays were not included in the composite reference standard to avoid incorporation bias. All data were entered into Microsoft Excel and analysed using R (version 3.5.1).

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The sponsor of the study, FIND, is taking legal responsibility for the overall conduct of the study. This is a sponsor-investigator study, and as such FIND was involved in many aspects of the study including the protocol, controlling and training on the products under investigation, analysis plans and analysis. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Of the 464 eligible children who were enrolled, 49 children were excluded from the main analysis because of the absence of a complete microbiological reference standard (n=12), urine samples not being available (n=36) and an invalid FujiLAM test (n=1) after repeat (Figure 1). Overall, urine samples from 415 children were included in the main analysis (167 from The Gambia, 67 from Mali, 81 from Nigeria and 100 from Tanzania). A total of 63 (15%) were classified as Confirmed TB, 113 (27%) as Unconfirmed TB, and 239 (58%) as Unlikely TB (Table 2).

Using the microbiological reference standard, the pooled sensitivity of FujiLAM was 64·9% (95% CI 43·7-85·2; 40 of 63) and 30.7% (8·6-61·6; 19 of 63) for AlereLAM (difference 34·2%; Figure 2). FujiLAM sensitivity was higher in the HIV-negative group (67·5% [41·8-88·0]) than children with HIV (54·8% [28·7-81·5]). Conversely, the sensitivity of AlereLAM was higher in children with HIV (36·6%, [13·8-70·4]) compared to the HIV negative group (26·6%, [1·2-66·4]), although, with wide and overlapping confidence intervals. In the analysis by age-group, FujiLAM showed higher sensitivity (61·8% [36·6-85·5] than AlereLAM (38·8% [0·4-98·9]; difference 23·0%) in children below five years of age, even though the confidence intervals were wide and overlapping. The difference in sensitivity between the assays was higher in patients aged five years and older. FujiLAM had a five-fold higher sensitivity than AlereLAM in the Gambia subgroup (50·0% [18·7-81·3] for FujiLAM *vs* 10·0% [0·3-44·5] for AlereLAM; difference 40·0%), a three-fold higher sensitivity in the Tanzania subgroup (70·6% [44·0-89·7] for FujiLAM *vs* 23·5% [6·8-49·9] for AlereLAM; difference 47·1%), and a two-fold higher sensitivity in the Mali subgroup (56·5% [34·5-76·8] for FujiLAM *vs* 26·1% [10·2-48·4] for AlereLAM), albeit with wide and overlapping confidence intervals. Both assays had higher point estimates for sensitivity in the Nigeria subgroup. The sensitivity for FujiLAM also remained higher than for AlereLAM in the height and weight subgroup analyses (Appendix, page 4).

There was an overall lower sensitivity of both assays with the composite reference standard compared to the microbiological reference standard (32·9% [24·6-41·9] for FujiLAM *vs* 20·2% [12·3-29·4] for AlereLAM; difference 12·7%), although with similarly overlapping confidence intervals. However, the sensitivity for FujiLAM remained higher than for AlereLAM in all subgroup analyses (Figure 2, Appendix, page 4). Of note, the confidence intervals of the comparisons between FujiLAM and AlereLAM were wide and overlapping in all subgroup analyses except in children with normal height.

Compared to the microbiological reference standard, there was a pooled specificity of 83·8% (76·5-89·4) for FujiLAM and 87·8% (79·0-93·7) for AlereLAM, with a difference of -4·0% and overlapping confidence intervals. Overall specificity dropped with the composite reference standard to 83·3% (70·3-93·4) for FujiLAM and increased to 90·0% (81·6-95·6) for AlereLAM, with a larger difference (-6·7%) but again overlapping and wide confidence intervals. The specificity for AlereLAM remained higher than for FujiLAM in all subgroup analyses except the Mali and Tanzania cohorts (Figure 2 and Appendix, page 4).

Discussion

In this report, we compared the performance of two urine LAM point-of-care tests in over 400 urine samples from children with presumed PTB, including children with HIV, originating from four different African countries. The sensitivity of FujiLAM was more than double that of the AlereLAM test when evaluated against a microbiological reference. FujiLAM also had higher point estimates of sensitivity than AlereLAM in all sub-group analyses, while maintaining comparable specificity. Both assays performed better when the microbiological reference standard was used as opposed to the composite reference standard. However, both FujiLAM and AlereLAM assays had an overall suboptimal sensitivity when compared with the WHO target product profile's minimum recommended sensitivity of ≥90% for a triage test or ≥66% for a diagnostic test.[8]

Although the point estimates of the sensitivity for FujiLAM were higher than for AlereLAM in the entire cohort and all sub-group analyses, the confidence intervals were wide and overlapping in most of our comparisons. The sample size in our study was small for sensitivity calculation resulting in the wide confidence intervals. This explains the overlapping confidence intervals despite the observed difference in point estimates.

There are few reports on the use of point-of-care LAM tests for TB diagnosis in children, none of which used FujiLAM. A previous study[11] reported a pooled sensitivity of 48·3% (33·7–59·2) for AlereLAM, which is not very different from that observed in our study. Like our study, their tests were conducted on frozen, bio-banked urine samples collected from children below 15 years with symptoms suggestive of PTB. Conversely, Gautam and colleagues reported a higher sensitivity of 73·2% for AlereLAM among children, of which 25·3% had microbiological confirmation for *M. tuberculosis*.[18] The sensitivity of a test may often vary with factors that influence the pre-test probability of disease.[19] These differences, such as the greater proportion of children above five years

old in their study (72% *vs* 53% in our study), make the results less comparable, and may explain the differences observed.

In our study, the point estimates of FujiLAM sensitivity ranged from as low as 25·6% in the Gambia cohort (with the composite reference standard) to as high as 90·9% (with the X-pert-only microbiological reference standard) in the Nigeria cohort. While the results vary widely, the consistently better performance for FujiLAM compared to AlereLAM suggests that the design of FujiLAM enables improved detection of urinary LAM. This is consistent with the findings reported in adults with HIV.[13,20]

It is interesting to note that the best performance of both assays was seen with the Nigerian samples, where the sensitivity was relatively higher compared to the other three countries. However, as Xpert Ultra alone was used as the reference standard in Nigeria, the patients identified as Confirmed TB in the Nigerian cohort were likely to be less paucibacillary than the patients in the other three cohorts where culture results had been available. As it is well established that the sensitivity of culture is higher than that of Xpert Ultra in children,[21,22] the use of Xpert Ultra alone as a reference standard would, therefore, have been biased towards children with higher bacillary loads in order to be detectable by Xpert Ultra. As a consequence, the urine samples from Xpert-positive children might contain a higher amount of LAM, and therefore FujiLAM (and AlereLAM) would have potentially diagnosed more cases amongst the Nigerian cohort. This could explain the inflated sensitivity estimates seen with the Nigerian samples. The Nigerian children also had a higher rate of stunting, possibly an indication that they were overall sicker patients. However, we cannot currently verify these hypotheses, and this study was designed to conduct side-by-side comparisons of two available LAM tests rather than evaluate the performance of these tests in clinical subgroups.

We observed a wide variation in the sensitivity of FujiLAM between the other three countries' cohorts. While we used the same inclusion criteria and clinical recruitment procedures at all the study sites, the variation in sensitivity could reflect the fact that our study sites in the four countries represent different health care levels including community-based recruitment setting in The Gambia, urban comprehensive health care facility in Tanzania, to tertiary hospital-based recruitment setting in Mali and Nigeria. This is supported by the differences in the proportions of TB disease in each cohort.[19]

We found that FujiLAM performed worse among the children with HIV than in the HIV negative group in our study, a finding that differs from results in adults.[13,20] Given the relatively small numbers of children with HIV in the entire cohort, and even smaller numbers in the subgroup analyses, we are reluctant to draw any firm conclusions on the performance in the context of HIV. The observed performance in children living with HIV could have been due to chance.

Even though the point estimates of FujiLAM specificity were lower than those for AlereLAM in most of the subgroup analyses, the differences in specificity were not significant. Also, FujiLAM had a

higher specificity than AlereLAM in the Mali and Tanzania cohorts. Based on analytical sensitivity studies[23], particularly in respect to rapid growing non-tuberculous mycobacteria, one would expect a higher specificity clinically as well for FujiLAM. This is not in line with our observations. Overall, specificity was lower than observed in adult patients,[13,20] which could have resulted from the higher amount of expected contamination from perineal flora in the process of urine collection in children versus adults.

The tests were performed on bio-banked urine specimens in a research laboratory setting. Connelly and colleagues[24] have suggested that fresh urine samples contain more detectable LAM than frozen samples, thus implying that the sensitivity of LAM-based tests would be improved with fresh samples. However, this finding has not been confirmed in studies using the FujiLAM test[25] and accordingly, we would not expect any significant change in the sensitivity between fresh and frozen samples.

The inclusion of urine samples of children from four different countries who were recruited using the same criteria is a strength of this study, as it allowed us to assess the performance of FujiLAM across different geographical settings. Additionally, the use of two reference standards further strengthens the study. Nevertheless, the heterogeneity of these cohorts, together with the difference in reference standard used in Nigeria, also represents one of its limitations.

Despite the sizeable number and the geographical spread of our cohort, we believe that it is premature to conclude on the ultimate utility of FujiLAM in clinical practice based on these results alone. The potential clinical value of this test could be seen in children from whom adequate sputum cannot be obtained, and who would not otherwise be diagnosed. As it uses a very small quantity of easily obtainable urine, it avoids invasive procedures at an acceptable opportunity cost. Its high specificity qualifies it as a rule-in test that could guide further diagnostic evaluation. We, therefore, recommend larger and appropriately designed studies to evaluate the potential utility of FujiLAM and to demonstrate its added value in terms of showing how many additional TB cases, which would otherwise have been missed, will be detected if LAM is combined with Xpert and culture.

In conclusion, FujiLAM showed substantially higher sensitivity than AlereLAM while maintaining comparable specificity. In the absence of the “dream test” advocated for by the WHO, FujiLAM has the potential to add value to the diagnosis of childhood TB.

Acknowledgement

This work was supported by a UKRI-GCRF and MRC program grant to BK (MR/P024270/1; MR/K011944/1) and additionally by the German Federal Ministry of Education and Research through KfW (grant number 2020 60 457), Global Health Innovative Technology Fund (GHIT grant number G2017-207). We thank the children and their parents for participating in the study. Additionally, we would like to acknowledge the invaluable contributions of the clinical, laboratory, data, and research support staff at the Medical Research Council Unit The Gambia at London School of Hygiene and

Tropical Medicine; University Clinical Research Centre (UCRC)-SEREFO Laboratory, University of Sciences, Techniques and Technologies of Bamako (USTTB), Bamako, Mali; Jos University Teaching Hospital, Jos, Nigeria; and NIMR-Mbeya Medical Research Centre, Mbeya, Tanzania.

Contributors

EN, TT, RS, BK, and CMD conceptualised and designed the study. EN, MPG and DJ analysed the samples and collected the data. EN and AM did the analysis and generated the estimates. EN drafted the initial version of the manuscript with input from TT, MPG, RS, AM, SGS, BK, and CMD. All authors contributed to the revision and correction on multiple iterations of the manuscript.

Conflicts of interests

This study was supported by FIND and conducted at the MRC Unit in the Gambia. All data were initially analysed by the MRC Unit staff and raw data and analysis results shared with FIND.

References

1. WHO. Global tuberculosis report 2019. World Health Organisation, Geneva 2019. Available from: https://www.who.int/tb/publications/global_report/en/. Date accessed Apr 3, 2020
2. Basu Roy R, Whittaker E, Seddon JA, Kampmann B. Tuberculosis susceptibility and protection in children. *Lancet Infect Dis*. 2019; **19**(3): e96–e108. doi:10.1016/S1473-3099(18)30157-9
3. WHO Stop TB partnership. The missing 3 million: reach, treat, cure everyone. World Health Organization 2016. Available from: http://www.stoptb.org/assets/documents/resources/factsheets/Stop%20TB%20infographic%20Missing%203%20Million.pdf. Date accessed Apr 3, 2020
4. Dodd PJ, Gardiner E, Coghlan R, Seddon JA. Burden of childhood tuberculosis in 22 high-burden countries: a mathematical modelling study. *Lancet Glob Health*. 2014; **2**(8): e453–e459. doi:10.1016/S2214-109X(14)70245-1
5. Zar HJ, Hanslo D, Apolles P, Swingler G, Hussey G. Induced sputum versus gastric lavage for microbiological confirmation of pulmonary tuberculosis in infants and young children: a prospective study [published correction appears in Lancet. 2005 Jun 4-10;365(9475):1926]. *Lancet*. 2005; **365** (9454):130–134. doi:10.1016/S0140-6736(05)17702-2
6. Zar HJ, Connell TG, Nicol M. Diagnosis of pulmonary tuberculosis in children: new advances. *Expert Rev Anti Infect Ther*. 2010; **8**(3):277–288. doi:10.1586/eri.10.9
7. WHO. High-priority target product profiles for new tuberculosis diagnostics. World Health Organization, Geneva 2014. Available from: https://www.who.int/tb/publications/tpp_report/en/. Date accessed Jul 25, 2019
8. Denkinger CM, Kik SV, Cirillo DM, et al. Defining the needs for next generation assays for tuberculosis. *J Infect Dis*. 2015; **211** Suppl 2(Suppl 2): S29–S38. doi:10.1093/infdis/jiu821
9. Nicol MP, Gnanashanmugam D, Browning R, et al. A Blueprint to Address Research Gaps in the Development of Biomarkers for Pediatric Tuberculosis. *Clin Infect Dis*. 2015; **61**Suppl 3(Suppl 3): S164–S172. doi:10.1093/cid/civ613
10. WHO. Lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis of active tuberculosis in people living with HIV. Policy update 2019. World Health Organization, Geneva, 2019. Available from: https://apps.who.int/iris/bitstream/handle/10665/329479/9789241550604-eng.pdf?sequence=1&isAllowed=y&ua=1. Date accessed Nov 29, 2019
11. Nicol MP, Allen V, Workman L, et al. Urine lipoarabinomannan testing for diagnosis of pulmonary tuberculosis in children: a prospective study. *Lancet Glob Health*. 2014; **2**(5): e278–e284. doi:10.1016/S2214-109X(14)70195-0
12. Marais BJ. Urine lipoarabinomannan testing in children with tuberculosis. *Lancet Glob Health*. 2014; **2**(5): e245–e246. doi:10.1016/S2214-109X(14)70207-4
13. Broger T, Sossen B, du Toit E, et al. Novel lipoarabinomannan point-of-care tuberculosis test for people with HIV: a diagnostic accuracy study. *Lancet Infect Dis*. 2019; **19**(8): 852–861.

doi:10.1016/S1473-3099(19)30001-5

14. Graham SM, Cuevas LE, Jean-Philippe P, et al. Clinical case definitions for classification of intrathoracic tuberculosis in children: an update. *Clin Infect Dis*. 2015; **61** Suppl 3(Suppl 3): S179–S187. doi:10.1093/cid/civ581
15. WHO. Child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development. World Health Organization, Geneva, Switzerland 2006. Available from: https://www.who.int/childgrowth/standards/Technical_report.pdf. Date accessed Jul 10, 2019
16. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015; **351**: h5527. Published 2015 Oct 28. doi:10.1136/bmj.h5527
17. Abott Laboratories. ALERE Determine TB LAM AG. Available from: https://www.alere.com/en/home/product-details/determine-tb-lam.html. Date accessed: Nov 14, 2019
18. Gautam H, Singla M, Jain R, Lodha R, Kabra SK, Singh UB. Point-of-care urine lipoarabinomannan antigen detection for diagnosis of tuberculosis in children. *Int J Tuberc Lung Dis*. 2019; **23**(6):714-719. doi:10.5588/ijtld.18.0364
19. Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ*. 2013; **185**(11):E537-E544. doi:10.1503/cmaj.121286
20. Bjerrum S, Broger T, Székely R, et al. Diagnostic accuracy of a novel and rapid lipoarabinomannan test for diagnosing tuberculosis among people with Human Immunodeficiency Virus. *Open Forum Infect Dis*. 2019; **7**(1): ofz530. doi:10.1093/ofid/ofz530
21. Nicol MP, Workman L, Prins M, et al. Accuracy of Xpert Mtb/Rif Ultra for the Diagnosis of Pulmonary Tuberculosis in Children. *Pediatr Infect Dis J*. 2018; **37**(10):e261-e263. doi:10.1097/INF.0000000000001960
22. Sabi I, Rachow A, Mapamba D, et al. Xpert MTB/RIF Ultra assay for the diagnosis of pulmonary tuberculosis in children: a multicentre comparative accuracy study. *J Infect*. 2018; **77**(4):321-327. doi:10.1016/j.jinf.2018.07.002
23. Sigal GB, Pinter A, Lowary TL, et al. A novel sensitive immunoassay targeting the 5-methylthio-d-xylofuranose-lipoarabinomannan epitope meets the WHO's performance target for tuberculosis diagnosis. *J Clin Microbiol*. 2018; **56**(12): e01338-18. doi:10.1128/JCM.01338-18
24. Connelly JT, Grant B, Munsamy V, Pym A, Somoskovi A. Lipoarabinomannan point-of-care tests: evaluation with fresh samples needed. *Lancet Infect Dis*. 2019; **19**(10): 1053. doi:10.1016/S1473-3099(19)30475-X
25. Broger T, Muyoyeta M, Kerkhoff AD, Denkinger CM, Moreau E. Tuberculosis test results using fresh versus biobanked urine samples with FujiLAM. *Lancet Infect Dis*. 2020; **20(1)**: 22–

23. doi:10.1016/S1473-3099(19)30684-X

Table and Figure titles and legend

Figure 1: Study flow diagram

***Note**: LAM = lipoarabinomannan; Ultra = Xpert MTB/RIF Ultra assay*; *FujiLAM = Fujifilm SILVAMP TB LAM assay; MRS = microbiological reference standard; CRS = composite reference standard*

Table 1: Revised classification of intrathoracic TB case definitions for diagnostic evaluation studies in children*

**All children were followed up for a period of at least six months*

***Note**: TB = tuberculosis; Mtb = Mycobacterium tuberculosis*

Table 2: Demographic and clinical characteristics of study participants

***Note**: IQR = inter-quartile range; TB = tuberculosis; HIV = human immunodeficiency virus*

Figure 2: Sensitivity, specificity and differences in sensitivity and specificity between FujiLAM and AlereLAM against the microbiological reference standard (MRS) and composite reference standard (CRS)

***Note**: TP = true positive; FP = false negative; FN = false positive; TN = true negative; ΔSn = difference in sensitivity; ΔSp = difference in specificity*

Appendix Table and Figure titles and legend

Supplementary Table 1: Supplementary Table 1: Study population, setting and location

Supplementary Table 2: Reference standard outcomes among diagnostic categories

** **Note**: NA = not applicable; FujiLAM = Fujifilm SILVAMP TB LAM assay; AlereLAM = Alere Determine LAM; Xpert Ultra = Xpert MTB/RIF Ultra assay*

Supplementary Figure 1: Sensitivity, specificity and differences in sensitivity and specificity between FujiLAM and AlereLAM against the microbiological reference standard (MRS) and composite reference standard (CRS) in different anthropometric groups

Note: TP = true positive; FP = false negative; FN = false positive; TN = true negative; ΔSn = difference in sensitivity; ΔSp = difference in specificity

Figure 1

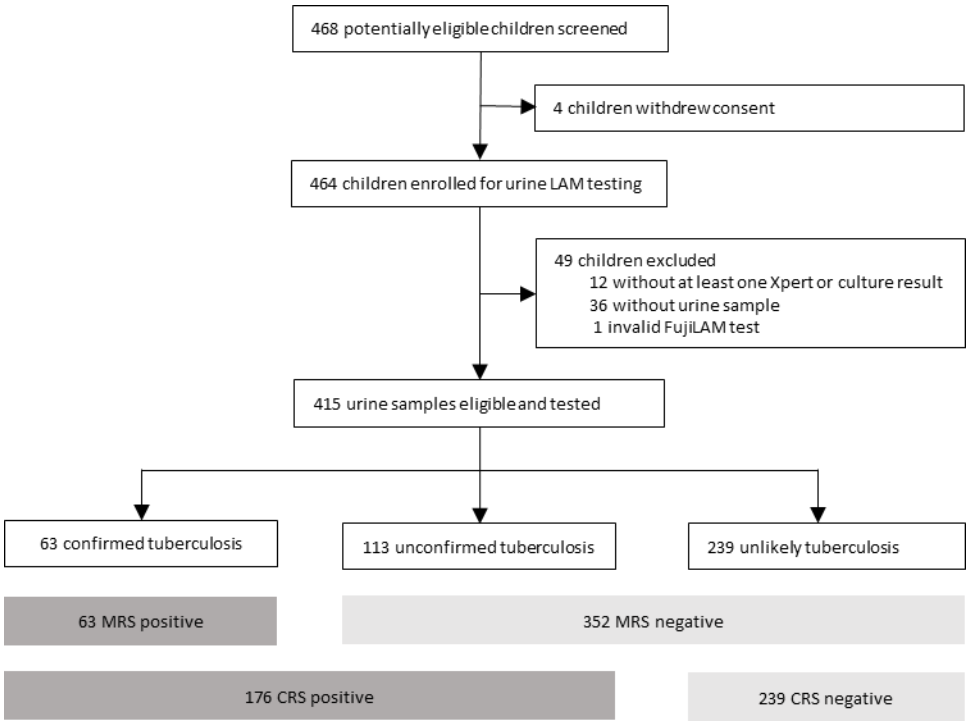
468 potentially eligible children screened
4 children withdrew consent
464 children enrolled for urine LAM testing
49 children excluded
12 without at least one Xpert or culture result
36 without urine sample
1 invalid FujiLAM test
415 urine samples eligible and tested
63 confirmed tuberculosis
113 unconfirmed tuberculosis
239 unlikely tuberculosis
63 MRS positive
352 MRS negative
176 CRS positive
239 CRS negative

Figure 2

MRS	Test	N	TP	FP	FN	TN	Sensitivity [95% CI]		Specificity [95% CI]	
Overall	FujiLAM	415	40	55	23	297	64·9%	[43·7 – 85·2]	83·8%	[76·5 – 89·4]
	AlereLAM	415	19	40	44	312	30·7%	[8·6 – 61·6]	87·8%	[79·0 – 93·7]
	ΔSn and ΔSp						34·2%		-4·0%	
CRS										
Overall	FujiLAM	415	58	37	118	202	32·9%	[24·6 – 41·9]	83·3%	[71·8 – 91·7]
	AlereLAM	415	36	23	140	216	20·2%	[12·3 – 29·4]	90·0%	[81·6 – 95·6]
	ΔSn and ΔSp						12·7%		-6·7%	

MRS	Test	N	TP	FP	FN	TN	Sensitivity [95% CI]		Specificity [95% CI]	
All HIV+	FujiLAM	61	8	11	7	35	54·8%	[28·7 – 81·5]	75·9%	[61·8 – 86·9]
	AlereLAM	61	5	9	10	37	36·6%	[13·8 – 70·4]	80·4%	[66·3 – 91·0]
	ΔSn and ΔSp						18·2%		-4·5%	
All HIV-	FujiLAM	344	31	40	15	258	67·5%	[41·8 – 88·0]	85·9%	[79·2 – 91·0]
	AlereLAM	344	14	30	32	268	26·6%	[1·2 – 66·4]	89·1%	[80·7 – 94·7]
	ΔSn and ΔSp						40·9%		-3·2%	
CRS	**Test**	**N**	**TP**	**FP**	**FN**	**TN**	**Sensitivity [95% CI]**		**Specificity [95% CI]**	
All HIV+	FujiLAM	61	14	5	30	12	31·9%	[18·9 – 47·0]	71·4%	[46·8 – 91·5]
	AlereLAM	61	13	1	31	16	29·3%	[16·3 – 44·6]	92·8%	[72·6 – 99·8]
	ΔSn and ΔSp						2·6%		-21·4%	
All HIV-	FujiLAM	344	43	28	86	187	33·2%	[23·7 – 43·5]	85·7%	[76·2 – 92·2]
	AlereLAM	344	23	21	106	194	15·3%	[1·7 – 37·5]	89·3%	[81·0 – 94·7]
	ΔSn and ΔSp						17·9%		-3·6%	

MRS	Test	N	TP	FP	FN	TN	Sensitivity [95% CI]		Specificity [95% CI]	
Age <5 years	FujiLAM	194	16	35	10	133	61·8%	[36·6 – 85·5]	78·5%	[69·1 – 86·0]
	AlereLAM	194	9	32	17	136	38·8%	[0·4 – 98·9]	80·5%	[68·3 – 89·4]
	ΔSn and ΔSp						23·0%		-2·0%	
Age >=5 years	FujiLAM	221	24	20	13	164	67·1%	[40·1 – 90·2]	88·8%	[82·1 – 93·6]
	AlereLAM	221	10	8	27	176	26·9%	[7·2 – 54·8]	95·2%	[89·7 – 98·3]
	ΔSn and ΔSp						40·2%		-6·4%	
CRS	**Test**	**N**	**TP**	**FP**	**FN**	**TN**	**Sensitivity [95% CI]**		**Specificity [95% CI]**	
Age <5 years	FujiLAM	194	28	23	55	88	33·3%	[19·8 – 48·3]	78·4%	[66·5 – 87·2]
	AlereLAM	194	20	21	63	90	23·3%	[10·0 – 39·9]	81·4%	[70·4 – 90·4]
	ΔSn and ΔSp						10·0%		-3·0%	
Age >=5 years	FujiLAM	221	30	14	63	114	32·7%	[22·4 – 44·4]	88·2%	[76·0 – 96·4]
	AlereLAM	221	16	2	77	126	17·3%	[9·6 – 27·4]	98·1%	[93·7 – 99·9]
	ΔSn and ΔSp						15·4%		-9·9%	

MRS	Test	N	TP	FP	FN	TN	Sensitivity [95% CI]		Specificity [95% CI]	
Gambia	FujiLAM	167	5	19	5	138	50·0%	[18·7 – 81·3]	87·9%	[81·8 – 92·6]
	AlereLAM	167	1	10	9	147	10·0%	[0·3 – 44·5]	93·6%	[88·6 – 96·9]
	ΔSn and ΔSp						40·0%		-5·7%	
Mali	FujiLAM	67	13	8	10	36	56·5%	[34·5 – 76·8]	81·8%	[67·3 – 91·8]
	AlereLAM	67	6	9	17	35	26·1%	[10·2 – 48·4]	79·5%	[64·7 – 90·2]
	ΔSn and ΔSp						30·4%		2·3%	
Nigeria	FujiLAM	81	10	18	1	52	90·9%	[58·7 – 99·8]	74·3%	[62·4 – 84·0]
	AlereLAM	81	8	6	3	64	72·7%	[39·0 – 94·0]	91·4%	[82·3 – 96·8]
	ΔSn and ΔSp						18·2%		-17·1%	
Tanzania	FujiLAM	100	12	10	5	73	70·6%	[44·0 – 89·7]	88·0%	[79·0 – 94·1]
	AlereLAM	100	4	15	13	68	23·5%	[6·8 – 49·9]	81·9%	[72·0 – 89·5]
	ΔSn and ΔSp						47·1%		6·0%	
CRS	**Test**	**N**	**TP**	**FP**	**FN**	**TN**	**Sensitivity [95% CI]**		**Specificity [95% CI]**	
Gambia	FujiLAM	167	11	13	32	111	25·6%	[13·5 – 41·2]	89·5%	[82·7 – 94·3]
	AlereLAM	167	4	7	39	117	9·3%	[2·6 – 22·1]	94·4%	[88·7 – 97·7]
	ΔSn and ΔSp						16·3%		-4·8%	
Mali	FujiLAM	67	20	1	37	9	35·1%	[22·9 – 48·9]	90·0%	[55·5 – 99·8]
	AlereLAM	67	14	1	43	9	24·6%	[14·1 – 37·8]	90·0%	[55·5 – 99·8]
	ΔSn and ΔSp						10·5%		0·0%	
Nigeria	FujiLAM	81	15	13	25	28	37·5%	[22·7 – 54·2]	68·3%	[51·9 – 81·9]
	AlereLAM	81	11	3	29	38	27·5%	[14·6 – 43·9]	92·7%	[80·1 – 98·5]
	ΔSn and ΔSp						10·0%		-24·4%	
Tanzania	FujiLAM	100	12	10	24	54	33·3%	[18·6 – 51·0]	84·4%	[73·1 – 92·2]
	AlereLAM	100	7	12	29	52	19·4%	[8·2 – 36·0]	81·3%	[69·5 – 89·9]
	ΔSn and ΔSp						13·9%		3·1%	

Table 1: Revised classification of intrathoracic TB case definitions for diagnostic evaluation studies in children*

Case definition	**Criteria**
Confirmed TB	Bacteriological confirmation of Mtb (culture and/or Xpert MTB/RIF assay) from at least 1 respiratory specimen
Unconfirmed TB	Bacteriological confirmation NOT obtained AND at least 2 of the following:
	• Symptoms/signs suggestive of tuberculosis
	• Chest radiograph consistent with tuberculosis
	• Close tuberculosis exposure
	• Positive response to tuberculosis treatment (requires documented positive clinical response to TB treatment—no time duration specified)
Unlikely TB	Bacteriological confirmation NOT obtained AND Criteria for "unconfirmed tuberculosis" NOT met

*All children were followed up for a period of at least six months

Table 2: Demographic and clinical characteristics of study participants

	Gambia		**Mali**		**Nigeria**		**Tanzania**		**All Patients**	
Variable	**(N=167)**		**(N=67)**		**(N=81)**		**(N=100)**		**(N=415)**	
Median age - years (IQR)	5.6	[2.3;9.3]	6	[2.6;11.0]	4.9	[1.7;8.9]	5.4	[2.3;8.8]	5.6	[2.3;9.3]
Age <5 years (%)	67	40.1	28	41.8	43	53.1	56	56.0	194	46.7
Age ≥5 years (%)	100	59.9	39	58.2	38	46.9	44	44.0	221	53.3
Males - no. (%)	98	58.7	44	65.7	30	37.0	53	53.0	225	54.2
Distribution in diagnostic categories										
Confirmed TB - no. (%)	12	7.2	23	34.3	11	13.6	17	17.0	63	15.2
Unconfirmed TB - no. (%)	31	18.6	34	50.7	29	35.8	19	19.0	113	27.2
Unlikely TB - no. (%)	124	74.3	10	14.9	41	50.6	64	64.0	239	57.6
HIV status										
Positive (%)	15	9.0	21	31.3	8	9.9	17	17.0	61	14.7
Negative (%)	148	88.6	46	68.7	68	84.0	82	82.0	344	82.9
Unknown (%)	4	2.4	0	0.0	5	6.2	1	1.0	10	2.4
Anthropometry										
Underweight (%)	74	44.3	40	59.7	36	44.4	31	31.0	181	43.6
Stunted (%)	34	20.4	17	25.4	45	55.6	38	38.0	134	32.3