Journal of Clinical Microbiology

Diagnostic accuracy of stool Xpert MTB/RIF for the detection of pulmonary tuberculosis in children: a systematic review and meta-analysis

Running title: Stool Xpert for diagnosing childhood TB

Emily MacLean[a,b], Giorgia Sulis[a,b], Claudia M. Denkinger[b,c,d], James C. Johnston[b,e,f], Madhukar Pai[a,b], Faiz Ahmad Khan[a,b,g,#]

[a] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Canada
[b] McGill International TB Centre, McGill University, Montreal, Canada
[c] Foundation for Innovative New Diagnostics, Geneva, Switzerland
[d] Department of Infectious Diseases, University of Heidelberg, Heidelberg, Germany
[e] Division of Respiratory Medicine, University of British Columbia, Vancouver, Canada
[f] BC Centre for Disease Control, Vancouver, Canada
[g] Respiratory Epidemiology & Clinical Research Unit, Research Institute of the McGill University Health Centre & Montreal Chest Institute, Montreal, Canada

[#] Corresponding author: Faiz Ahmad Khan
Email: faiz.ahmadkhan@mcgill.ca
Phone: 1-514-398-1934, ext 32117

Abstract

Invasive collection methods are often required to obtain samples for the microbiologic evaluation of children with presumptive pulmonary tuberculosis (PTB). Nucleic-acid amplification testing of easier to collect stool samples could be a non-invasive method of diagnosing PTB. We conducted a systematic review and meta-analysis to evaluate the diagnostic accuracy of testing stool with the Xpert MTB/RIF assay ('stool Xpert') for childhood PTB. Four databases were searched for publications from January 2008 to June 2018. Studies assessing the diagnostic accuracy amongst children of stool Xpert compared to a microbiological reference standard of conventional specimens tested by mycobacterial culture or Xpert were eligible. Bivariate random-effects meta-analyses were performed to calculate pooled sensitivity and specificity of stool Xpert against the reference standard. From 1589 citations, 9 studies (n=1681) were included. Median participant ages ranged from 1.3 to 10.6 years. Protocols for stool processing and testing varied substantially, with differences in reagents and methods of homogenization and filtering. Against the microbiological reference standard, pooled sensitivity and specificity of stool Xpert were 67% (95%CI:52-79) and 99% (95%CI:98-99), respectively. Sensitivity was higher among children with HIV (79%; 95%CI:68-87; versus 60%; 95%CI:44-74 among HIV-uninfected). Heterogeneity was high. Data were insufficient for subgroup analyses amongst children under age 5, the most relevant target population. Stool Xpert could be a non-invasive method of ruling-in PTB in children, particularly those with HIV. However, studies focused on children under 5 are needed, and generalizability of the evidence is limited by the lack of a standardized stool preparation and testing protocol.

Introduction

At least 1 million incident tuberculosis (TB) cases and 230,000 TB-related deaths are estimated to have occurred among children in 2017, accounting for approximately 10% of total cases and 15% of deaths (1). Pulmonary TB (PTB) is the most common form of childhood TB (2). Xpert MTB/RIF® (Xpert) (Cepheid, USA), an automated cartridge-based PCR assay, is currently recommended by the World Health Organization (WHO) as the initial diagnostic test in presumptive PTB cases for adults and children (3). Minimal sample preparation is required, and test results are produced within 2 hours. In a meta-analysis that pooled sputum smear positive and negative subjects, Xpert on respiratory samples had sensitivity of 62% (95% credible interval: 51-73) and specificity of 98 (95% credible interval: 97-99). Xpert on sputum is thus more sensitive than smear microscopy. Moreover, Xpert has several operational advantages over mycobacterial culture, the gold standard for TB diagnosis (4). However, in children under 5 years old—and particularly in those under 2—the collection of sputum specimens is difficult and often requires invasive methods that are challenging to implement in resource-limited settings (e.g. nasopharyngeal/nasogastric aspiration, bronchoscopy) and not widely available (2). Further, as pediatric TB is typically paucibacillary, the sensitivity of currently deployed tests is diminished in children versus adults (5).

Mycobacterium tuberculosis-containing sputum may be swallowed, particularly during sleep, and acid-fast bacilli have been shown to survive digestion and are detectable in stool (6, 7). As such, stool may represent a more acceptable and feasible alternative to conventional specimens for the evaluation of suspected childhood PTB. The use of Xpert on stool has not been included in recommendations by WHO, nor has any claim been made by the manufacturer regarding stool. However, several groups have now developed preprocessing methods in order to use Xpert on stool for the diagnosis of childhood TB.

Journal of Clinical Microbiology

We performed a systematic review and meta-analysis of the diagnostic performance of Xpert using stool samples for PTB in children.

Methods

Protocol and registration

The protocol for this systematic review and meta-analysis was registered at the International Prospective Register of Systematic Reviews (PROSPERO) (CRD42017079836).

Search strategy and Information sources

PubMed, EMBASE, Scopus, and the Cochrane Library were systematically searched from January 1, 2008 until June 15, 2018. The search strategy was developed with a medical librarian and based on key validated terms for “children” and “Xpert”, as well as “tuberculosis” with no filters applied. The full search strategies for each database are presented in Supplementary Text S1. Experts in TB diagnostics were consulted to identify relevant papers that may have been missed by the search strategy. Citations of reviews and included publications were also searched.

Eligibility criteria

Publications in English, French, Italian, Mandarin, Spanish, and Portuguese, of any design and sampling strategy, and of any enrolment timing (prospective, retrospective, cross-sectional) were eligible for inclusion. Conference proceedings and abstracts, commentaries, editorials, and reviews were excluded, as were studies with a sample size less than 10. To be included, eligible studies must have reported the diagnostic performance of stool Xpert in patients under 16 years old, as compared to a microbiological reference standard for the diagnosis of PTB. Studies that did not explicitly state their focus was PTB were eligible if the

types of specimens used for the reference standard were those that are typically used for PTB diagnosis (e.g. gastric aspirate). Studies that used banked sputum and stool specimens originally collected from children were also eligible.

Study screening and selection

Search results were imported into a citation manager and duplicates were removed. Two authors (E.M., G.S.) independently screened citations by title and abstract per pre-defined eligibility criteria, followed by full-text review for all selected studies. Results disagreed upon were discussed, and a third reviewer consulted if necessary (F.A.K.).

Data extraction

A data extraction form was piloted by two reviewers (E.M., G.S.) with critical input from a third (C.M.D.). Two reviewers (E.M., G.S.) independently extracted results using a standardized form (Supplemental Text S2) from all included studies. After data extraction, results were compared, and disagreements discussed until a consensus was reached. Study authors were contacted for missing performance data, clarification regarding reference standard definitions, and sample preparation techniques. Using these data and figures indicated in the publications, we reconstructed two-by-two tables for stool Xpert performance compared to the microbiological reference standard and, where applicable, the clinical reference standard.

Risk of bias assessment

The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool (8) was used to assess each included study's risk of bias. No formal assessment of publication bias was made, as traditional methods such as funnel plots and regression tests are not helpful for diagnostic studies (9).

Reference standards

Acceptable microbiological reference standards were mycobacterial culture, or Xpert MTB/RIF, performed on specimens that are conventionally used to diagnose childhood PTB (nasogastric aspirates, gastric lavage, nasopharyngeal aspirates, expectorated sputum). No studies included stool mycobacterial culture in their diagnostic work-up. Stool Xpert was not included in the reference standard.

Childhood PTB is often clinically diagnosed (i.e. without microbiologic confirmation). As such, we also examined the performance of stool Xpert compared to clinical reference standards that are compatible with updated international guidelines (5). Studies that followed these guidelines used a combination of signs and symptoms, chest radiography, epidemiologic history, and tuberculin skin test results, to classify children as “likely TB”, “unconfirmed TB”, and “unlikely TB” (Supplementary Table S1). For our purposes, we dichotomized these outcomes into “likely/possible TB” and “unlikely TB”.

Statistical Analysis

Data from reconstructed two-by-two tables were used to calculate sensitivity and specificity and associated 95% confidence intervals (CIs). In cases of empty cells in two-by-two tables, a zero correction was made by replacing the cell with 0.5. Aggregate data meta-analyses were performed with bivariate random effect hierarchical models (10) to estimate pooled sensitivity and specificity for stool Xpert compared to the microbiologic reference standard, and separately, compared to the clinical reference standard. We also estimated pooled sensitivity and specificity stratified by HIV status. Results from individual studies and pooled estimates were presented on forest plots. To assess between-study heterogeneity, we used the I^2-statistic (11). In a sensitivity analysis, we estimated pooled sensitivity and specificity after excluding studies that used Xpert MTB/RIF but not mycobacterial culture of conventional

specimens as the microbiological reference standard. All analyses were conducted using the *Midas* package in STATA (STATA 15, Stata Corp., USA (12)). The study is reported following PRISMA guidelines (Supplementary Table S2) (13).

Results

Search results

Our search identified 1589 unique citations from which 34 studies were selected for full-text review, and 9 studies met inclusion criteria (Figure 1).

Study and participant characteristics

Study and patient characteristics are presented in Table 1. Among the 9 studies we included, African countries were most well-represented (7/9), whereas 2 studies recruited participants from Asia. One study had multiple sites across two continents, whereas the others were single-country studies. In total, 1681 children from 9 studies were included in our meta-analysis of stool Xpert's diagnostic performance compared to a microbiologic reference standard, and 869 children from 5 studies were included in the comparison against a clinical reference standard. Prevalence of microbiologically confirmed cases per study ranged widely, from 2.6% (21) to 54% (14). The prevalence of clinically confirmed or unconfirmed cases was much higher, ranging from 35% (16) to 100% (15). Supplementary Table S1 provides details on clinical reference standard definitions of included studies.

Studies enrolled children from 0 to 16 years. The ratio of females to males was generally balanced. Participants with a documented history of TB disease contact, when reported (5/9 studies), ranged from 12% (17) to 56% (22). Most studies did not include information about tuberculin skin test (TST) results. Two studies only included children with HIV

(17, 18) and two restricted enrolment to HIV negative children (16, 19); the remainder had a mixed population.

Sample processing

Table 2 shows the sample preparation steps utilized in each study. In one study (22), two sample preparation methods were attempted, with results ultimately pooled. Most studies (6/9) obtained one stool sample from enrolled children, typically within 24 hours of obtaining respiratory samples. Samples were either used immediately or stored for later use, except for one study (18) which used some samples immediately and some after freezing, and a second study (22) which stored samples collected at the child's home and immediately used those collected at the healthcare center. As information on sample storage was not available for all studies, sub-group analysis could not be performed per sample storage method.

The mass of stool utilized, and its collection method, varied: 0.15g of bulk stool (16); 0.15g sterile loop (15); flocked rectal swab (20); 0.5g (19); 0.6g (14); 2g (18); 5g (22). A diluent solution, such as PBS or distilled water or sucrose solution, was added to the stool before a homogenization, in variable quantities, typically followed by vortexing. Most studies (6/9) reported a period of sample settling before further work-up. Final sample preparation methods were quite varied, but included either centrifugation or filtering through syringe filter or gauze, primarily to remove large particles, before final addition of the sample into the Xpert cartridge (Table 2).

Quality assessment

Figure 2 displays the overall risk of bias and applicability concerns of the 9 studies included in our meta-analysis. Supplementary Figure S1 presents the individual studies' quality assessment results. In the patient selection domain (Figure 2), five studies were at low risk of bias and one study (14) was at high risk of bias due to its use of a case-control design,

whereas the remaining eight were either cross-sectional or cohort studies. Risk of bias was high for one study because of convenience sampling (16), and unclear in two studies because of an unclear sampling strategy and the inappropriate exclusions of certain children (15, 19). With respect to applicability, the majority of studies (Table 1) included children who presented with symptoms suggestive of TB. Two studies (17, 18) only included children with HIV and, because it is known that Xpert performs differentially in those who are HIV-infected (23), these were scored for applicability concerns as high. One study (14) only tested samples from confirmed TB cases and non-cases, which does not represent a typical clinical scenario, so we also rated applicability concerns as high.

The conduct of the index test generally was at low risk of bias, as Xpert is an automated assay with a predefined cut-off of detection that produces a binary response. However, since there is no standardized operating protocol for stool samples and no internationally-recommended procedure for sample storage and processing, applicability concerns regarding the index test's conduct are unclear (Figure 2).

In light of the inherent limitations of microbiologic tests for diagnosing childhood PTB, we classified 8/9 studies as having an unclear risk of bias with respect to correctly classifying the target condition despite having used culture as the reference test. The exception was one study which was scored as high risk of bias as its microbiological reference standard did not include culture. Culture and Xpert are both automated assays, so we scored the risk of bias as low regarding test interpretation. Additionally, all studies' reference standards were performed in regional or central reference laboratories, so we expect bias from operator error to be of low concern. Applicability concerns were uniformly unclear.

We scored the risk of bias as low for all studies with respect to the appropriateness of the time interval between index test and reference standard, as all studies reported running stool Xpert within 7 days of specimen collection (Figure 2).

Meta-analysis of diagnostic accuracy

For the comparison against the microbiological reference standard, sensitivities of stool Xpert varied from 32% (22) to 85% (14), while specificity was uniformly very high (Figure 3A). Pooled sensitivity was 67% (95% CI [52-79]) and pooled specificity was 99% (95% CI [98-99]). I^2 values for sensitivity and specificity were 83% (95% CI [72-93]) and 62% (95% CI [35-90]), respectively, indicating high between-study heterogeneity, particularly for sensitivity. For the clinical reference standard comparison, the pooled sensitivity of stool Xpert was 22% (95% CI [9.0-44]) while specificity was 100% (95% CI [66-100]) (Figure 3B).

Although 7/9 studies included children with HIV, only 5/9 studies provided sufficient information to construct two-by-two tables (14, 15, 17, 19, 20) (2 of these studies enrolled only children with HIV (17, 19)) (Figure 3C). One study (14) did not provide sufficient information to calculate specificity amongst children with HIV. Data from children that were HIV-negative were available from 5 studies (15, 16, 18, 20, 21) (Figure 3D). Using the microbiologic reference standard, among children with HIV, sensitivity of stool Xpert was 79% (95% CI [68-87]) and pooled specificity was 99% (95% CI [94-100]) (Figure 3C); amongst those without HIV, sensitivity was 60% (95% CI [44-79]) and specificity 99% (95% CI [97-100]) (Figure 3D). For both sensitivity and specificity, I^2 values were lower in HIV stratified analyses as compared to when all studies were pooled (Table 3), suggesting that HIV partially explained the between-study heterogeneity.

Results of the sensitivity analysis in which we excluded the study that did not use mycobacterial culture as part of the reference standard (14) are presented in Supplementary figure S2. Pooled sensitivity and specificity estimates combining all studies and stratified by HIV status were all similar to those estimated in our main analyses, as was between-study heterogeneity. Pooled estimates from our main analysis and from this sensitivity analysis are summarized in Table 3.

We undertook two post-hoc sensitivity analyses. In the first, we sought to determine whether the quantity of stool used for testing was associated with diagnostic accuracy (assuming greater mass might increase sensitivity). Studies were too few to estimate pooled accuracy stratified by stool mass used, however, visual inspection of forest plots found no obvious trend to support a minimum quantity (Supplementary Figure S3). In the second sensitivity analysis, we evaluated whether the burden of TB in the country where a study was conducted was associated with accuracy of stool Xpert. As shown in Supplementary Figure S4, there was no clear trend to suggest such an association.

Discussion

In this systematic review and meta-analysis, we found that the sensitivity and specificity of stool Xpert (67% (95% CI [52-79]) and 99% (95% CI [98-99]), respectively) for the diagnosis of microbiologically-confirmed childhood PTB were comparable to what has been reported for Xpert on respiratory specimens (62% (95% credible interval [51-73]) and 98% (95% credible interval [97-99]), respectively) (4). Sensitivity and specificity varied by HIV status. As stool collection is noninvasive, this is of substantial interest for the medical evaluation of children with suspected PTB—but a number of limitations of the existing evidence highlight the need for more research, and greater standardization of testing, before policy formulation.

Amongst the most important limitations of the evidence base is the lack of data on performance in the subpopulation of children in whom stool Xpert is of greatest potential clinical utility —those under age 5, and especially the subgroup under age 2. Only one study compared accuracy between age categories, and a cutoff of 10 years old was used (15).

We observed substantial between-study heterogeneity in diagnostic accuracy, mostly for sensitivity. Different approaches to participant selection likely contributed to this, in particular the use of case-control (14) and non-consecutive sampling (16, 19) which are at a

higher risk of introducing bias into a study. Data also suggested that heterogeneity was partly explained by differences in the prevalence of HIV infection. The higher sensitivity of stool Xpert among children with HIV has also been observed for other specimen types in this population (4, 24), perhaps as a result of more severe TB disease in HIV/TB coinfected children.

We found substantial variability in protocols for performing stool Xpert, with each study taking a unique approach. Differences were seen at all steps: 1) at stool collection, different methods of sampling, numbers of specimens, and volumes of stool were used; 2) differing reagents were added to stool samples before homogenization, and all studies utilized different additional reagents; 3) dissimilar filtration methods and decontamination steps were adopted. Future studies should ensure, at minimum, complete reporting of protocols for stool collection processing and testing. A standardized protocol would be of value, as would a standardized stool collection and processing kit.

Our systematic review and meta-analysis has a number of strengths. First, all included studies reported using a microbiological reference standard for the comparison of stool Xpert, and 8 out of 9 studies used liquid or solid culture. While the imperfect nature of any reference standard for diagnosing pediatric TB means that the true number of affected children is always unknown, the accuracy of stool Xpert against microbiological confirmation is likely a closer estimation of its true accuracy than its performance compared to the clinical reference standard (as symptoms of PTB are non-specific). Second, by systematically assessing each study's sample preparation and processing techniques we found substantial variability in methods of performing stool Xpert, and were also able to identify obstacles to implementation. For example, most protocols required at least one centrifugation step, which is inauspicious in terms of translating this assay to a lower healthcare system level. Lastly, we utilized a sensitive and validated search strategy that covered six languages.

The present work also has some limitations. First, data were insufficient and there were too few studies for us to perform stratified or meta-regression analyses to assess most demographic-related potential causes of observed heterogeneity. Hence we suggest that in addition to HIV-stratified results, future studies of stool Xpert should also ensure reporting is stratified by age, gender, and extent of radiographic disease. Second, while we identified a wide variability in sampling and stool processing, we could not explore these as sources of heterogeneity or determine if any processing work-flows were potentially superior. Third, we did not include one study concerning stool Xpert in children (25) that was published after our systematic search was completed and therefore not included in our meta-analysis. However, including it in our pooled analyses did not significantly alter sensitivity or specificity estimates (Supplementary Figure S5). Finally, our pooled estimates came from study populations with a high prevalence of TB, hence it is possible that these estimates may not be generalisable to settings of lower TB burden.

Given that these preliminary studies of stool Xpert suggest high specificity and moderate sensitivity, its potential role in the diagnostic pathway would be as a first-line rule-in test, rather than as a triage test to rule-out PTB. Studies assessing whether stool Xpert has value as an add-on test in combination with currently deployed assays will be useful, as will studies assessing the effect of repeat testing on sensitivity.

Conclusion

Preliminary data suggest Xpert on stool specimens may be potentially useful as a rule-in test, but a standardized stool sample preparation protocol is lacking, and the accuracy of stool Xpert in children under 5 years old—the subgroup in whom the test could bring the most added value—remains largely unknown.

Journal of Clinical Microbiology

Funding

This work was supported by the Foundation for Innovative New Diagnostics (FIND), Geneva, Switzerland [grant number OCC2015-412]. FAK receives salary support from the Fonds de Récherche du Québec Santé.

Acknowledgements

The authors are grateful to Samuel Schumacher for his early critical comments regarding study protocol development.

We attest that all authors of this research paper have directly participated in the planning, execution, or analysis of the study and have seen and approved the manuscript.

Conflicts of Interest

Emily MacLean – no conflict

Giorgia Sulis – no conflict

Claudia M. Denkinger – Dr. Denkinger is employed by FIND. FIND is a non-for-profit foundation, whose mission is to find diagnostic solutions to overcome diseases of poverty in LMICs. It works closely with the private and public sectors and receives funding from some of its industry partners. It has organizational firewalls to protect it against any undue influences in its work or the publication of its findings. All industry partnerships are subject to review by an independent Scientific Advisory Committee or another independent review body, based on due diligence, TTPs and public sector requirements. FIND catalyzes product development, leads evaluations, takes positions, and accelerates access to tools identified as serving its mission. It

provides indirect support to industry (e.g., access to open specimen banks, a clinical trial platform, technical support, expertise, laboratory capacity strengthening in LMICs, etc.) to facilitate the development and use of products in these areas. FIND also supports the evaluation of prioritized assays and the early stages of implementation of WHO-approved (guidance & PQ) assays using donor grants. In order to carry out test validations and evaluations, has product evaluation agreements with several private sector companies for the diseases FIND works in which strictly define its independence and neutrality vis-a-vis the companies whose products get evaluated, and describes roles and responsibilities.

James C. Johnston – no conflict

Madhukar Pai – Dr. Pai reports no commercial or financial conflicts. He serves on the Scientific Advisory Committee of FIND, Geneva, a non-profit foundation that works on diagnostics.

Faiz Ahmad Khan – no conflict

References

1. World Health Organization (WHO). 2018. Global Tuberculosis Report 2018. World Health Organization, Press W, Geneva, Switzerland.
2. Marais BJ, Gie RP, Hesseling AC, Schaaf HS, Lombard C, Enarson DA, Beyers N. 2006. A refined symptom-based approach to diagnose pulmonary tuberculosis in children. Pediatrics 118:e1350-9.
3. World Health Organization. 2013. Policy Update. Automated real-time nucleic acid amplification technology for rapid and simultaneous detection of tuberculosis and rifampicin resistance: Xpert MTB/RIF assay for the diagnosis of pulmonary and extrapulmonary TB in adults and children. World Health Organization, Geneva, Switzerland.
4. Detjen AK, DiNardo AR, Leyden J, Steingart KR, Menzies D, Schiller I, Dendukuri N, Mandalakas AM. 2015. Xpert MTB/RIF assay for the diagnosis of pulmonary tuberculosis in children: a systematic review and meta-analysis. Lancet Respir Med 3:451-61.
5. Graham SM, Cuevas LE, Jean-Philippe P, Browning R, Casenghi M, Detjen AK, Gnanashanmugam D, Hesseling AC, Kampmann B, Mandalakas A, Marais BJ, Schito M, Spiegel HM, Starke JR, Worrell C, Zar HJ. 2015. Clinical Case Definitions for Classification of Intrathoracic Tuberculosis in Children: An Update. Clin Infect Dis 61Suppl 3:S179-87.
6. Cordova J, Shiloh R, Gilman RH, Sheen P, Martin L, Arenas F, Caviedes L, Kawai V, Soto G, Williams DL, Zimic M, Escombe AR, Evans CA. 2010. Evaluation of molecular tools for detection and drug susceptibility testing of Mycobacterium tuberculosis in stool specimens from patients with pulmonary tuberculosis. J Clin Microbiol 48:1820-6.

Journal of Clinical Microbiology

7. Donald PR, Schaaf HS, Gie RP, Beyers N, Sirgel FA, Venter A. 1996. Stool microscopy and culture to assist the diagnosis of pulmonary tuberculosis in childhood. J Trop Pediatr 42:311-2.

8. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM. 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155:529-36.

9. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. 2010. Chapter 10: Analysing and Presenting Results. *In* Collaboration TC (ed), Cochrane Handbok for Systematic Reviews of Diagnostic Test Accuracy, Version 1.0 ed. The Cochrane Collaboration.

10. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. 2005. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 58:982-90.

11. Higgins JP, Thompson SG, Deeks JJ, Altman DG. 2003. Measuring inconsistency in meta-analyses. Bmj 327:557-60.

12. Dwamena B. 2007. MIDAS: Stata module for meta-analytical integration of diagnostic test accuracy studies, vol S456880.

13. Moher D, Liberati A, Tetzlaff J, Altman DG. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 151:264-9, w64.

14. Banada PP, Naidoo U, Deshpande S, Karim F, Flynn JL, O'Malley M, Jones M, Nanassy O, Jeena P, Alland D. 2016. A novel sample processing method for rapid detection of tuberculosis in the stool of pediatric patients using the Xpert MTB/RIF assay. PLoS ONE 11 (3) (no pagination).

15. Chipinduro M, Mateveke K, Makamure B, Ferrand RA, Gomo E. 2017. Stool Xpert(R) MTB/RIF test for the diagnosis of childhood pulmonary tuberculosis at primary clinics in Zimbabwe. Int J Tuberc Lung Dis 21:161-166.

16. Hasan Z, Shakoor S, Arif F, Mehnaz A, Akber A, Haider M, Kanji A, Hasan R. 2017. Evaluation of Xpert MTB/RIF testing for rapid diagnosis of childhood pulmonary tuberculosis in children by Xpert MTB/RIF testing of stool samples in a low resource setting. BMC Res Notes 10:473.

17. LaCourse SM, Pavlinac PB, Cranmer LM, Njuguna IN, Mugo C, Gatimu J, Stern J, Walson JL, Maleche-Obimbo E, Oyugi J, Wamalwa D, John-Stewart G. 2018. Stool Xpert MTB/RIF and urine lipoarabinomannan for the diagnosis of tuberculosis in hospitalized HIV-infected children. Aids 32:69-78.

18. Marcy O, Ung V, Goyet S, Borand L, Msellati P, Tejiokem M, Nguyen Thi NL, Nacro B, Cheng S, Eyangoh S, Pham TH, Ouedraogo AS, Tarantola A, Godreuil S, Blanche S, Delacourt C. 2016. Performance of Xpert MTB/RIF and Alternative Specimen Collection Methods for the Diagnosis of Tuberculosis in HIV-Infected Children. Clinical Infectious Diseases 62:1161-1168.

19. Moussa H, Bayoumi FS, Mohamed AM. 2016. Gene Xpert for Direct Detection of Mycobacterium Tuberculosis in Stool Specimens from Children with Presumptive Pulmonary Tuberculosis. Ann Clin Lab Sci 46:198-203.

20. Nicol MP, Spiers K, Workman L, Isaacs W, Munro J, Black F, Zemanay W, Zar HJ. 2013. Xpert MTB/RIF testing of stool samples for the diagnosis of pulmonary tuberculosis in children. Clinical Infectious Diseases 57:e18-e21.

21. Orikiriza P, Nansumba M, Nyehangane D, Bastard M, Mugisha IT, Nansera D, Mwanga-Amumpaire J, Boum Y, 2nd, Kumbakumba E, Bonnet M. 2018. Xpert MTB/RIF diagnosis of childhood tuberculosis from sputum and stool samples in a high TB-HIV-prevalent setting. Eur J Clin Microbiol Infect Dis doi:10.1007/s10096-018-3272-0.

22. Walters E, Van Der Zalm MM, Palmer M, Bosch C, Demers AM, Draper H, Goussard P, Schaaf HS, Friedrich SO, Whitelaw A, Warren R, Gie RP, Hesseling AC. 2017. Xpert MTB/RIF on Stool Is Useful for the Rapid Diagnosis of Tuberculosis in Young Children with Severe Pulmonary Disease. Pediatric Infectious Disease Journal 36:837-843.

23. Steingart KR, Schiller I, Horne DJ, Pai M, Boehme CC, Dendukuri N. 2014. Xpert(R) MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. Cochrane Database Syst Rev doi:10.1002/14651858.CD009593.pub3:Cd009593.

24. World Health Organization. 2016. The use of the Xpert MTB/RIF assay for the diagnosis of TB. Meeting Report. World Health Organization, Geneva, Switzerland.

25. Walters E, Scott L, Nabeta P, Demers AM, Reubenson G, Bosch C, David A, van der Zalm M, Havumaki J, Palmer M, Hesseling AC, Ncayiyana J, Stevens W, Alland D, Denkinger C, Banada P. 2018. Molecular detection of Mycobacterium tuberculosis from stool in young children using a novel centrifugation-free processing method. J Clin Microbiol doi:10.1128/jcm.00781-18.

Figure legends

Fig.1: PRISMA study flow diagram.

Fig. 2: QUADAS-2 risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across the 9 included studies.

Fig. 3A: Forest plots of stool Xpert's diagnostic performance compared to a microbiological reference standard of culture or Xpert positivity on respiratory samples. Two studies (17, 18) presented results from "intention-to-treat" (ITT) analyses, where any child who produced any sample was included, as well as "per protocol" analyses, where only children who produced all requested samples were included. In these instances, we meta-analysed the ITT results to avoid selection bias.

Fig. 3B: Forest plots of stool Xpert's diagnostic performance compared to a clinical reference standard of "likely/possibly TB" or "unlikely TB".

Fig. 3C: Forest plots of diagnostic performance of stool Xpert in children with HIV compared to a microbiological reference standard.

Fig. 3D: Forest plots of diagnostic performance of stool Xpert in HIV-negative children compared to a microbiological reference standard.

451 **Tables**

452 **Table 1**: Features of included studies and participants. Studies that included separate

453 comparisons of stool Xpert for microbiological and clinical reference standards have two rows.

Journal of Clinical Microbiology

Study	Location	No. eligible children	Age in years (range, median [IQR])	No. patients (%)					Clinical features reported	EPTB status, No. EPTB (%)	Reference standard	Samples used for reference standard	Total included in analysis	No. microbiologically confirmed (%)	No. clinically confirmed / unconfirmed cases (%)	No. clinically unlikely TB (%)	No. contaminated cultures (%)
				Females	TB history	TB contact history	TST positive	HIV-positive									
Banada 2016 (14)	South Africa	40	0-15, NR	21/38 (55)	NR	16/38 (42)	NR	16/38 (42)	Cough, EP symptoms, Weight loss	PTB only	**Xpert**	IS, GA	37	20 (54)	-	-	NR
Chipinduro 2017 (15)	Zimbabwe	218	5-16, 10.6 [8-13]	123/218 (56)	17/218 (7.8)	51/218 (23)	NR	111/198 (56)	Cough, Weight loss, Night sweats, Fever, Appetite loss	PTB only[a]	**Culture[c] /Xpert**	IS	218	19 (8.7)	-	-	NR
											CRS[b]	-	32	-	32 (100)	0 (0)	NR
Hasan 2017 (16)	Pakistan	50	0-15, 6.8 [2-9]	22/50 (44)	NR	27/50 (54)	NR	0/50 (0)	Cough, EP symptoms, Weight loss	PTB only	**Culture[d] /Xpert**	Sputum, GA	49	11 (22)	-	-	NR
											CRS[b]	-	49	-	17 (35)	32 (65)	NR
Lacourse 2018 (17)	Kenya	165	0-12, 2 [1.1-4.8]	75/165 (45)	NR	20/162 (12)	7/151 (4.6)	165/165 (100)	Cough, Lethargy, Fever, Failure to thrive	PTB only[a]	**Culture[f] /Xpert**	Sputum, GA	147	11 (7.5)	-	-	NR
											CRS[b]	-	165	-	85 (52)	80 (48)	NR
Marcy 2016 (18)	Burkina Faso, Cambodia, Cameroon, Vietnam	272	0-13, 7.2 [4.1-7.2]	132/272 (49)	49/272 (18)	58/272 (21)	50/272 (18)	272/272 (100)	Cough, Weight loss, Lethargy, Fever, Broad spectrum Abx failure, CXR abnormality	PTB only[a]	**Culture[e]**	GA, IS, NS, string	272	27 (10)	-	-	NR
											CRS[b]	-	272	-	245 (90)	27 (10)	NR
Moussa 2016 (19)	Egypt	115	1-16, NR	45/115 (39)	NR	29/115 (25)	13/67 (19)	0/115 (0)	Cough, Weight loss, Night sweats, Fever, CXR abnormality	PTB only	**Culture[c]**	Sputum, IS	115	36 (31)	-	-	0/115 (0)

Nicol 2013 (20)	South Africa	115	1-15, 2.6 [1.6-4.8]	NR	0/115 (0)	NR	NR	17/115 (15)	Cough, Weight loss, CXR abnormality	PTB only	**Culture**[d]	IS	115	17 (15)	-	-	NR
Orikiriza 2018 (21)	Uganda	357	1-14, NR	178/392 (45)	8/392 (2.0)	76/391 (19)	99/383 (26)	121/388 (31)	Cough, Weight loss, Night sweats, Lethargy, Fever	PTB only[a]	**Culture**[e] **/Xpert**	Sputum, IS	349	9 (2.6)	-	-	6/357 (1.7)
Walters 2017 (22)	South Africa	379	0-13, 1.3 [0.8-2.4]	184/379 (49)	27/379 (7.1)	214/379 (56)	82/294 (28)	51/379 (13)	Cough, Weight loss, Fever	Mix of EPTB and PTB, 35/379 (9.2)	**Culture**[d] **/Xpert**	GA, IS, NA, string	379	72 (19)	-	-	NR
											CRS[b]	-	351	-	242 (69)	109 (31)	NR

Table 1 Footnotes: [a] =implied only pulmonary TB cases based on collection of respiratory samples only; [b] = definitions of each clinical reference standard are given in Supplementary Table S1; [c] = Lowenstein-Jensen solid culture; [d] = BACTEC MGIT liquid culture; [e] = both Lowenstein-Jensen solid cultures and MGIT liquid culture; [f] = MGIT liquid culture, then positive samples were sub-cultured on Lowenstein-Jensen for 3 additional weeks. Abbreviations: Abx = antibiotics; CRS = clinical reference standard; CXR = chest x-ray; EP = extrapulmonary; EPTB = extrapulmonary TB; GA = gastric aspirate; IQR = interquartile range; IS = induced sputum; NA = nasopharyngeal aspirate; No. = number of; NR = not reported; TST = tuberculin skin test.

464 **Table 2**: **Details of stool sample storage and processing for each of the included studies.**

Journal of Clinical Microbiology

Study	No. samples collected, mass	Stool sample collection timing	Imme -diate use?	Storage details	Stool mass used for Xpert	First reagent(s) added to stool	Homogen -isation	Specimen settling	Additional reagents and or filtering / processing	Pellet processing	Final sample into cartridge
Banada 2016	1, 5g	NR	No	4°C for 7 days	0.6g	2mL processing buffer (AL buffer, 10% povidone), 2mL Xpert buffer	Vortex with glass beads	30min at RT	All syringe filtered	No pellet	2mL added to cartridge
Chipinduro 2017	1, 5g	Within 24hr of respiratory sample	No	4°C for max 2 days	0.15g using sterile loop	2.4mL PBS	Vortex	20min at RT	1mL supernatant taken, centrifuged at 3200rpm for 15min	Pellet resuspended in 1mL PBS	Diluted 2:1 in buffer, added to cartridge
Hasan 2017	1, NR	Within 24hr of respiratory sample	No	2-8°C for NR days, taken to 3e hospital, stored at -80°C	0.15g	2.4mL PBS	Vortex	20min at RT	1mL supernatant taken, centrifuged at 3500rpm for 15min	Pellet resuspended in 1mL PBS	Diluted 2:1 in buffer, added to cartridge
Lacourse 2018	1, 2-15g	Within 24hr of respiratory sample	Yes	NA	NR	Equal volume PBS	Manual homogen-isation	12 to 48h at 2-5°C	All filtered through fine filter, vortexed; added to equal volume NaOH-NALC; NRmL PBS added to 40 mL and centrifuged, twice	Pellet resuspended in 1.4mL PBS by vortex	Diluted 2:1 in buffer, added to cartridge
Marcy 2016	NR, 0.5g	NR	Both	Some frozen at NR for NR days	0.5g	10mL Sheather's solution (28% sucrose)	Manual homogen-isation, Vortex 30 sec	NR	All filtered through funnel gauze; centrifuged at 100g for 1min;	No pellet	0.5mL supernatant, 1.8mL buffer added to cartridge; sit 15min at RT; shake; run
Moussa 2016	2, 2g	NR	Yes	NA	2g	10mL distilled H_2O	Vortex	NR	NRmL supernatant taken, centrifuged at 4000rpm for 20min	Pellet decontaminated in 10mL 3% NALC-NaOH for 15min at RT; added to 40mL PBS; centrifuged 20min; pellet resuspended in 1mL PBS	Diluted 2:1 in buffer, added to cartridge
Nicol 2013	1, NR	"At baseline"	No	-80°C within 2hr for max 6 months	0.15g using FLOQ Swabs	2.4mL PBS	Vortex	20min at RT	1mL supernatant taken, centrifuged at 3200rpm for 15min	Pellet resuspended in 1mL PBS	Diluted 2:1 in buffer, added to cartridge
Orikiriza 2018	1, NR	NR	Yes	NA	NR	Saline solution	Vortex	5min at RT	5mL mixture taken, added to NaOH-NALC, vortexed, stand for 20min; PBS added to 50mL and centrifuged at 3000g for 20min at 4°C	Pellet decontaminated with NaOH-NALC method; respun; pellet resuspended in 1.5mL unspecified buffer	0.5mL added to cartridge
Walters 2017	1, 0.3-5g	Within 7 days of respiratory sample	Both	2-8°C for max 3 days if collected at home	<5g	20mL PBS	Vortex	No	5mL mixture taken, added to NALC-NaOH	"Concentration"	Diluted 2:1 in buffer, added to cartridge
	1, 0.3-5g	Within 7 days of respiratory sample	Both	2-8°C for max 3 days if collective at home	1-4g	10mL PBS	Vortex	No	All centrifuged at 3000g at 4°C for 20min	Pellet resuspended in 10mL by vortex for 20sec; centrifuged at 2000g for 1sec, keep supernatant	1mL supernatant added to cartridge

465

466

467 **Table 2 Footnotes**: Abbreviations – 3e = tertiary; max = maximum; NA = not applicable; No. = number of; NR = not reported; PBS =

468 phosphate buffered saline; RT = room temperature; NALC-NaOH = N-Acetyl-l-Cysteine–Sodium Hydroxide; tx = anti TB treatment.

469 **Table 3**: **Results of meta-analyses for estimated stool Xpert sensitivity and specificity.**

	Main results			Sensitivity analysis excluding study that did not use culture as reference standard		
	No. studies included (no. children included)	**Pooled sensitivity (95% CI); I^2 statistic (95% CI)**	**Pooled specificity (95% CI); I^2 statistic (95% CI)**	**No. studies included (no. children included)**	**Pooled sensitivity (95% CI); I^2 statistic (95% CI)**	**Pooled specificity (95% CI); I^2 statistic (95% CI)**
Stool Xpert against microbiological reference standard	9* (1681)	67% (52-79); 83 (72-93)	99% (98-99); 62 (35-90)	8† (1644)	64% (49-76); 81 (69-93)	99% (98-100); 61 (31-91)
Stool Xpert against clinical reference standard	5** (869)	22% (9.0-44); 95 (92-98)	100% (66-100); 78 (59-97)	Not applicable	Not applicable	Not applicable
Stool Xpert against microbiological reference standard in children with HIV	5*** (395)	79% (68-87); 0 (0-100)	99% (94-100); 35 (0-99)	5†† (379)	80% (68-88); 0 (0-100)	99 (94-100); 51 (0-100)
Stool Xpert against microbiological reference standard in HIV-negative children	7**** (974)	61% (40-79); 39 (0-100)	99% (98-100); 56 (13-100)	Not applicable	Not applicable	Not applicable

470

471 I^2-statistic was used to quantify the effect of between study heterogeneity. Abbreviations: CI =

472 confidence interval; no. = number of. References: *(14-22); **(15-18, 22); ***(14, 15, 17, 19,

473 20); ****(14-16, 18, 20-22); †(15-22); ††(15, 17, 19, 20)

474

475

476

477

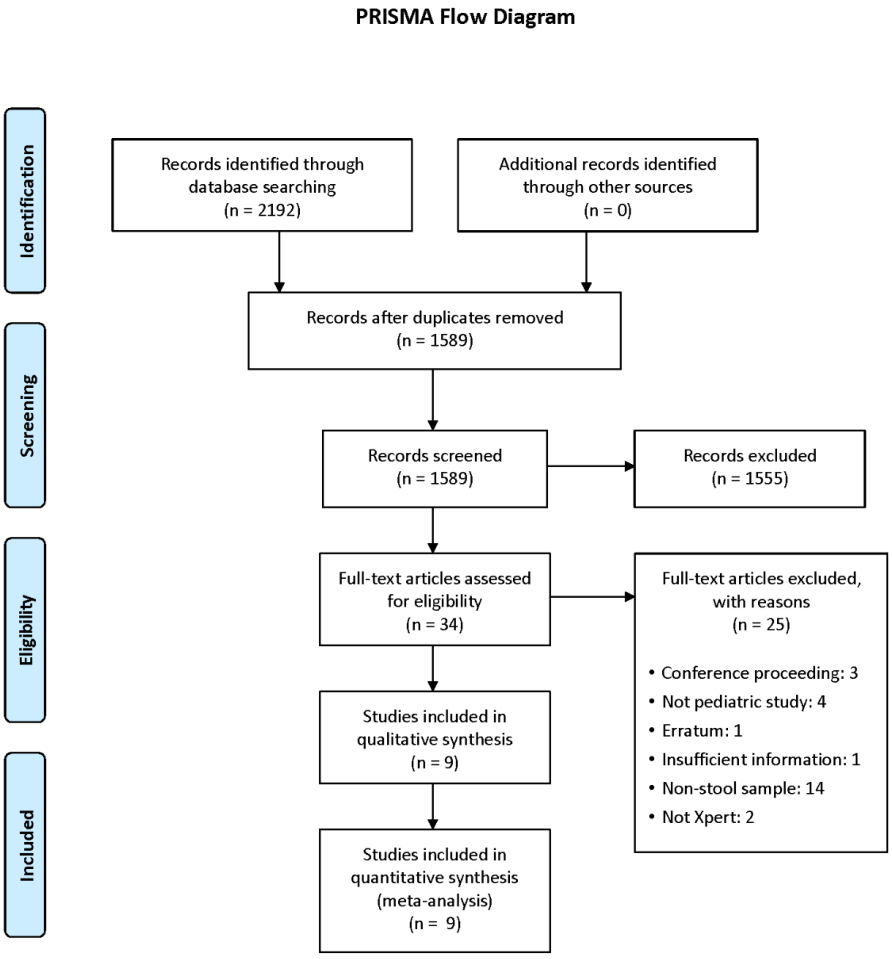

Fig.1: PRISMA study flow diagram.

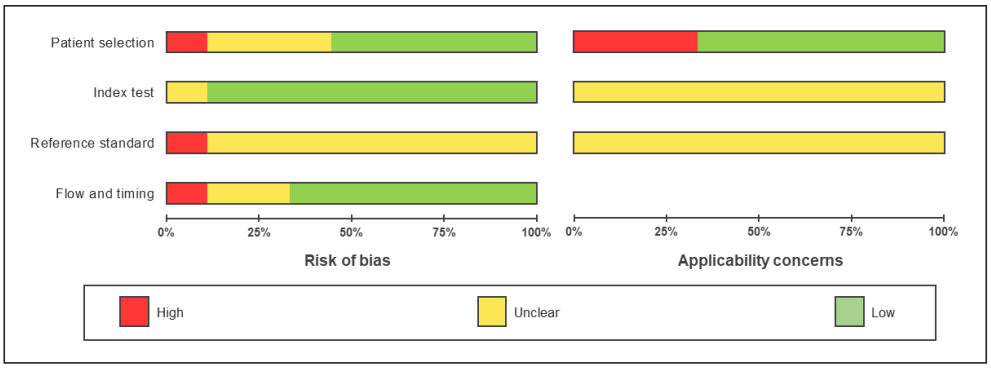

Fig. 2: QUADAS-2 risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across the 9 included studies.

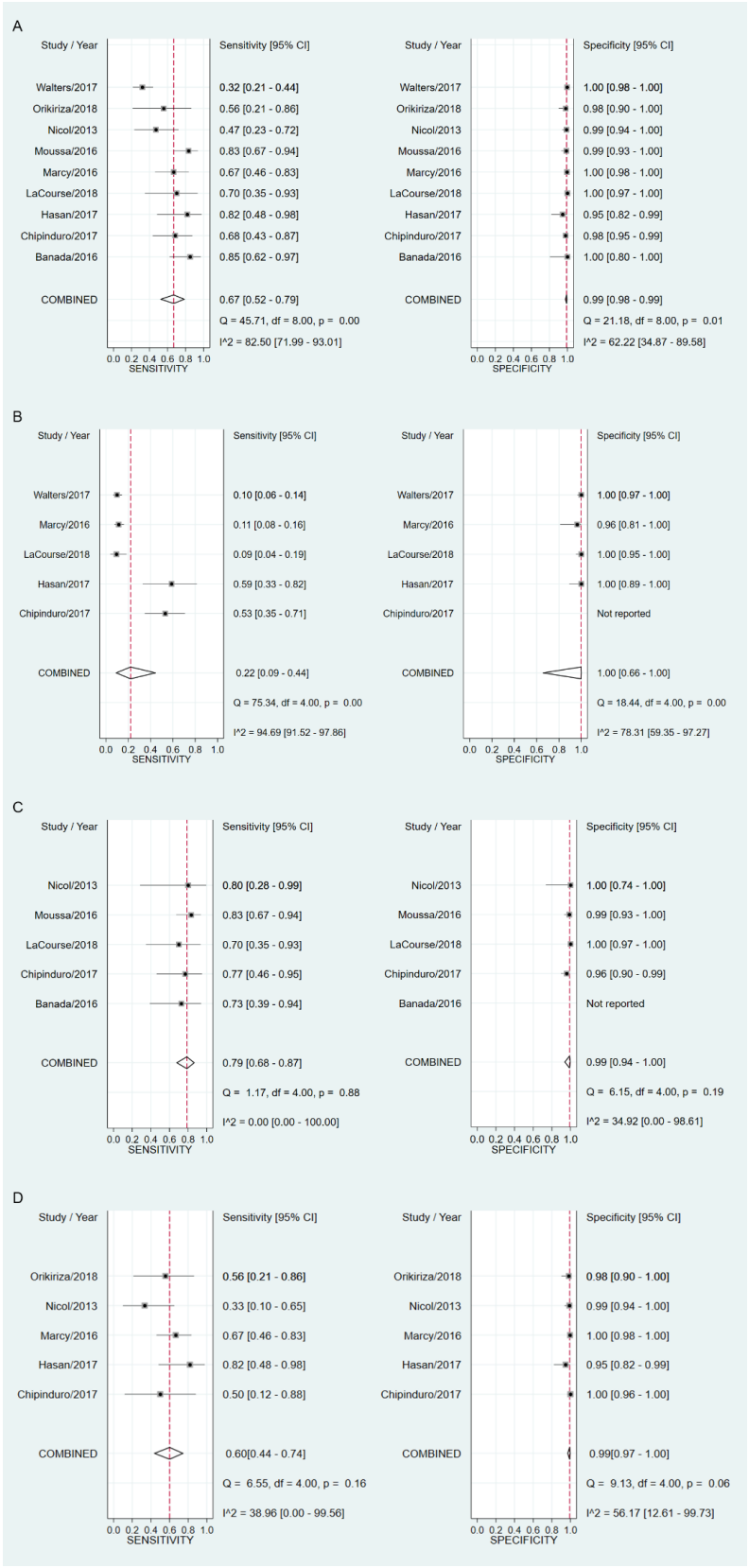

Fig. 3A: Forest plots of stool Xpert's diagnostic performance compared to a microbiological reference standard of culture or Xpert positivity on respiratory samples. Two studies (17, 18) presented results from "intention-to-treat" (ITT) analyses, where any child who produced any sample was included, as well as "per protocol" analyses, where only children who produced all requested samples were included. In these instances, we meta-analysed the ITT results to avoid selection bias.

Fig. 3B: Forest plots of stool Xpert's diagnostic performance compared to a clinical reference standard of "likely/possibly TB" or "unlikely TB".

Fig. 3C: Forest plots of diagnostic performance of stool Xpert in children with HIV compared to a microbiological reference standard.

Fig. 3D: Forest plots of diagnostic performance of stool Xpert in HIV-negative children compared to a microbiological reference standard.

Journal of Clinical Microbiology
JCM
Journal of Clinical Microbiology