Tuberculosis 119 (2019) 101878

Contents lists available at ScienceDirect

Tuberculosis

journal homepage: http://www.elsevier.com/locate/tube

Diagnostic accuracy of molecular detection of Mycobacterium tuberculosis in pediatric stool samples: A systematic review and meta-analysis

Annelies W. Mesman [a], Carly Rodriguez [a], Emily Ager [b], Julia Coit [a], Letizia Trevisi [a], Molly F. Franke [a,*]

[a] Harvard Medical School, Department of Global Health and Social Medicine, Boston, MA, USA
[b] Harvard TH Chan School of Public Health, Department of Social and Behavioral Sciences, Boston, MA, USA

ARTICLE INFO

Keywords:
Children
Alternative sample
Review
Stool test

ABSTRACT

Background: Stool is a promising specimen option to diagnose pediatric tuberculosis (TB), but studies have reported a wide range of test sensitivities. We conducted a meta-analysis to assess the accuracy of Xpert MTB/RIF or 'in-house' molecular tests on stool samples against culture or Xpert MTB/RIF on respiratory samples or clinically-diagnosed unconfirmed TB and aimed to identify factors that contribute to the heterogeneity of reported sensitivity.

Methods: We searched EMBASE and Pubmed databases and conference abstract books for studies reporting molecular stool testing against a clinical or microbiological reference standard among children.

Results: We identified 16 studies that included 2,481 children in stool test analyses. Pooled specificity was 98% [95%CI: 96–99], pooled sensitivity was 57% [95%CI: 40–72] against culture and 3% [95%CI: 2–6] among children with clinically-diagnosed, unconfirmed TB. There was much heterogeneity. Sensitivity was higher among children with a smear-positive sputum test. Rifampin resistance in stool was reported in two studies and detected in 5/14 children (36%).

Conclusion: Our results suggest molecular stool tests have potential as diagnostic rule-in tests, but it is challenging to optimize sensitivity due to between-study variation in methodology and test procedures. Therefore, we recommend future research with rigorous study design and standardized results reporting.

1. Introduction

Tuberculosis (TB) is the global leading infectious cause of death. Children account for 10% of the global disease burden with an estimated one million new pediatric TB cases and 234,000 childhood TB deaths in 2017 [1]. Rapid TB case detection is critical for timely treatment initiation but is challenging in children because they often present with paucibacillary disease and may be unable to expectorate sputum for microbiological confirmation of *Mycobacterium tuberculosis* (*Mtb*). Smear microscopy has little value due to low sensitivity, and while culture has higher sensitivity in children, sensitivity is still suboptimal, at about 30%–40% in children with TB disease. Consequently, TB diagnosis in children often relies on clinical manifestations [2–4]. Since the World Health Organization endorsed Xpert MTB/RIF (Cepheid, Sunnyville CA, USA) for TB diagnosis, this molecular test has been widely used to test pediatric respiratory samples. With a sensitivity of 62% among culture-confirmed pediatric cases, Xpert MTB/RIF performs better than microscopy, but is suboptimal to culture [5]. Due to these limitations there is a need for a non-sputum based test to diagnose TB in children.

For children unable to expectorate sputum, procedures including sputum induction, gastric or nasopharyngeal aspiration or the string test may be conducted to obtain a respiratory specimen for testing [6,7]. Some of these methods are invasive and none are routinely available across resource-limited settings. Because sputum may be swallowed and excreted, TB molecular diagnostic studies have focused on the evaluation of stool as an alternative, non-invasive specimen for pediatric TB diagnosis. These studies have reported a wide range of test sensitivities, reflecting the absence of a standard protocol and heterogeneous patient populations. A recent systematic review of stool Xpert MTB/RIF test accuracy confirmed a high degree of heterogeneity in nine pediatric studies [8]. The potential sources of heterogeneity in TB molecular diagnostic studies of stool have not been adequately explored.

We conducted a systematic review on the detection of *Mtb* in stool samples from children suspected of having pulmonary TB using Xpert

* Corresponding author.
E-mail address: molly_franke@hms.harvard.edu (M.F. Franke).

https://doi.org/10.1016/j.tube.2019.101878
Received 24 August 2019; Received in revised form 18 October 2019; Accepted 19 October 2019
Available online 23 October 2019
1472-9792/

MTB/RIF or other in-house molecular assays. Our aim was to estimate sensitivity and specificity of stool assays in comparison to the gold standard test, solid or liquid culture, and two other commonly-used TB reference standards: Xpert MTB/RIF on respiratory samples and clinically-diagnosed, unconfirmed TB. Additionally, we aimed to investigate sources of heterogeneity in available data (i.e., smear-status, timing of sample collection) and collected Xpert MTB/RIF drug resistance data to estimate sensitivity of rifampin (RIF) resistance in stool.

2. Materials and methods

We followed the PRISMA diagnostic test accuracy (DTA) guidelines [9] in all steps of the process (Supplementary Fig. S1).

2.1. Search strategy

We conducted a Pubmed and EMBASE search on September 12, 2018, using terms for child, tuberculosis and stool (Supplementary Fig. S2). We also searched the 2017 and 2018 International Union Against Tuberculosis and Lung Disease World Conference on Lung Health abstract books and the reference lists of included studies and review articles.

2.2. Eligibility

We included original data studies published after 1990, written in English, Spanish, French, and German, with no restrictions on study design, setting, or comorbidities. Studies eligible for inclusion were those reporting results for molecular detection of *Mtb* from stool (i.e., the index test) in children (0–15 years) with presumed pulmonary TB as compared to microbiological test results in sputum or gastric aspirate (GA).

2.3. Data extraction

After removing duplicate entries, two reviewers (AM, JC) independently selected titles and abstracts for full text review, based on inclusion and exclusion criteria. Full text reviews were conducted by reviewers AM and EA. In case of disagreement, discussion with a third reviewer (MF) was sought to reach consensus. Throughout all review steps, we contacted authors for clarifications regarding data and study design or requested additional information. This occurred when, for example, data were not disaggregated by age or sample type, or full reports of data presented in abstracts were not yet published.

Data from each included article were independently extracted by two of three authors (AM, EA, CR) in Microsoft Excel using a standardized format based on the Review Manager (RevMan version 5.3) DTA format (The Cochrane Collaboration, Copenhagen, Denmark). We extracted data for diagnostic two by two tables, as well as information regarding sample collection, including number of samples collected per child, sample volume tested and timing of collection relative to TB treatment start. We also gathered data on test performance, such as invalid test results and drug resistance results. To summarize available data, we also retrieved characteristics of the study setting and population, such as age, sex and HIV-status.

To obtain estimates for stool test sensitivity and specificity, index test (i.e., molecular diagnostic assay on stool) and reference test (i.e., solid or liquid culture on sputum or GA) results were aggregated on a per-child basis. We made this decision based on clinical practice, in which multiple tests and sample types are often used to diagnose TB in children. We additionally retrieved sputum or GA Xpert MTB/RIF results and, when available, results from smear microscopy and whether a clinical diagnosis was made in the absence of microbiological confirmation. We stratified sensitivity estimates by whether the diagnosis was microbiologically confirmed (e.g., by culture or Xpert MTB/RIF on sputum or GA) or solely based on clinical criteria in the absence of microbiological confirmation (i.e., clinically-diagnosed, unconfirmed cases). We additionally collected Xpert MTB/RIF stool RIF-resistance test results (index test), and culture-based phenotypic RIF-sensitivity results or Genotype line probe assay on respiratory samples (reference test).

2.4. Quality assessment

Two reviewers independently assessed the quality of included studies using the most recent Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool [10]. Discrepancies between reviewers were resolved by consensus. With this tool we evaluated risk of bias within four domains: (1) patient selection, (2) index test, (3) reference standard, and (4) flow and timing. Per guidelines, we evaluated concerns regarding applicability in the first three domains. Bias or applicability concerns were qualified as low, unclear, or high [10]. We did not assess publication bias because there is no validated method to examine this for diagnostic studies [9].

2.5. Accuracy analysis

For accuracy estimates, true positives (tp), true negatives (tn), false positives (fp) and false negatives (fn) were entered in RevMan. We constructed forest plots with sensitivity and specificity estimates and 95% confidence intervals. To visualize the impact of the index assay (i.e., Xpert MTB/RIF versus in-house) and reference standard (i.e., microbiologically confirmed versus clinically-diagnosed unconfirmed TB) on test accuracy, we presented sensitivity and specificity plots stratified by these variables. We excluded invalid or non-interpretable test results from sensitivity and specificity analyses. When an article included a comparison between two stool testing methods, we included the results that yielded highest sensitivity [11,12].

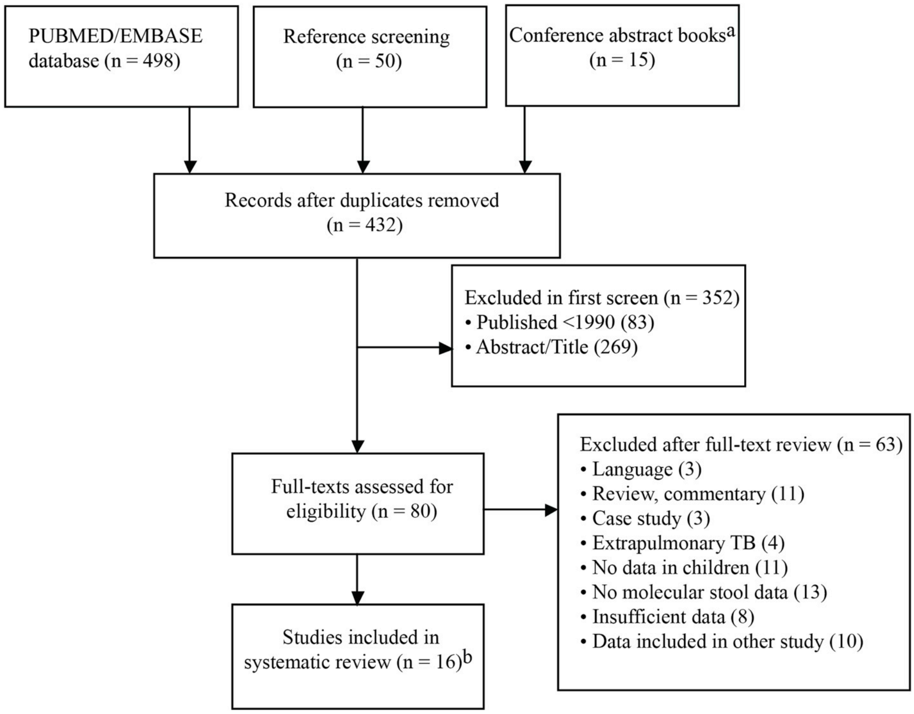
We conducted meta-analyses using a bivariate random effect model to obtain pooled sensitivity and specificity and I^2 estimates as a measure for heterogeneity, all with 95% confidence intervals [13]. These analyses were performed in STATA software (version 15, STATACorp) with the Midas command. Where the default *xtmelogit* function was not applicable, we used the *gllamm* function. To obtain pooled sensitivity estimates for subgroups that were too small for bivariate analyses e.g., smear status, clinically-diagnosed unconfirmed TB, we used a univariate random effect model via the Metan command. For studies with 0% or 100% sensitivity or specificity we applied a continuity correction of 0.5 [14].

3. Results

3.1. Characteristics of included studies

We identified 432 unique studies through database and reference searches, of which, after an abstract and title screen, we assessed 80 full-text articles for eligibility (Fig. 1). We subsequently excluded 63 studies, most of which did not contain stool results based on molecular tests from children, including studies in which data from children could not be disaggregated from adults. Other studies were excluded because they did not contain sufficient data (including abstracts of unpublished papers) or because the data overlapped with those presented in another report. In one report [15], confirmation of TB was only conducted by smear microscopy of respiratory samples. Although we had not explicitly specified the requirement for culture or Xpert MTB/RIF in our eligibility criteria, we did not include this study in our quantitative analysis because of the limited sensitivity of the smear test. The remaining 16 studies included a total of 2,481 children. All of the included studies were written in English.

Table 1 summarizes the characteristics of the included studies. All studies were conducted in low or middle-income countries; eight took place in high-burden TB countries [1]. One multi-country study included participants from both high-burden and non-high-burden countries



[a] 2017 and 2018 World Conference on Lung Health (Int Union against Tuberculosis and Lung Disease)
[b] 17 studies identified and one study excluded based on reference test

Fig. 1. Flowchart of included and excluded studies.

Table 1
Characteristics of the included studies.

Study	Country	TB cases	TB controls	Participant stool analysis (N)	Female (%)	Median age (years)	HIV infection (%)
Banada 2016 [11]	South Africa[a]	Xpert-confirmed	symptomatic	38	55.3	NR[a]	42.1
Chipinduro 2017 [30]	Zimbabwe[a]	micro-confirmed[b]; unconfirmed TB	symptomatic	218	56.4	10.6	51.0
DiNardo 2018 [18]	Eswatini	micro-confirmed; unconfirmed TB	asymptomatic	143	65 controls; cases NR	Est[c] 6.8	32.0
Hasan 2017 [26]	Pakistan[a]	micro-confirmed; unconfirmed TB	symptomatic	49	44.0	6.8	NR
LaCourse 2018 [22]	Kenya[a]	Culture-confirmed; unconfirmed TB	symptomatic	150	45.4	2	100.0
Marcy 2016 [16]	Burkina Faso, Vietnam[a], Cameroon, Cambodia[a]	Culture-confirmed	symptomatic	267	49.0	7.2	100.0
Memon 2018 [28]	India[a]	Culture-confirmed; unconfirmed TB	–	100	59.0	11	NR
Mesman 2019 [20]	Peru	Culture-confirmed; unconfirmed TB	symptomatic	259	48.0	5.1	NR
Moussa 2016 [17]	Egypt	Culture-confirmed	symptomatic	115	39.0	Est >6	0.0
Nicol 2013 [29]	South Africa[a]	Culture-confirmed; unconfirmed TB	symptomatic	115	NR	2.6	14.8
Oberhelman 2010 [19]	Peru	Culture-confirmed; unconfirmed TB	asymptomatic	451	47.0	Est 4	0.0
Orikiriza 2018 [23]	Uganda	micro-confirmed; unconfirmed TB	–	71	45.4	NR	31.2
Sauter 2018 [27],[b]	Cameroon	micro-confirmed; unconfirmed TB	symptomatic	73	51.8	8	100
Walters 2012 [25]	South Africa[a]	Culture-confirmed; unconfirmed TB	symptomatic	14	65.0	1.4	8.6
Walters 2017 [24]	South Africa[a]	micro-confirmed; unconfirmed TB	symptomatic	379	48.5	1.3	13.5
Wolf 2008 [12]	Peru	Culture-confirmed	asymptomatic	39	25.0	5.1	0.0

Abbreviations: [a]NR: not reported; [b]microbiological confirmation (via either Xpert or culture); [c]est: estimated.
[a] High-burden TB countries.
[b] Abstract; data obtained via personal communication.

[16]. Annual TB incidence rates in the countries where the remaining studies were conducted ranged from 13/100,000 (Egypt) to 308/100,000 (Eswatini) [17,18].

In most studies (13/16; 81%), children with TB symptoms were consecutively enrolled and subsequently categorized as TB cases with microbiologically-confirmed or clinically-diagnosed unconfirmed TB, or as controls in the absence of a diagnosis. The remaining three studies had a case-control design in which controls were healthy children [12,18,19]. The median age of participants ranged from 1.3 to 11 years. The proportion of children who were HIV-positive was reported in 13 studies and ranged from 0% to 100% (interquartile range: 8.6%–42%).

Both index and reference test characteristics were heterogeneous. With respect to the index test, most studies (12/16, 75%) used Xpert MTB/RIF for stool testing (Table 2); four studies used other molecular tests (referred to as: 'in-house') and in all four *Mtb* detection was based on IS6110 amplification. These studies all used different DNA extraction methods: Chelex/fast [12], MP Fast DNA kit for soil (MP Biochemicals, Solon, OH) [18], Akonni TruTip [20] and Triton X-100 [19,21]. Stool sample input volume ranged from 50 to 5000 mg, and was not recorded in four studies [22–25]. Three studies manually homogenized the sample before further processing [22,25,26]. Also other stool processing procedures we did not include in our analysis, such as decontamination, or concentration steps, were highly variable for both in-house and Xpert MTB/RIF studies. Microbiological confirmation of TB disease usually included culture analysis (15/16, 94%), and most studies (9/16, 56%) additionally had Xpert MTB/RIF results available, performed for either diagnostic or research purposes. One study relied solely on confirmation by Xpert MTB/RIF (Table 2).

Only 7/16 (44%) study protocols required that stool samples be collected prior to initiation of anti-TB treatment [12,16,17,20,22,27,28]. Of those studies without this requirement, one allowed enrollment of children up to four weeks post treatment initiation. The actual time of stool sample collection relative to treatment initiation sometimes differed from the intended collection time and was not always reported (Table 2).

3.2. Quality assessment of included studies

The assessment of methodological quality of included studies is summarized in Fig. 2. Summaries per-study are shown in Supplementary Fig. 3. We considered five studies to have high risk for bias in the patient selection domain due either to a case-control study design with healthy controls [12,18,19] or lack of a control group [23,26]. Within the index test domain, we did not consider the lack of blind testing as a source of bias because automated molecular test results and differences in turn-around-times between the index test and culture reference test are unlikely to affect interpretation of results.

One study had not specified the positivity threshold of the index test before study initiation, which we considered an unclear risk of bias [20]. Within the reference test domain, one study scored as having a high risk

Table 2
Test characteristics.

Study	Index test	Actual collection stool sample relative to start ATT[a]	Stool samples collected (N)	Stool volume (mg [range])	Stool processing method	Microbiological confirmation test	Sample type reference test
Banada 2016 [11]	Xpert	<5 days	1	600; 1200	Bead vortexing	Xpert	IS[b]/GA[c]
Chipinduro 2017 [30]	Xpert	NR[d]	1	150	Vortexing	CU[e] (LJ[f]); Xpert	IS
DiNardo 2018 [18]	In-house	<14 days	1	50	Bead beating [MP Fast DNA Soil Kit]	CU; Xpert	IS/ES[g]/GA
Hasan 2017 [26]	Xpert	before ATT	1	150	Vortexing	CU (MGIT[h]); Xpert	ES/GA
LaCourse 2018 [22]	Xpert	NR	1	NR	Manual homogenization	CU (MGIT, LJ); Xpert	IS/ES/GA
Marcy 2016 [16]	Xpert	before ATT	1	500	Vortexing	CU (LJ, MGIT); Xpert	ES/GA
Memon 2018 [28]	Xpert	NR	1	200	Manual homogenization	CU (MGIT); Xpert	IS/GA
Mesman 2019 [20]	In-house	<14 days	2	500	Magnetically-induced vortexing [TruTip extraction]	CU (MGIT)	IS/ES/GA
Moussa 2016 [17]	Xpert	NR	2	2 cm^3	Vortexing	CU (LJ)	ES/IS
Nicol 2013 [29]	Xpert	NR	1	150	Vortexing	CU (MGIT); Xpert	IS
Oberhelman 2010 [19]	In-house	before ATT	2	100	Homogenization (method NR)	CU (MODS[i], LJ)	GA
Orikiriza 2018 [23]	Xpert	<7 days	1	NR [1000–5000]	Vortexing	CU (LJ, MGIT); Xpert	ES/IS
Sauter 2018 [27]	Xpert	before ATT	1	Est 1000-3000	NA[j]	CU (LJ, MGIT); Xpert	Spm/GA
Walters 2012 [25]	Xpert	<7 days	1	NR	Manual homogenization	CU (MGIT)	GA
Walters 2017 [24]	Xpert	<7 days	1	NR [<5000]	Vortexing	CU (MGIT); Xpert	ES/IS/GA
Wolf 2008 [12]	In-house	before ATT	2	200	FastPrep mechanical homogenization	CU (MODS)	GA

[a] ATT: anti-TB treatment.
[b] IS: induced sputum.
[c] GA: gastric aspirate.
[d] NR: not reported.
[e] CU: culture.
[f] LJ: Lowenstein-Jensen (solid).
[g] ES: expectorated sputum.
[h] MGIT: Mycobacteria growth indicator tube (liquid).
[i] MODS: Microscopic observation drug susceptibility.
[j] NA: Data Not Available.

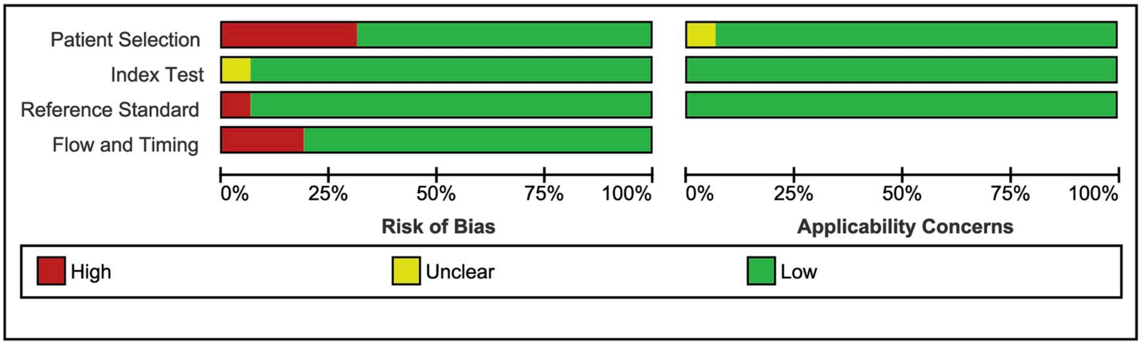


Fig. 2. QUADAS-2 summarized risk of bias and applicability concerns for the included studies.

of bias because it only compared the stool results of all children to Xpert MTB/RIF on respiratory samples, and did not conduct the best available reference standard (i.e. culture) [11]. We had few applicability concerns, but considered applicability unclear in one study [22], which was nested within an HIV project and included all hospitalized children, rather than basing inclusion on TB symptoms.

One of the questions the flow and timing domain assesses is the interval between the index and reference test. Instead of the test interval, we assessed the interval between the index test and treatment start, which could impact test outcomes. We considered studies with over 72 h between index test and treatment start at high risk of bias, except when the authors demonstrated that late collection did not affect the outcomes [11,20,24]. Of the three studies that accepted stool samples that were collected more than 72 h post-treatment start, one reported that late sample collection negatively affected *Mtb* detection in stool [18], whereas the impact was not examined in the two other studies [23,26].

3.3. Diagnostic accuracy

Fourteen studies estimated stool test sensitivity against culture; a total of 212/1910 (11%) children from these studies had culture-confirmed TB. Stool test sensitivity varied across these studies from 12% to 100% with Xpert MTB/RIF stool tests and 19%–67% with other in-house tests, all with wide confidence intervals (Fig. 3A). Pooled sensitivity was 57% [95%CI: 40–72] (Table 3). The specificity for stool testing was high, with a pooled estimate of 98% [95%CI: 96–99]. We did not conduct meta-regression to compare Xpert MTB/RIF with in-house stool assays because of the small number of studies with in-house tests and between-study differences in both index test methodology and participant selection (Table 1, Table 2); the substantial heterogeneity is reflected in the high I^2 statistics (Table 3). Separate pooled sensitivity for studies using stool Xpert MTB/RIF was 63% [95%CI: 42–80], whereas the pooled estimate for in-house tests was 42% [95%CI: 24–61] (Supplementary Fig. 4); heterogeneity remained high for both groups.

We additionally plotted sensitivity and specificity against other reference standards: Xpert MTB/RIF (Fig. 3B), any microbiological confirmation (either Xpert MTB/RIF/culture; Fig. 3C) and clinically-diagnosed unconfirmed TB (Fig. 3D). We hypothesized that the Xpert MTB/RIF reference test would lead to higher index test sensitivity, because it is less sensitive than culture. However, the pooled sensitivity estimates for the alternative reference standards of Xpert MTB/RIF or any microbiological confirmation were similar to the pooled sensitivity estimate with culture as the reference: 60% [95%: 36–80] and 53% [95%CI: 39–67], respectively (Table 3). Notably, these separate analyses do not include the same studies. For example, a large study by Walters et al. reported a sensitivity of 35% [95%CI: 24–48], which was only included in the combined microbiological confirmed analysis [24].

Among children with clinically-diagnosed unconfirmed TB, sensitivity was very low, ranging from 0 to 17% across studies, with a pooled estimate of 3% [95%CI: 2–6] (Fig. 3D, Table 3). Specificity against this reference standard however was very high, with estimates of 99%–100% for Xpert MTB/RIF stool tests [24,25,27–30] and 96%–98% for in-house stool tests [18–20].

We did not identify any study-level factors, such as median age, percentage of HIV-infected participants or number of stool samples analyzed, associated with stool test sensitivity (Supplementary Table 1); however, we observed that studies in which stool samples were collected after treatment start all reported sensitivities at or below the median.

In a subgroup analysis stratified by smear-status among children with confirmed disease (Fig. 4), sensitivity was much higher among smear-positive confirmed cases. In half of the studies, stool test sensitivity was 100% among smear-positive children (pooled 80%, 95%CI: 60–92, Table 3). In two studies sensitivity among smear-positive cases was below 50%; this included the study of Memon et al. with the lowest stool test sensitivity of 13% [95% CI: 3–32] that enrolled children until four weeks post treatment initiation [26]. Compared to the smear-positive cases, sensitivity among smear-negative children was much lower in all studies (pooled 33%, 95%CI: 20–49, Table 3).

3.4. Additional xpert MTB/RIF test performance parameters

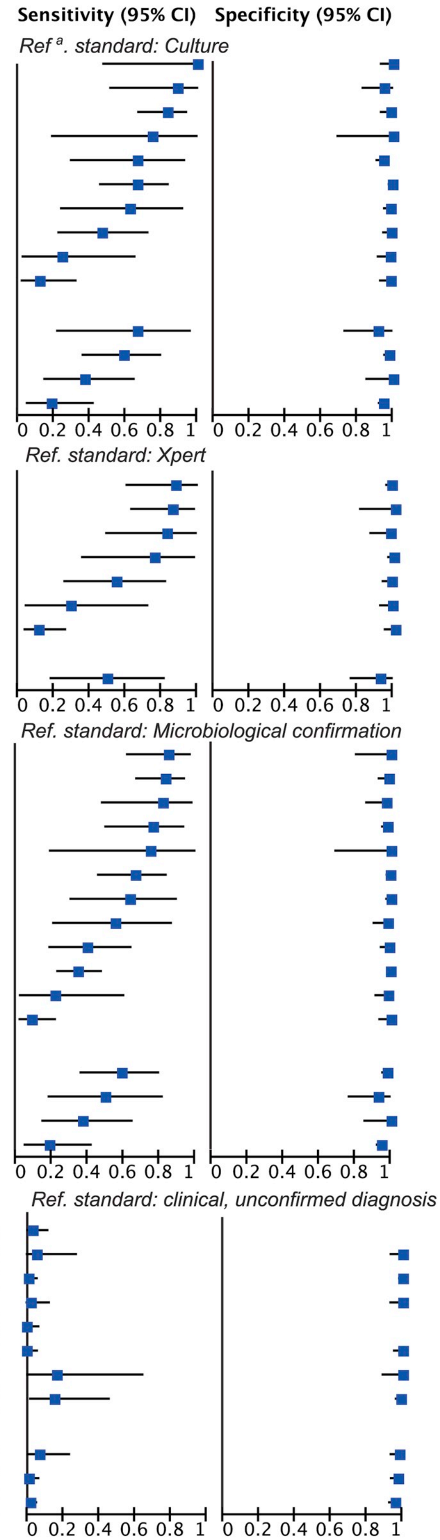
For stool Xpert MTB/RIF studies we assessed two additional performance parameters: the percentage of invalid or erroneous test results and RIF resistance testing. The percentage of invalid outcomes was generally low. In 6/12 studies (50%), all children had valid stool result (Supplementary Table 2). The number of invalid results from one study was unknown because the authors combined invalid results with withdrawals [16]. In the six studies that reported invalid test results, the percentage ranged from 1.3% to 7.9%. Notably, 4/5 studies without invalid results used low stool input volumes up to 150 mg [17,27,29,30]; (the fifth study did not record volume [25]), whereas the input volume ranged from 200 to 5000 mg in the studies that reported invalid outcomes.

Drug resistance was reported in 9/12 (75%) of stool Xpert MTB/RIF studies (Supplementary Table 2). Most of these studies (7/9, 78%) detected no drug resistance in either respiratory or stool samples (100% concordance) [11,25–30]. Only 14 children across two studies were diagnosed with RIF-resistant TB after phenotypic drug susceptibility testing (DST) or confirmation via Genotype line probe assay. Xpert MTB/RIF stool testing resistance sensitivity was 100% and 25% in these two studies respectively and confirmed RIF resistance in 5/14 children (36%) [17,24]. Studies did not report RIF resistance results detected by Xpert MTB/RIF on respiratory samples.

4. Discussion

In this review, we aimed to estimate the accuracy of molecular *Mtb* testing in pediatric stool samples following a careful evaluation of the literature and found consistently high specificity and a pooled sensitivity estimate of 57% against culture.

	Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
A	<u>Xpert</u>						
	Orikiriza 2018	5	0	0	52	1.00 [0.48, 1.00]	1.00 [0.93, 1.00]
	Hasan 2017	8	2	1	38	0.89 [0.52, 1.00]	0.95 [0.83, 0.99]
	Moussa 2016	30	1	6	78	0.83 [0.67, 0.94]	0.99 [0.93, 1.00]
	Walters 2012	3	0	1	10	0.75 [0.19, 0.99]	1.00 [0.69, 1.00]
	Chipinduro 2017	6	11	3	198	0.67 [0.30, 0.93]	0.95 [0.91, 0.97]
	Marcy 2016	18	1	9	223	0.67 [0.46, 0.83]	1.00 [0.98, 1.00]
	LaCourse 2018	5	2	3	137	0.63 [0.24, 0.91]	0.99 [0.95, 1.00]
	Nicol 2013	8	1	9	97	0.47 [0.23, 0.72]	0.99 [0.94, 1.00]
	Sauter 2018	2	1	6	62	0.25 [0.03, 0.65]	0.98 [0.91, 1.00]
	Memon 2018	3	1	21	73	0.13 [0.03, 0.32]	0.99 [0.93, 1.00]
	<u>In-house</u>						
	DiNardo 2018	4	2	2	22	0.67 [0.22, 0.96]	0.92 [0.73, 0.99]
	Mesman 2019	13	5	9	228	0.59 [0.36, 0.79]	0.98 [0.95, 0.99]
	Wolf 2008	6	0	10	23	0.38 [0.15, 0.65]	1.00 [0.85, 1.00]
	Oberhelman 2010	4	23	17	407	0.19 [0.05, 0.42]	0.95 [0.92, 0.97]
B	<u>Xpert</u>						
	Chipinduro 2017	13	4	2	199	0.87 [0.60, 0.98]	0.98 [0.95, 0.99]
	Banada 2016 (0.6)	17	0	3	17	0.85 [0.62, 0.97]	1.00 [0.80, 1.00]
	Hasan 2017	9	1	2	37	0.82 [0.48, 0.98]	0.97 [0.86, 1.00]
	LaCourse 2018	6	1	2	139	0.75 [0.35, 0.97]	0.99 [0.96, 1.00]
	Nicol 2013	7	2	6	99	0.54 [0.25, 0.81]	0.98 [0.93, 1.00]
	Sauter 2018	2	1	5	63	0.29 [0.04, 0.71]	0.98 [0.92, 1.00]
	Memon 2018	4	0	33	61	0.11 [0.03, 0.25]	1.00 [0.94, 1.00]
	<u>In-house</u>						
	DiNardo 2018	5	2	5	26	0.50 [0.19, 0.81]	0.93 [0.76, 0.99]
C	<u>Xpert</u>						
	Banada 2016	17	0	3	17	0.85 [0.62, 0.97]	1.00 [0.80, 1.00]
	Moussa 2016	30	1	6	78	0.83 [0.67, 0.94]	0.99 [0.93, 1.00]
	Hasan 2017	9	1	2	37	0.82 [0.48, 0.98]	0.97 [0.86, 1.00]
	Chipinduro 2017	13	4	4	197	0.76 [0.50, 0.93]	0.98 [0.95, 0.99]
	Walters 2012	3	0	1	10	0.75 [0.19, 0.99]	1.00 [0.69, 1.00]
	Marcy 2016	18	1	9	223	0.67 [0.46, 0.83]	1.00 [0.98, 1.00]
	LaCourse 2018	7	0	4	137	0.64 [0.31, 0.89]	1.00 [0.97, 1.00]
	Orikiriza 2018	5	1	4	54	0.56 [0.21, 0.86]	0.98 [0.90, 1.00]
	Nicol 2013	8	1	12	94	0.40 [0.19, 0.64]	0.99 [0.94, 1.00]
	Walters 2017	23	1	43	284	0.35 [0.24, 0.48]	1.00 [0.98, 1.00]
	Sauter 2018	2	1	7	61	0.22 [0.03, 0.60]	0.98 [0.91, 1.00]
	Memon 2018	4	0	39	55	0.09 [0.03, 0.22]	1.00 [0.94, 1.00]
	<u>In-house</u>						
	Mesman 2019	13	5	9	228	0.59 [0.36, 0.79]	0.98 [0.95, 0.99]
	DiNardo 2018	5	2	5	26	0.50 [0.19, 0.81]	0.93 [0.76, 0.99]
	Wolf 2008	6	0	10	23	0.38 [0.15, 0.65]	1.00 [0.85, 1.00]
	Oberhelman 2010	4	23	17	407	0.19 [0.05, 0.42]	0.95 [0.92, 0.97]
D	<u>Xpert</u>						
	Orikiriza 2018	2	0	58	0	0.03 [0.00, 0.12]	Not estimable
	Sauter 2018	1	0	17	53	0.06 [0.00, 0.27]	1.00 [0.93, 1.00]
	Walters 2017	1	0	96	181	0.01 [0.00, 0.06]	1.00 [0.98, 1.00]
	Nicol 2013	1	0	42	52	0.02 [0.00, 0.12]	1.00 [0.93, 1.00]
	Memon 2018	0	0	55	0	0.00 [0.00, 0.06]	Not estimable
	LaCourse 2018	0	0	63	74	0.00 [0.00, 0.06]	1.00 [0.95, 1.00]
	Hasan 2017	1	0	5	31	0.17 [0.00, 0.64]	1.00 [0.89, 1.00]
	Chipinduro 2017	2	2	11	184	0.15 [0.02, 0.45]	0.99 [0.96, 1.00]
	<u>In-house</u>						
	DiNardo 2018	2	2	26	103	0.07 [0.01, 0.24]	0.98 [0.93, 1.00]
	Mesman 2019	1	4	83	149	0.01 [0.00, 0.06]	0.97 [0.93, 0.99]
	Oberhelman 2010	4	10	188	227	0.02 [0.01, 0.05]	0.96 [0.92, 0.98]



a: Reference

Fig. 3. Forest plots of stool test accuracy against culture (A), Xpert (B), microbiological confirmation (C) or clinically-diagnosed unconfirmed TB (D) reference standard. Plots display Xpert stool test studies at the top and in-house stool test studies at the bottom of the plots.

Stool test sensitivity varied immensely across studies, from 13% to 100% as compared to culture, due to major heterogeneity in study design, sample processing methods and test procedures. Therefore, all pooled estimates should be interpreted with great caution. Within studies however, sensitivity was consistently highest among children with smear-positive TB. In fact, sensitivity among these children was 100% in the majority of studies [17,20,23,29,30], as well as in the study excluded from the meta-analysis because the authors only compared stool to smear microcopy [15]. These outcomes align with previous reports that demonstrated higher sensitivity of Xpert MTB/RIF respiratory [5,31] and non-respiratory samples [32] among adults and children with a positive smear status. The difference between our pooled

Table 3
Pooled sensitivity and specificity estimates of stool tests (Xpert MTB/RIF and In-house) against different reference tests obtained by meta-analyses.

Reference test			Sensitivity		Specificity	
	Studies (n)	Model	Pooled (95%CI)	I^2 (95%CI)	Pooled (95%CI)	I^2 (95%CI)
Culture	14	bivariate	0.57 (0.40–0.72)	77.1 (65.5–88.8)	0.98 (0.96–0.99)	76.6 (64.6–88.6)
Xpert	8	bivariate	0.60 (0.36–0.80)	86.3 (78.0–94.5)	0.98 (0.97–0.99)	28.6 (0.0–85.9)
Micro-confirmed[a]	16	bivariate	0.53 (0.39–0.67)	83.1 (75.7–90.5)	0.99 (0.98–0.99)	83.1 (75.7–90.5)
Clinically-diagnosed Unconfirmed TB	11	univariate	0.03 (0.02–0.06)	29.8 (p = 0.16)	0.98 (0.96–0.99)	35.2 (p = 0.11)
Subgroup analysis						
Micro-confirmed SM+	10	univariate	0.80 (0.60–0.92)	50.0 (p = 0.04)	–	–
Micro-confirmed SM-	10	univariate	0.33 (0.20–0.49)	67.4 (p = 0.001)	–	–

[a] Microbiologically-confirmed via culture or Xpert.

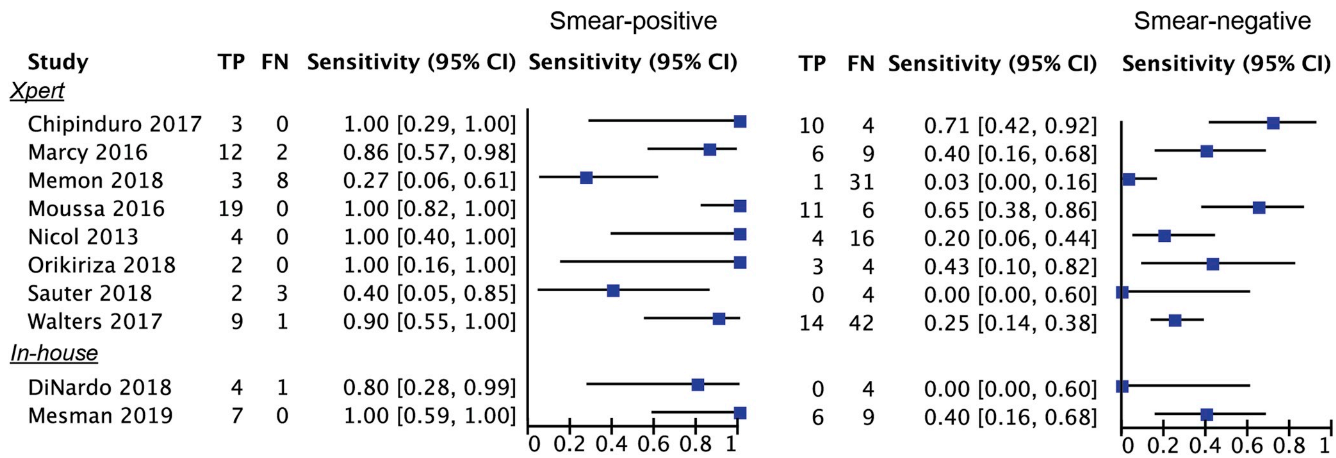


Fig. 4. Forest plots with sensitivity of molecular stool tests among microbiologically-confirmed children stratified by smear status.

sensitivity estimates of 80% among smear-positive and 33% among smear-negative children is more striking than HIV-infected versus uninfected (80% and 60% respectively) [8]. This indicates that the frequency of children with a smear positive respiratory sample, often older children with adult-like disease, may affect the sensitivity of stool assays and highlights the importance of stratifying sensitivity estimates by smear status.

We compared stool test accuracy against three reference standards. The molecular stool test had a limited value for children with clinically-diagnosed unconfirmed TB (pooled 3%), for whom a novel diagnostic TB tool is mostly needed. This is comparable to the 2% sensitivity of Xpert MTB/RIF on sputum among this group [5]. We hypothesized that using a reference of Xpert MTB/RIF on respiratory samples could overestimate the sensitivity of stool testing relative to culture of respiratory specimens because it is less sensitive than culture, but found similar results. This could have occurred for several reasons. First, if children had received any anti-TB treatment (or fluoroquinolones, reported as bias concern by Orikiriza [23]), this could have resulted in a negative culture result and positive Xpert MTB/RIF result because the latter test may detect nonviable *Mtb* [23,26]. Secondly, children with positive respiratory Xpert MTB/RIF results, may have had missing culture results [18,23,27]. Thirdly, some children had respiratory sample Xpert MTB/RIF-positive, but culture-negative results [28,29], which leads to a decreased sensitivity in the overall microbiological confirmed analysis. While we did not find evidence to support our hypothesis and pooled stool sensitivity against Xpert MTB/RIF [60%] was similar to that against culture [57%], our pooled estimates must be interpreted with caution given the heterogeneous nature of the studies included.

We also observed variability in the timing of stool sample collection relative to treatment initiation and found that time of collection is a possible risk for bias in accuracy estimates. DiNardo et al. showed that late collection negatively impacted detection [18]. In the study with the lowest sensitivity against culture [13%] including among children who tested smear-positive [27%], stool samples were collected up to four weeks post treatment start, which was the possible cause for low test sensitivity [26]. Unfortunately, lack of individual patient data hampered confirmation of the association.

In contrast to the between-study variation in sensitivity, specificity of stool testing was consistently very high and therefore stool testing could be used as rule-in TB test (i.e. a positive stool test result would lead to a TB diagnosis, but a negative result cannot rule out TB). Of note, high specificity for a TB diagnostic test is crucial given the long and potentially toxic nature of TB treatment. In this regard, the specificity of molecular detection of *Mtb* in stool is promising, and compares favorably with the 82% specificity of the urine-based TB-LAM lateral flow test [33]. Notably, in multiple studies that recorded treatment responses of all participants, 'false positive' stool results (against respiratory culture or Xpert MTB/RIF reference standard) belonged to children who had received an unconfirmed TB diagnosis and clinically improved after receiving anti-TB treatment [27–29]. Assuming these children truly did have TB, including clinically-diagnosed, unconfirmed TB cases in the control group could result in underestimates of test specificity.

We conducted the first molecular stool test review that included assays other than Xpert MTB/RIF. While pooled sensitivity in the Xpert MTB/RIF stool test group [63%] was higher than in the in-house test group [42%], heterogeneity was high for both groups. The small number of studies with in-house molecular tests, the differences in sample processing, extraction methodology and study design (3/4 in-house studies were case-control), made it impossible to compare outcomes across the two assay types. Individual in-house tests have demonstrated potential to rival Xpert MTB/RIF. For example, one in-house tests reported a limit of detection of 96 bacilli/50 mg stool [18] (to compare: sputum Xpert MTB/RIF's limit of detection is 131 bacilli/ml) and successfully detected *Mtb* in 17% of children with clinically-diagnosed unconfirmed TB.

The major limitation of this study reflects that of the existing body of evidence related to the molecular detection of *Mtb* in stool: a limited number of studies coupled with heterogeneous patient populations, testing procedures, and protocols; these factors limited our ability to examine the key factors that drive assay sensitivity. We were unable to account for all variation in stool sample collection and processing methods that could impact sensitivity and specificity, or could influence the percentage of invalid test outcomes due to clogging of Xpert MTB/RIF cartridges [34]. For example, our results suggest a possible association between stool sample volume and invalid test results which is supported by the reduction of invalid results after decreasing the sample volume from 1200 to 600 mg [11]. However, in many studies volume varied among participants and was not recorded, which prevented a participant-level sub-analysis. Secondly, we conducted analyses by-child and did not stratify for the number of index or reference samples or the type(s) of reference sample, because these individual patient data were in not available in multiple studies.

Based on our analyses we developed recommendations for future studies of TB diagnostic tests in children (Table 4). A major challenge in determining the accuracy of stool testing is the absence of standardized test procedures; studies that compared multiple procedures in parallel, such as sample volume [11], processing [34], or extraction methods [12, 15], reported different outcomes, indicating the relevance of each of these factors. An optimal procedure should be identified through both literature review and experimental studies. Secondly, the lack of a reliable gold standard for pediatric TB diagnoses and consequently small sample size of confirmed TB cases emphasizes the relevance of a rigorous design and data collection for pediatric diagnostic TB studies. Lastly, gaps need to be filled before implementation, including drug resistance testing on stool samples. We identified only two studies that detected or reported any drug resistance via molecular stool testing, with only 25% RIF resistance sensitivity in the largest study [24]. Because another study by the same group (not included in our review) used different sample processing methods and did not confirm drug resistance in three children diagnosed with multi-drug resistant TB [35], this testing should ideally be included in test optimization and extended beyond RIF resistance. Also, the position of stool tests within the diagnostic algorithm should be investigated. Our accuracy estimates (high specificity, lower sensitivity) suggest a role as initial rule-in screening test, without replacing confirmation on respiratory samples. However, the added value of stool tests to other non-sputum based tests remains unknown. For example, stool Xpert MTB/RIF confirmed TB disease in a different patient group than did the urine TB-LAM flow test [22], for which the recommended use is restricted to patients living with HIV [33]. Moreover, Xpert MTB/RIF studies should confirm results with the Xpert MTB/RIF Ultra test, which has higher sensitivity and lower specificity compared to Xpert MTB/RIF [36] and will replace the current test in the near future.

In conclusion, our results indicate that stool is a promising sample type for a rule-in TB test in children. Standardization of testing procedures and rigorous study design will be important to overcome or understand heterogeneity in sensitivity outcomes and develop a valuable molecular test.

Author contributions statement

AM and MF conceptualized and designed the study; AM, EA, JC, CR selected articles and extracted data; AM, LT performed the statistical analyses; AM wrote the first draft of the manuscript; All authors contributed to manuscript revision, read and approved the submitted version.

Funding

This work was supported by the National Institutes of Health (NIH) under the Center of Excellence in Translational Research (CETR) grant U19 AI109755.

Table 4
Recommendations (see text for clarifications).

Recommendations for pediatric stool studies	
Work towards standardized procedures	• Define optimal sample processing, extraction and amplification procedures to increase sensitivity and limit the number of invalid outcomes via literature review and experimental studies

Rigorous study design	• Collect stool samples pre-treatment
	• Record and report factors that facilitate data interpretation and comparison across studies, such as index test sample volume, reference standard (test and sample type), treatment responses of clinically-diagnosed unconfirmed cases
	• Results should be stratified by the following characteristics:
	  • smear status
	  • clinically-diagnosed, unconfirmed TB
	  • reference sample type (gastric aspirate, spontaneous sputum, induced sputum)
	  • number of samples tested per child
	  • reference test (Xpert MTB/RIF versus culture)
	  • stool procedures (e.g., volume, processing/ extraction details)

Filling gaps towards implementation	• Investigate drug resistance detection in stool • Investigate position of stool test in algorithm relative to other screening/diagnostic tools • Focus Xpert MTB/RIF studies on Xpert MTB/RIF Ultra

Declaration of competing interest

MFF was a paid consultant on a National Institutes of Health Small Business Innovation Research program award granted to Akonni Biosystems Inc (SBIR #HHSN272201700063C).

Acknowledgements

We are grateful to all authors who shared data and information with us, and to Javier Zamora (Clinical Biostatistics Unit, Ramón y Cajal Hospital) who supported us in conducting the univariate meta-analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tube.2019.101878.

References

[1] World Health Organization. Global tuberculosis report. 2018.
[2] Starke JR. Pediatric tuberculosis: time for a new approach. Tuberculosis 2003;83:208–12.
[3] Newton S, Brent A, Anderson S, Whittaker E, Kampmann B. Paediatric tuberculosis. Lancet Infect Dis 2008;8:498–510. PMCID: PMC2804291.
[4] DiNardo AR, Detjen A, Ustero P, Ngo K, Bacha J, Mandalakas AM. Culture is an imperfect and heterogeneous reference standard in pediatric tuberculosis. Tuberculosis 2016;101:S105–8.
[5] Detjen AK, DiNardo AR, Leyden J, et al. Xpert MTB/RIF assay for the diagnosis of pulmonary tuberculosis in children: a systematic review and meta-analysis. Lancet Respir Med 2015;3:451–61. PMCID: PMC4756280.
[6] Nicol MP, Zar HJ. New specimens and laboratory diagnostics for childhood pulmonary TB : progress and prospects. Paediatr Respir Rev 2011;12:16–21.
[7] Tafur KT, Coit J, Leon SR, et al. Feasibility of the string test for tuberculosis diagnosis in children between 4 and 14 years old. BMC Infect Dis 2018:1–8.
[8] MacLean E, Sulis G, Denkinger CM, Johnston JC, Pai M, Khan FA. Diagnostic accuracy of stool Xpert MTB/RIF for the detection of pulmonary tuberculosis in children: a systematic review and meta-analysis. J Clin Microbiol 2019;57:e02057.18. PMCID: PMC6535592.
[9] McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, PRISMA DTA Group. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies. Jama 2018;319:388.
[10] Whiting P, Rutjes A, Westwood M, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36.
[11] Banada PP, Naidoo U, Deshpande S, et al. A novel sample processing method for rapid detection of tuberculosis in the stool of pediatric patients using the Xpert MTB/RIF assay. PLoS One 2016;11:1–13.

[12] Wolf H, Mendez M, Gilman RH, et al. Diagnosis of pediatric pulmonary tuberculosis by stool PCR. Am J Trop Med Hyg 2008;79:893–8.
[13] Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005;58:982–90.
[14] Ma X, Nie L, Cole SR, Chu H. Statistical methods for multivariate meta-analysis of diagnostic tests: an overview and tutorial. Stat Methods Med Res 2016;25: 1596–619.
[15] Haile Welday S, Nyerere Kimang A, Muthoni Kabera B, et al. Stool as appropriate sample for the diagnosis of Mycobacterium tuberculosis by gene xpert test. Open J Respir Dis 2013;4:83–9.
[16] Marcy O, Ung V, Goyet S, et al. Performance of xpert MTB/RIF and alternative specimen collection methods for the diagnosis of tuberculosis in HIV-infected children. Clin Infect Dis 2016;62:1161–8.
[17] Moussa HS, Bayoumi FS, Mohamed AMA. Gene xpert for direct detection of Mycobacterium tuberculosis in stool specimens from children with presumptive pulmonary tuberculosis. Ann Clin Lab Sci 2016;46:198–203.
[18] DiNardo AR, Kay AW, Maphalala G, et al. Diagnostic and treatment monitoring potential of a stool-based quantitative polymerase chain reaction assay for pulmonary tuberculosis. Am J Trop Med Hyg 2018;99:310–6.
[19] Oberhelman RA, Soto-Castellares G, Gilman RH, et al. Diagnostic approaches for paediatric tuberculosis by use of different specimen types, culture methods, and PCR: a prospective case-control study. Lancet Infect Dis 2010;10:612–20.
[20] Mesman AW, Soto M, Coit J, et al. Detection of Mycobacterium tuberculosis in pediatric stool samples using TruTip technology. BMC Infect Dis 2019;19:563. PMCID: PMC6598370.
[21] Dalovisio JR, Montenegro-james S, Kemmerly SA, et al. Comparison of the amplified Mycobacterium tuberculosis (MTB) direct test , amplicor MTB PCR , and IS6110-PCR for detection of MTB in respiratory specimens. 1996. p. 1099–106.
[22] Lacourse SM, Pavlinac PB, Cranmer LM, et al. Stool Xpert MTB/RIF and urine lipoarabinomannan for the diagnosis of tuberculosis in hospitalized HIV-infected children. AIDS 2018;32:69–78. PMCID: PMC5812271.
[23] Orikiriza P, Nansumba M, Nyehangane D, et al. 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 2018;37:1465–73.
[24] Walters E, van der Zalm MM, Palmer M, et al. Xpert MTB/RIF on stool is useful for the rapid diagnosis of tuberculosis in young children with severe pulmonary disease. Pediatr Infect Dis J 2017;36:837–43. PMCID: PMC5558052.
[25] Walters E, Gie RP, Hesseling AC, Friedrich SO, Diacon AH, Gie RP. Rapid diagnosis of pediatric intrathoracic tuberculosis from stool samples using the xpert MTB/RIF assay: a pilot study. Pediatr Infect Dis J 2012;31:1316.
[26] Memon SS, Sinha S, Sharma S, Kabra S, Lodha R, Soneja M. Diagnostic accuracy of xpert MTB/RIF assay in stool samples in intrathoracic childhood tuberculosis. J Tuberc Ther 2018;3.
[27] Hasan Z, Shakoor S, Arif F, et al. 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 2017;10:473.
[28] Sauter F, Laah S, Zemsi A, Ngouanom A, Sander M. Potential of the urine lateral flow lipoarabinomannan assay to improve the diagnosis of tuberculosis in symptomatic HIV-positive children in Cameroon. Abstr B 49th World Conf Lung Heal Union 2018;22 Suppl2:S184.
[29] Nicol MP, Spiers K, Workman L, et al. Xpert MTB/RIF testing of stool samples for the diagnosis of pulmonary tuberculosis in children. Clin Infect Dis 2013;57: e18–21.
[30] Chipinduro M, Mateveke K, Makamure B, Ferrand RA, Gomo E. Stool Xpert MTB/RIF test for the diagnosis of childhood pulmonary tuberculosis at primary clinics in Zimbabwe. Int J Tuberc Lung Dis 2017;21:161–6. PMCID: PMC5234433.
[31] Chang K, Lu W, Wang J, et al. Rapid and effective diagnosis of tuberculosis and rifampicin resistance with Xpert MTB/RIF assay: a meta-analysis. J Infect 2012;64: 580–8.
[32] Maynard-smith L, Larke N, Peters JA, Lawn SD. Diagnostic accuracy of the Xpert MTB/RIF assay for extrapulmonary and pulmonary tuberculosis when testing non-respiratory samples : a systematic review. 2014.
[33] World Health Organization. The use of lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis and screening of active tuberculosis in people living with HIV. 2015.
[34] Denkinger CM, Schumacher SG, Boehme CC, Dendukuri N, Pai M, Steingart KR. Xpert MTB/RIF assay for the diagnosis of extrapulmonary tuberculosis: a systematic review and meta-analysis. Eur Respir J 2014;44:435–46.
[35] Walters E, Scott L, Nabeta P, et al. Molecular detection of mycobacterium tuberculosis from stools in young children by use of a novel centrifugation-free processing method. J Clin Microbiol 2018;56:e00781. –18. PMCID: PMC6113478.
[36] Dorman SE, Schumacher SG, Alland D, et al. Xpert MTB/RIF Ultra for detection of Mycobacterium tuberculosis and rifampicin resistance: a prospective multicentre diagnostic accuracy study. Lancet Infect Dis 2018;18:76–84.