Accuracy of upper respiratory tract samples to diagnose Mycobacterium tuberculosis: a systematic review and meta-analysis

Helen R. Savage, Hannah M Rickman, Rachael M Burke, Maria Lisa Odland, Martina Savio, Beate Ringwald, Luis E Cuevas, Peter MacPherson.

Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, United Kingdom (HR Savage MBChB, Prof LE Cuevas MTropMed, DSc, P MacPherson PhD)

London School of Hygiene & Tropical Medicine, London, United Kingdom (HM Rickman MBBChir, RM Burke MRCP, P MacPherson PhD)

Malawi Liverpool-Wellcome Programme, Blantyre, Malawi (HM Rickman MBBChir, RM Burke MRCP, ML Odland PhD, P MacPherson PhD)

The Department of Public Health and Nursing, The Norwegian University of Science and Technology (ML Odland PhD)

The LIGHT Consortium, Liverpool School of Tropical Medicine, Liverpool, United Kingdom (M Savio MPhil, B Ringwald PhD)

Corresponding author: Dr Helen R Savage, Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, L3 5QA, United Kingdom. Helen.savage@lstmed.ac.uk

Keywords: Tuberculosis, diagnosis, systematic review, upper respiratory tract, swab, oral

Word count: Abstract, 297; Manuscript, 3335

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Structured summary (300 words)

Background

Pulmonary tuberculosis (PTB) due to Mycobacterium tuberculosis (Mtb) can be challenging to diagnose because of difficulty obtaining samples, and suboptimal sensitivity of existing tests. We investigated the performance characteristics and diagnostic accuracy of upper respiratory tract tests for diagnosing PTB and hypothesised they would have sufficient accuracy and utility to improve PTB diagnosis.

Methods

A systematic review and meta-analysis was conducted by searching MEDLINE, Cinahl, Web of Science, Global Health, and Global Health Archive databases up to 27/5/2022 to identify studies that reported on the accuracy of upper respiratory tract sampling for TB diagnosis compared to microbiological reference standards. We used a random-effects meta-analysis with a bivariate hierarchical model to estimate pooled sensitivity and specificity, stratified by sampling method. Bias was assessed using QUADAS-2 criteria. Study registered with PROSPERO (CRD42021262392).

Findings

9860 titles were screened for inclusion, 263 studies were assessed for full text review, and 63, comprising 111 test comparisons published between 1933 and 2022 were included in the systematic review (49 in meta-analysis). For laryngeal swabs, pooled sensitivity was 57.8% (95% CI 50.5-65.0%), specificity was 93.8% (95% CI 88.4-96.8%) and diagnostic odds ratio (DOR) was 20.7 (95% CI 11.1-38.8). Nasopharyngeal aspirate sensitivity was 65.2% (95% CI 52.0-76.4%), specificity was 97.9% (95% CI 96.0-99.0%) and DOR was 91.0 (95% CI 37.8-218.8). Oral swabs sensitivity was 58.9% (95% CI 42.9-73.2%), specificity was 85.5% (95% CI 72.9-92.8%), and DOR was 8.47 (95% CI 3.5-20.4).

Interpretation
Upper respiratory tract sampling holds promise to expand access to TB diagnosis, including for people who can’t produce sputum. Exploring historical methods using modern microbiological techniques may further increase the options for alternative sample types. Prospective studies are needed to optimise accuracy and utility of sampling methods in clinical practice.

Funding

HRS is funded by the MRC through the MRC DTP programme at LSTM [Grant number MR/N013514/1].
Research in context

Evidence before this study

Globally in 2021, an estimated 4.2 million of 10.6 million people with incident tuberculosis (TB) disease went undiagnosed, emphasising the urgent need for new diagnostic methodologies. Most TB diagnostics are performed on sputum samples, but people who need TB tests are often unable to produce sputum. Upper respiratory tract sampling for TB diagnosis was widely used historically and holds promise to expand non-sputum-based diagnosis.

Added value of this study

We systematically reviewed and synthesised through meta-analysis diagnostic accuracy evaluations of upper respiratory tract sampling for TB. Historically, upper respiratory tract sampling for TB diagnosis was commonly used, with 39/63 studies conducted before 1970, although in recent years there has been a resurgence of interest in oral sampling. We show that upper respiratory tract samples have acceptable sensitivity and specificity compared to sputum culture, and, if testing is optimised using newer molecular and culture-based methods, may be capable of meeting WHO target produce profiles.

Implications of all the available evidence

Upper respiratory tract sampling methodologies for TB (oral sampling, and sampling from the larynx and nasopharynx) may hold promise to expand access to TB diagnosis, including for people who can’t produce sputum. These sampling strategies can be optimised using modern microbiological techniques to increase access to diagnostics for TB.
Introduction

Pulmonary disease due to *Mycobacterium tuberculosis* (*Mtbt*) can be challenging to diagnose because of lack of access to testing services, difficulty obtaining samples, and suboptimal sensitivity of tests. In 2021, of the estimated 10.6 million people with incident tuberculosis (TB) over 4.2 million went undiagnosed (1), and of those diagnosed with pulmonary TB (PTB) only 63% were bacteriologically confirmed. Bacterial confirmation is important to ensure correct diagnosis and for identification of drug resistance, so that the most effective treatment regime can be identified (1).

The most common sample used for PTB diagnosis is sputum, which can be tested by smear microscopy, culture, or using nucleic acid amplification tests (NAAT) such as Xpert (Cepheid) or Truenat (Molbio). These methods rely on people producing a sputum sample for analysis. However not everyone being investigated for PTB can produce sputum. For adults in the WHO consolidated guidelines on TB (2) sputum or induced sputum are the recommended sample types for diagnosis of PTB and, if HIV positive, urine using the ALERE-LAM. No alternative non-sputum options are recommended for a microbiological diagnosis of PTB to allow confirmation and resistance testing.

Children commonly do not produce sputum and therefore the WHO recommends alternative sample types for diagnosis of PTB, these include induced sputum, gastric aspirate, gastric lavage, nasopharyngeal aspirate (NPA) and stool (3). Gastric aspirate has a sensitivity of 73% (Xpert MTB/RIF) compared to microbiological reference standard (32% compared to composite reference standard), but is invasive, requires fasting and early morning testing with low caregiver acceptability (2,3). NPA is less invasive however still requires specialist equipment (for suction) and has moderate caregiver acceptability, sensitivity against a microbiological reference standard is 46% (Xpert MTB/RIF, based on four studies) (2,3).
Stool is a newly recommended specimen type in the 2021 update, has high acceptability, and is non-invasive; sensitivity of stool Xpert MTB/RIF is 61% against microbiological reference standard (16% against clinical reference standard) (2,3). However stool testing requires waiting for a bowel movement.

Alternative accessible sample types that can be processed with both existing and novel tests, are patient-centred, and can be collected at the time of consultation, are urgently required. WHO Target Product Profiles (TPPs) have defined the minimal and optimal diagnostic accuracy standards for tests to diagnose TB in clinical settings, as well as defining desirable characteristics, including using non-sputum samples such as urine, stool, oral mucosal transudates, saliva, exhaled air, or blood from a fingerstick (4). Apart from stool, currently available alternative sample types, such as gastric aspirate or in adults broncho-alveolar lavage are invasive, need fasting and often admission to a hospital with specialist equipment. This can be a barrier to microbiological diagnosis both in terms of availability of the tests, costs, and accessibility.

In Europe in the early 1900s, physicians investigating patients for TB encountered similar concerns and laryngeal swabs were investigated as an alternative sample type for diagnosis of PTB (5–7); however with the decline of TB in Europe this practice fell out of use. More recently, when looking for alternatives to sputum in children, NPA (a procedure where a catheter is inserted into the nostril and suction applied to produce a sample) has been used, and oral mucosa specimens (samples taken by swabbing either the inside of the cheek or the tongue) from South Africa, Moldova and Kenya have successfully been used to diagnose TB (8–10). Sampling of the upper respiratory tract (from the mouth and nose, tonsils, down to the level of the larynx/vocal cords) which is non/minimally invasive, and can be performed
quickly as an outpatient, offers a potential alternative to expand access to microbiological TB
diagnosis in those that cannot produce sputum.

We therefore set out to systematically appraise the evidence for the performance
characteristics and diagnostic accuracy of upper respiratory tract sampling for diagnosing
active PTB disease. We hypothesised that upper respiratory tract tests would have sufficient
accuracy and utility to increase access to TB diagnosis.

Methods

We undertook a systematic review and meta-analysis; the published protocol is available
online (https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=262392). We
developed a search strategy with information specialists at the Liverpool School of Tropical
Medicine library (Supplemental Table S1). We searched the following databases up to
27/5/2022: MEDLINE; Cinahl; Web of Science; Global Health; and Global Health Archive.

We included studies that evaluated the accuracy of upper respiratory tract sampling (index
tests) for a microbiological (culture and nucleic acid amplification tests [NAAT] including
automated platforms and laboratory PCR) diagnosis of TB disease, compared to a
microbiological reference standard using either sputum or gastric lavage. We included cohort,
cross-sectional, and randomised controlled studies (either published in peer-reviewed
manuscripts, or as pre-prints) that recruited participants from any community or clinical
setting; we placed no restrictions on publication language. We excluded: studies where the
index or reference standard used histological or biomarker-based testing; post-mortem
studies; studies in non-human animal species; case reports; clinical guidelines; and studies
testing for latent TB infection (where active TB disease was not tested for).
Titles and abstracts were imported into a Rayyan.ai database(11), and screened by one reviewer (HRS), with a random subset of 10% checked for agreement by a blinded second reviewer (HMR or RMB). The full text of selected manuscripts was then independently assessed by two reviewers (two of HRS, HMR, RMB) for inclusion; in cases of disagreement, the third reviewer acted as an arbitrator. Risk of bias was assessed using the QUADAS-2 tool (12).

Data were extracted using a piloted extraction form into a study database. Manuscripts published in languages other than English were translated prior to full text review and data extraction.

For each included manuscript, we extracted data on: year; author; country; type of publication; original language; report variable if multiple data sets; study design; setting; population; swab/device used; technique of sampling; site of sampling; number of samples per participant; timing of swabs; swab preparation; transport; storage; method of testing; test cut-off; participants; inclusion criteria; number of samples sets taken; number included in analysis; sex; HIV status; median age; index test positive; index test negative; missing result; reference test positive; reference test negative; missing result; true positive; control population; number of samples sets taken, index test positive, and index test negative.

If a manuscript contained multiple index tests, reference tests or cohorts of patients, each data set was included as a separate report. As participants within a particular study may have been included in more than one index test comparison, we reported the number of sample sets included in the analysis for each test comparison undertaken.

**Data synthesis and analysis**
We summarised study characteristics and, separately for each upper respiratory tract sampling methodology, graphed forest plots and calculated the pooled sensitivity and specificity of the index test compared to reference standards using a bivariate hierarchical random effects model, fitted using the lme4 package in R (version 4.2.1) based on the Cochrane handbook for Systematic Reviews of Diagnostic Test Accuracy(13). If studies were of children a second model, using a composite reference standard, including microbiological and clinical cases of PTB was performed (as in the recent WHO consolidated guidelines (2,3)). Definitions of how clinical cases were defined are provided in the Supplementary Material. Data and code to reproduce analysis are available at: https://osf.io/9nuvq/?view_only=a00f5d373c27461c8b19517673c383aa. This study was registered in PROSPERO (CRD42021262392).

Funding source

HRS is funded by the MRC through the MRC DTP programme at LSTM [Grant number MR/N013514/1]. This research was funded in whole, or in part, by the Wellcome Trust [Grant number 200901/Z/16/Z]. This work was supported by UK Foreign, Commonwealth and Development Office (“Leaving no-one behind: transforming gendered pathways to health for TB”) and partially funded by UK aid from the UK government (to MS, BR, LEC, and PM); however, the views expressed do not necessarily reflect the UK government’s official policies. RMB and HMR are supported by Wellcome PhD Fellowships (203905/Z/16/Z and 225482/Z/22/Z). For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.
Results

9860 studies were identified during the initial search period up to 31/01/2021, and a further 933 when extended to 27/5/2022 (Figure 1). We removed 769 duplicates, leaving 9880 studies that were screened for inclusion. We reviewed the full text of 263 manuscripts. Overall, 63 studies were included in the systematic review, comprising 111 reports on index test comparisons; 49 were included in meta-analyses. Full details of excluded studies and reasons for exclusion are in Supplemental Table S2.

Figure 1: PRISMA 2020 Flow diagram for systematic review.

From included manuscripts (Table 1; Supplementary Table S3 includes full methodological details and results for each study), we classified types of upper respiratory tract sampling into four groups: laryngeal swabs (32 studies); NPAs (nine studies); oral swabs (13 studies); and other (mouthwash: three studies; nasal swabs: one study; saliva: two studies; other mucosa/dental samples: three studies). Studies were published between 1933 and 2022 from South Africa (nine studies), Norway (seven studies), UK (seven studies), Peru (four studies), Uganda (four studies), Canada (three studies), India (three studies), USA (three studies), Australia (two studies), Kenya (two studies), and one study from each of Brazil, Chile, China, Czechoslovakia, Denmark, Finland, France, Germany, Hungary, Italy, Japan, Malawi, Moldova, Mozambique, Spain, South East Asia (not specified), Sweden, Taiwan, and Turkey.

Overall, we included data on 23767 index test samples from participants in hospitals (24 studies); TB sanatoria (nine studies); Chest clinics (nine studies); TB hospitals (eight studies); hospital outpatients (five studies); primary care clinics (three studies); mass screening/asymptomatic/contact interventions (two studies); outpatient treatment centres (one study), prisons (one study) and studies where the setting was unknown (four studies). Eleven
studies were in children aged 16 years or younger. In studies that reported demographic data, 56.5% of participants were male (data from 18 studies) and 16.7% were HIV-positive (with 19 out of 29 studies conducted after 1981 presenting data on HIV prevalence in study participants).
### Table 1: Study Characteristics

<table>
<thead>
<tr>
<th>Year</th>
<th>Author</th>
<th>Country</th>
<th>Study design</th>
<th>Setting</th>
<th>Technique and Device</th>
<th>Site of sampling</th>
<th>Participants</th>
<th>IndeX test analysis</th>
<th>Referenc e test</th>
<th>Inclu ded in analysis</th>
<th>M al e</th>
<th>HIV positive (if study after 1981)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1933</td>
<td>Kereszturi (14)</td>
<td>USA</td>
<td>Diagnostic s evaluation</td>
<td>Hospital</td>
<td>Throat swab</td>
<td></td>
<td>Children with TB</td>
<td>Culture</td>
<td>Gastric lavage culture</td>
<td>10</td>
<td>N R</td>
<td></td>
</tr>
<tr>
<td>1937</td>
<td>Allodi (15)</td>
<td>Italy</td>
<td>Diagnostic s evaluation</td>
<td>Clinic</td>
<td>Nichrome swab passed to larynx with 'fenestrated' cannula using a mirror.</td>
<td>Clinic patients</td>
<td>Culture</td>
<td>Gastric lavage culture</td>
<td>40</td>
<td>N R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1941</td>
<td>Nassau (6)</td>
<td>UK</td>
<td>Diagnostic s evaluation</td>
<td>Sanatorium</td>
<td>Sterilised angulated swab with cotton wool, passed to larynx and cough.</td>
<td>Discharged 1940 – unable to expectorate</td>
<td>Culture</td>
<td>Sputum culture culture</td>
<td>166</td>
<td>N R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Author(s)</td>
<td>Country</td>
<td>Study Type</td>
<td>Institution</td>
<td>Specimen Collection Method</td>
<td>Specimen Type</td>
<td>Response</td>
<td>Status</td>
<td>Culture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>---------</td>
<td>------------</td>
<td>-------------</td>
<td>---------------------------</td>
<td>---------------</td>
<td>----------</td>
<td>--------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1948</td>
<td>Forbes (16)</td>
<td>UK</td>
<td>Diagnosis evaluation</td>
<td>Tuberculosis Hospital</td>
<td>Sterilised angulated nichrome swab with cotton wool, passed to larynx and cough.</td>
<td>Inpatient sputum culture negative</td>
<td>Culture</td>
<td>Gastric lavage culture</td>
<td>100</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1948</td>
<td>Hounslow (7)</td>
<td>UK</td>
<td>Diagnosis evaluation</td>
<td>Tuberculosis Hospital</td>
<td>Sterilised angulated nichrome swab with cotton wool, passed to larynx and cough.</td>
<td>Hospital patients – unable to expectorate</td>
<td>Culture</td>
<td>Gastric lavage culture</td>
<td>193</td>
<td>N R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td>Duggan (17)</td>
<td>Canada</td>
<td>Diagnosis evaluation</td>
<td>Sanatorium</td>
<td>Sterilised angulated swab with cotton wool, passed to larynx and cough.</td>
<td>Inpatients and Outpatients both adults and children</td>
<td>Culture</td>
<td>Gastric lavage culture</td>
<td>100</td>
<td>N R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td>Renoux (18)</td>
<td>France</td>
<td>Diagnosis evaluation</td>
<td>Sanatorium</td>
<td>Sterilised angulated nichrome swab with cotton wool, passed to larynx and cough.</td>
<td>Sanatorium patients – sputum negative</td>
<td>Culture</td>
<td>Gastric lavage culture</td>
<td>56</td>
<td>N R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1951</td>
<td>Armstrong (19)</td>
<td>Canada</td>
<td>Diagnosis evaluation</td>
<td>Hospital</td>
<td>Sterilised angulated swab with cotton wool, passed to larynx and cough.</td>
<td>Hospital patients on TB treatment</td>
<td>Culture</td>
<td>Gastric lavage culture</td>
<td>1024</td>
<td>N R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1951</td>
<td>Szabo (20)</td>
<td>Hungary</td>
<td>Diagnosis</td>
<td>Not stated</td>
<td>Not described</td>
<td>Smear negative TB patients</td>
<td>Culture</td>
<td>NR</td>
<td>2174</td>
<td>N R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Author</td>
<td>Country</td>
<td>Study Type</td>
<td>Sampling Method</td>
<td>Diagnosis</td>
<td>Culture</td>
<td>prescriptions</td>
<td>N R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>---------</td>
<td>------------</td>
<td>-----------------</td>
<td>-----------</td>
<td>---------</td>
<td>--------------</td>
<td>----</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1953</td>
<td>Chaves</td>
<td>USA</td>
<td>Diagnosis</td>
<td>Chest clinic</td>
<td>Sterilised angulated nichrome swab with cotton wool, passed to larynx and cough.</td>
<td>Clinic patients presumptive TB</td>
<td>Gastric lavage culture</td>
<td>1418 N R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1954</td>
<td>Frostad</td>
<td>Norway</td>
<td>Diagnosis</td>
<td>TB Hospital</td>
<td>Not described</td>
<td>TB hospital patients</td>
<td>Gastric lavage culture</td>
<td>1500 N R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1955</td>
<td>Campbell</td>
<td>Australia</td>
<td>Diagnosis</td>
<td>Chest clinic</td>
<td>Sterilised angulated swab with cotton wool, passed to larynx and cough.</td>
<td>Chest clinic – unable to expectorate</td>
<td>Sputum culture</td>
<td>578 N R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1955</td>
<td>Lees</td>
<td>UK</td>
<td>Diagnosis</td>
<td>Mass screening/a symptomatic</td>
<td>Not described</td>
<td>Mass screening admitted for swabs</td>
<td>Bronchial lavage culture</td>
<td>144 N R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1955</td>
<td>Lind</td>
<td>Sweden</td>
<td>Diagnosis</td>
<td>Not stated</td>
<td>Stainless wire with calcium alginate advanced to glottis and cough</td>
<td>Patients with lesions on XR</td>
<td>Gastric lavage culture</td>
<td>121 N R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1955</td>
<td>Wallace</td>
<td>UK</td>
<td>Diagnosis</td>
<td>Hospital</td>
<td>Either cotton wool AC or alginate BDE</td>
<td>Inpatients on TB treatment</td>
<td>Sputum culture</td>
<td>163 N R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Author(s)</td>
<td>Location</td>
<td>Disease Evaluation</td>
<td>Sample Collection Method</td>
<td>Sample Source</td>
<td>Number of Samples</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>----------</td>
<td>--------------------</td>
<td>--------------------------</td>
<td>---------------</td>
<td>-------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1956</td>
<td>Edwards (27)</td>
<td>English</td>
<td>Diagnostic evaluation</td>
<td>Chest clinic</td>
<td>Not described</td>
<td>Outpatients – presumptive TB</td>
<td>Culture</td>
<td>NR</td>
<td>1019</td>
<td>N</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>1956</td>
<td>Engbaek (28)</td>
<td>Denmark</td>
<td>Diagnostic evaluation</td>
<td>Hospital</td>
<td>Sterilised angulated swab with cotton wool, passed to larynx and cough.</td>
<td>Outpatients – unable to expectorate</td>
<td>Culture</td>
<td>Gastric lavage culture</td>
<td>1231</td>
<td>N</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>1956</td>
<td>Tonge (29)</td>
<td>Australia</td>
<td>Diagnostic evaluation</td>
<td>Chest clinic</td>
<td>Sterilised angulated swab with cotton wool, passed to larynx and cough.</td>
<td>Outpatients – presumed and known TB, unable to expectorate</td>
<td>Culture</td>
<td>Gastric lavage culture</td>
<td>465</td>
<td>N</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>1962</td>
<td>Hsing (30)</td>
<td>Taiwan</td>
<td>Diagnostic evaluation</td>
<td>Hospital</td>
<td>Sterilised angulated swab with cotton wool, passed to larynx and cough.</td>
<td>Hospital patients</td>
<td>Culture</td>
<td>Gastric lavage culture</td>
<td>1320</td>
<td>N</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>1962</td>
<td>Velu (31)</td>
<td>India</td>
<td>Diagnostic evaluation</td>
<td>Outpatient treatment study</td>
<td>Sterilised angulated swab with cotton wool, passed to larynx and cough.</td>
<td>Outpatients – bacteriologically confirmed TB</td>
<td>Culture</td>
<td>Sputum culture</td>
<td>2809</td>
<td>N</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>1962</td>
<td>Kertay (32)</td>
<td>Germany</td>
<td>Diagnostic evaluation</td>
<td>State TB institute</td>
<td>Not described</td>
<td>TB hospital patients treated for TB</td>
<td>Culture</td>
<td>Sputum culture</td>
<td>212</td>
<td>N</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>1965</td>
<td>Hauge (33)</td>
<td>Norway</td>
<td>Diagnostic evaluation</td>
<td>Hospital</td>
<td>Not described</td>
<td>Hospital patients – PTB</td>
<td>Culture</td>
<td>NR</td>
<td>157</td>
<td>N</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Author</td>
<td>Country</td>
<td>Study Design</td>
<td>Site</td>
<td>Specimen Type</td>
<td>Specimen Collection Method</td>
<td>Patient Description</td>
<td>Ref.</td>
<td>Culture</td>
<td>Sr.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>---------</td>
<td>--------------</td>
<td>-----</td>
<td>---------------</td>
<td>---------------------------</td>
<td>-------------------</td>
<td>-----</td>
<td>---------</td>
<td>-----</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1966</td>
<td>Pechacek (5)</td>
<td>Czech Republic</td>
<td>Diagnostic evaluation</td>
<td>TB hospital</td>
<td>Not described</td>
<td>TB hospital patients – referred as PTB some already treated</td>
<td>Cultu re</td>
<td>Sputum culture</td>
<td>525</td>
<td>N R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1968</td>
<td>Lloyd (34)</td>
<td>English</td>
<td>Diagnostic evaluation</td>
<td>Hospital</td>
<td>Stainless steel wire, wet cotton wool, bent to angle</td>
<td>Paediatric inpatients with PTB</td>
<td>Cultu re</td>
<td>Gastric lavage culture</td>
<td>60</td>
<td>N R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1970</td>
<td>Mankiewicz (35)</td>
<td>Canada</td>
<td>Diagnostic evaluation</td>
<td>Screening/Contacts/A Symptomatic</td>
<td>Alginate wool sterilised angulated swab, into larynx and cough</td>
<td>Outpatients – unable to expectorate</td>
<td>Cultu re</td>
<td>NR</td>
<td>1199</td>
<td>N R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>Thakur (36)</td>
<td>India</td>
<td>Diagnostic evaluation</td>
<td>Hospital outpatients</td>
<td>Sterilised angulated swab, passed to larynx and cough</td>
<td>Outpatients – children under 14 probable TB</td>
<td>Cultu re</td>
<td>NR</td>
<td>51</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1948</td>
<td>Gilje(37)</td>
<td>Norway</td>
<td>Diagnostic evaluation</td>
<td>Sanatorium</td>
<td>Sterilised angulated swab with cotton wool, passed to larynx and cough (Laryngeal streak)</td>
<td>Inpatients – unable to expectorate</td>
<td>Cultu re</td>
<td>Gastric lavage culture</td>
<td>400</td>
<td>N R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td>Lundar (38)</td>
<td>Norway</td>
<td>Diagnostic evaluation</td>
<td>Sanatorium</td>
<td>Sterilised angulated swab with cotton wool, passed to larynx and cough (Laryngeal streak)</td>
<td>Inpatients – unable to expectorate</td>
<td>Cultu re</td>
<td>Gastric lavage culture</td>
<td>304</td>
<td>N R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Name</td>
<td>Country</td>
<td>Study Type</td>
<td>Institution</td>
<td>Procedure Description</td>
<td>Diagnosis Method</td>
<td>Test Type</td>
<td>Test Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>---------------</td>
<td>---------</td>
<td>--------------</td>
<td>-------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td>-----------</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1951</td>
<td>Gilje(39)</td>
<td>Norway</td>
<td>Diagnostic</td>
<td>Sanatorium</td>
<td>Sterilised angulated swab with cotton wool, passed to larynx and cough (Laryngeal streak)</td>
<td>Inpatients – unable to expectorate</td>
<td>Culture</td>
<td>863</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1952</td>
<td>Roald (40)</td>
<td>Norway</td>
<td>Diagnostic</td>
<td>Sanatorium</td>
<td>Sterilised angulated swab with cotton wool, passed to larynx and cough (Laryngeal streak)</td>
<td>Inpatients</td>
<td>Culture</td>
<td>483</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1952</td>
<td>Smedsrud (41)</td>
<td>Norway</td>
<td>Diagnostic</td>
<td>Sanatorium</td>
<td>Sterilised angulated swab with cotton wool, passed to larynx and cough (Laryngeal streak)</td>
<td>Inpatients</td>
<td>Culture</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1953</td>
<td>Laes (42)</td>
<td>Finland</td>
<td>Diagnostic</td>
<td>Sanatorium</td>
<td>Sterilised angulated swab with cotton wool, passed to larynx and cough (Laryngeal streak)</td>
<td>Inpatients – negative Ziehl-Nielsen</td>
<td>Culture</td>
<td>232</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nasopharyngeal aspirate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>Franchi (43)</td>
<td>Peru</td>
<td>Diagnostic</td>
<td>Hospital</td>
<td>Feeding catheter</td>
<td>Children presumptive TB</td>
<td>Culture/PCR</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>Owens (</td>
<td>Uganda</td>
<td>Diagnostic</td>
<td>Hospital</td>
<td>Graduated</td>
<td>Oropharynx</td>
<td>Culture/Inducer</td>
<td>94</td>
<td>57 44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Study</td>
<td>Country</td>
<td>Setting</td>
<td>Type of Evaluation</td>
<td>Collection Device</td>
<td>Site of Collection</td>
<td>Study Type</td>
<td>Diagnostic Test(s)</td>
<td>Culture Ed</td>
<td>Induced Sputum Culture Ed</td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>-------------</td>
<td>-----------------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>---------------------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>2018</td>
<td>Cakir 45</td>
<td>Turkey</td>
<td>Hospital</td>
<td>Suction catheter</td>
<td>Oropharynx</td>
<td>Children presumptive TB</td>
<td>Culture</td>
<td>Gastric aspirate culture</td>
<td>40</td>
<td>18</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>Zar 46</td>
<td>South Africa</td>
<td>Hospital</td>
<td>Sterile catheter with mucus trap</td>
<td>Nasopharynx</td>
<td>Children presumptive TB</td>
<td>Culture/Xpert MTB/RIF</td>
<td>Induced sputum culture</td>
<td>535</td>
<td>N 117</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>Zar 47</td>
<td>South Africa</td>
<td>Primary care clinic</td>
<td>Sterile catheter with mucus trap</td>
<td>Nasopharynx</td>
<td>Children presumptive TB</td>
<td>Xpert MTB/RIF</td>
<td>Induced sputum culture</td>
<td>384</td>
<td>18 31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>Hanrahan 48</td>
<td>South Africa</td>
<td>Primary care clinic</td>
<td>Fasted 2 hours sterile catheter and mucus trap</td>
<td>Nasopharynx</td>
<td>Children presumptive TB</td>
<td>Culture/Xpert MTB/RIF</td>
<td>Induced sputum culture</td>
<td>105</td>
<td>63 21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>Zar 49</td>
<td>South Africa</td>
<td>Hospital</td>
<td>Sterile catheter with mucus trap</td>
<td>Nasopharynx</td>
<td>Hospitalised presumptive PTB</td>
<td>Xpert Ultra</td>
<td>Induced sputum</td>
<td>195</td>
<td>N 32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Study</td>
<td>Country</td>
<td>Study Design</td>
<td>Sample Group</td>
<td>Sampling Method</td>
<td>Test</td>
<td>Gene Expression</td>
<td>Sensitivity (%)</td>
<td>Specificity (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>---------</td>
<td>--------------</td>
<td>--------------</td>
<td>-----------------</td>
<td>------</td>
<td>-----------------</td>
<td>----------------</td>
<td>----------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2021</td>
<td>Song (50)</td>
<td>Kenya</td>
<td>Cross sectional</td>
<td>Inpatient and Outpatient</td>
<td>French tubing and mucus trap</td>
<td>Nasopharynx nx</td>
<td>Children presumptive TB</td>
<td>Xpert MTB/RIF / MGIT</td>
<td>Gastric aspirate MGIT</td>
<td>294</td>
<td>14</td>
<td>73</td>
</tr>
<tr>
<td>2020</td>
<td>Osorio(51)</td>
<td>Mozambique</td>
<td>Retrospective cross-sectional</td>
<td>Hospital</td>
<td>Not stated</td>
<td>Nasopharynx nx</td>
<td>Children with severe acute malnutrition</td>
<td>Xpert MTB/RIF / MGIT</td>
<td>NR</td>
<td>45</td>
<td>24</td>
<td>10</td>
</tr>
</tbody>
</table>

**Oral swab**

<table>
<thead>
<tr>
<th>Year</th>
<th>Study</th>
<th>Region</th>
<th>Study Design</th>
<th>Sample Group</th>
<th>Sampling Method</th>
<th>Test</th>
<th>Gene Expression</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>Wood (52)</td>
<td>South East Asia</td>
<td>Conference abstract</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>54</td>
</tr>
<tr>
<td>2015</td>
<td>Wood(53)</td>
<td>South Africa</td>
<td>Case control</td>
<td>TB clinic</td>
<td>Omnischwab</td>
<td>Inside of cheek 10 seconds</td>
<td>TB clinic patients</td>
<td>PCR</td>
<td>Sputum Xpert MTB/RIF</td>
</tr>
<tr>
<td>2019</td>
<td>Luabeya (8)</td>
<td>South Africa</td>
<td>Case control</td>
<td>TB Clinic</td>
<td>Omnischwab x 3, Puritan Purflock Ultra x 1 cheek</td>
<td>Tongue swabs</td>
<td>TB clinic patients</td>
<td>PCR</td>
<td>Sputum Xpert MTB/RIF</td>
</tr>
<tr>
<td>2019</td>
<td>Mesman (54)</td>
<td>Peru</td>
<td>Case control</td>
<td>TB clinic</td>
<td>Omnischwab in lysis buffer, Omnischwab in PBS, FTA</td>
<td>Cheek</td>
<td>TB clinic patients</td>
<td>Xpert MTB/RIF</td>
<td>Sputum MGIT</td>
</tr>
<tr>
<td>Year</td>
<td>Authors</td>
<td>Country</td>
<td>Study Design</td>
<td>Collection Method</td>
<td>Specimen</td>
<td>Sample Preparation</td>
<td>Diagnosis</td>
<td>PCR</td>
<td>Sputum Culture</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>---------</td>
<td>--------------</td>
<td>-------------------</td>
<td>----------</td>
<td>--------------------</td>
<td>-----------</td>
<td>-----</td>
<td>----------------</td>
</tr>
<tr>
<td>2019</td>
<td>Nicol (55)</td>
<td>South Africa</td>
<td>Prospective diagnostic evaluation</td>
<td>Hospital</td>
<td>Omniswab/Puritan Purflock</td>
<td>Cheek (right and left)</td>
<td>Presumptive pulmonary TB</td>
<td>PCR</td>
<td>Induced sputum Xpert MTB/RIF</td>
</tr>
<tr>
<td>2020</td>
<td>Flores (56)</td>
<td>Peru</td>
<td>Prospective diagnostic evaluation</td>
<td>Not stated</td>
<td>Omniswab in sterile lysis buffer or EasiCollect FTA card</td>
<td>Inside of both cheeks 10 seconds</td>
<td>Presumptive pulmonary TB</td>
<td>PCR</td>
<td>Bacteriologically confirmed</td>
</tr>
<tr>
<td>2020</td>
<td>Mesman (57)</td>
<td>Peru</td>
<td>Case control</td>
<td>Health centres</td>
<td>Omniswab, Easicollect FTA card</td>
<td>Inside of each cheek 10 seconds</td>
<td>Culture confirmed TB</td>
<td>PCR</td>
<td>Sputum culture</td>
</tr>
<tr>
<td>2020</td>
<td>Molina-Moya (58)</td>
<td>Moldova</td>
<td>Cross sectional</td>
<td>Hospital</td>
<td>Primeswab</td>
<td>Swab cheek 7-8 seconds</td>
<td>Clinical diagnosis presumptive TB</td>
<td>PCR</td>
<td>Sputum culture</td>
</tr>
<tr>
<td>2020</td>
<td>Lima (59)</td>
<td>Brazil</td>
<td>Case control</td>
<td>Prison</td>
<td>Not stated</td>
<td>Tongue 10 seconds</td>
<td>Sputum Xpert positive</td>
<td>PCR</td>
<td>Sputum Xpert Ultra</td>
</tr>
<tr>
<td>2021</td>
<td>Ealand (60)</td>
<td>South Africa</td>
<td>Case control</td>
<td>Hospital</td>
<td>Copan swab</td>
<td>Tongue 6-8 seconds</td>
<td>Young children less than equal 5 clinically diagnosed with TB</td>
<td>Culture, Auramine/spoliotyp</td>
<td>Gastric aspirate culture/Xp</td>
</tr>
<tr>
<td>Year</td>
<td>Author (Ref)</td>
<td>Location</td>
<td>Study Design</td>
<td>Specimen Type</td>
<td>Sample Collection Method</td>
<td>Test</td>
<td>Diagnostic Method</td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td>----------</td>
<td>--------------</td>
<td>---------------</td>
<td>--------------------------</td>
<td>------</td>
<td>------------------</td>
<td>------------</td>
<td>-----------</td>
</tr>
<tr>
<td>2021</td>
<td>Song (61)</td>
<td>China</td>
<td>Prospective diagnostic evaluation</td>
<td>Swab</td>
<td>Dorsum tongue 7-8 times</td>
<td>&gt; 16 years symptoms suggestive pulmonary TB</td>
<td>TB-LAMP</td>
<td>101</td>
<td>69</td>
</tr>
<tr>
<td>2021</td>
<td>Wood (62)</td>
<td>Uganda</td>
<td>Prospective diagnostic evaluation</td>
<td>Copan FLOQSwab</td>
<td>Front 2/3 tongue 15-20 seconds</td>
<td>Presumptive TB</td>
<td>qPCR/Culture</td>
<td>103</td>
<td>N</td>
</tr>
<tr>
<td>2022</td>
<td>LaCours e (10)</td>
<td>Kenya</td>
<td>Cross-sectional</td>
<td>OmniSwab</td>
<td>Left and Right buccal surface</td>
<td>Presumptive TB</td>
<td>qPCR</td>
<td>Sputum culture/Xpert Ultra MTB/RIF</td>
<td>100</td>
</tr>
</tbody>
</table>

**Mouth wash**

<table>
<thead>
<tr>
<th>Year</th>
<th>Author (Ref)</th>
<th>Location</th>
<th>Study Design</th>
<th>Specimen Type</th>
<th>Sample Collection Method</th>
<th>Test</th>
<th>Diagnostic Method</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive Predictive Value</th>
<th>Negative Predictive Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1955</td>
<td>Rogers (63)</td>
<td>USA</td>
<td>Diagnostic evaluation</td>
<td>Mouthwash</td>
<td>Pharynx</td>
<td>Proven or presumptive TB</td>
<td>Culture</td>
<td>Sputum culture</td>
<td>43</td>
<td>N</td>
<td>R</td>
</tr>
<tr>
<td>1994</td>
<td>Evans (64)</td>
<td>UK</td>
<td>Case control</td>
<td>Mouthwash</td>
<td>Pharynx</td>
<td>Active TB and controls</td>
<td>Lysed rRNA PCR</td>
<td>Sputum culture</td>
<td>39</td>
<td>N</td>
<td>NR</td>
</tr>
<tr>
<td>Year</td>
<td>Authors (Reference)</td>
<td>Country</td>
<td>Study Type</td>
<td>Site</td>
<td>Sample Collection</td>
<td>Diagnostic Test</td>
<td>Test Site</td>
<td>Outpatients and Inpatients</td>
<td>PCR</td>
<td>Sputum culture</td>
<td>2009 Davis (65)</td>
</tr>
<tr>
<td>------</td>
<td>---------------------</td>
<td>---------</td>
<td>------------</td>
<td>------</td>
<td>-------------------</td>
<td>----------------</td>
<td>----------</td>
<td>---------------------------</td>
<td>------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Nasal swab</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>Warn Dorff (66)</td>
<td>Malawi</td>
<td>Case control</td>
<td>Hospital</td>
<td>Cotton wool nasal swab</td>
<td>Nose</td>
<td>Inpatients and contacts</td>
<td>PCR Sputum culture</td>
<td>52 NR NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saliva</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>Shenai (67)</td>
<td>South Africa/Korea</td>
<td>Diagnostic evaluation</td>
<td>Hospital</td>
<td>Salivette saliva collector</td>
<td>Mouth</td>
<td>Adult PTB diagnosed patients</td>
<td>Xpert MTB/ RIF Sputum culture</td>
<td>26 NR 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>Gonzalez (68)</td>
<td>Spain</td>
<td>Case control</td>
<td>Hospital</td>
<td>Spitting method into sterile container</td>
<td>Mouth</td>
<td>Adult presumptive TB</td>
<td>Culture Sputum culture</td>
<td>32 21 NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>Eguchi (69)</td>
<td>Japan</td>
<td>Case control</td>
<td>Hospital</td>
<td>Unstimulated saliva in test tube</td>
<td>Mouth</td>
<td>Patients with TB</td>
<td>Culture/PCR Culture</td>
<td>75 67 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>Palakuru (70)</td>
<td>India</td>
<td>Case control</td>
<td>Outpatient clinic</td>
<td>Unstimulated saliva in sterile vial</td>
<td>Mouth</td>
<td>Pulmonary TB</td>
<td>PCR Acid-fast bacilli positive</td>
<td>25 NR 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>Balcells (71)</td>
<td>Chile</td>
<td>Cross sectional</td>
<td>Outpatient clinic</td>
<td>Bilateral nasopharyngeal</td>
<td>Nasopharyngeal</td>
<td>Adult contacts of smear</td>
<td>Xpert MTB N/A 77 34 NR</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Laryngeal swabs

Forty-one reports that evaluated the accuracy of laryngeal swabs were included, of which 23 contributed data to the meta-analysis (Supplementary Table S4). Studies were published between 1941 and 1968, with sample sizes ranging from 10 to 2809. A variety of angulated metal swabs with cotton wool or other absorbent materials were used to sample from the larynx, and processed using acidic preparations (Table 1 and Supplemental Table S3).

Samples (both index and reference) were then cultured using methodologies available at the time of the study including modified Petragnani’s medium, Lowenstein-Jensen solid medium, liquid oleic acid albumin media and guinea pig inoculation. In all studies, the reference test used was culture of sputum or gastric lavage fluid (using either expectorated sputum [eight studies] or gastric lavage [15 studies]). Random-effects model estimated sensitivity was 57.8% (95% CI: 50.5-65.0%), specificity 93.8% (95% CI: 88.4-96.8%), diagnostics odds ratio (DOR) 20.7 (95% CI: 11.1-38.8), positive likelihood ratio (PLR) 9.3 (95% CI: 5.1-17.0), and negative likelihood ratio 0.45 (95% CI 0.38-0.53) – Figure 2 (SROC Supplementary Figure S1).

Figure 2: Sensitivity and specificity of laryngeal sampling for active pulmonary tuberculosis, with random effects meta-analysis
Nasopharyngeal aspirates

Nine studies that included data on the accuracy of NPA were included, providing 17 test comparisons, of which 10 were included in the meta-analysis. All studies were conducted among children and published between 1998 and 2021, with the number of sample sets ranging from 64 to 535. Multiple index testing methodologies (non-automated culture [four reports], MGIT [one report], Xpert MTB/RIF [three reports], Xpert Ultra [one report], PCR [one report]) were carried out on samples and each was included as a separate comparison.

For the microbiological reference test induced sputum culture (6 reports) or gastric aspirate culture (4 reports) were used – Supplementary Table S5. The proportion of children with a microbiologically confirmed diagnosis of PTB ranged from 3% (48) to 38% (43) within the included studies. Model-estimated sensitivity was 65.2% (95%CI 52.0-76.4%), specificity 97.9% (95% CI 96.0-99.0%), DOR 91.0 (95% CI 37.8-218.8), PLR 32.2 (95% CI 15.8-66.1) and NLR 0.35 (95%CI 0.25-0.51) – Figure 3 (SROC, Supplementary Figure 2).

Figure 3: Sensitivity and specificity of nasopharyngeal aspirate for active pulmonary tuberculosis with microbiological reference standard, with random effects meta-analysis

It is made available under a CC-BY 4.0 International license .
A second model was calculated of reports that included study data of a composite reference standard (both microbiological and clinical diagnosis of PTB, see Supplementary Table 6 for inclusion criteria), with 12 comparisons included in the analysis (Supplementary Table S7). The random effects model gave an estimate of sensitivity 9.1% (95%CI 5.6-14.6%), specificity 99.9% (95% CI 93.6-99.9%), DOR 168.4 (95%CI 1.57-17959.1), PLR 153.1 (95% CI 1.43-16343.5) and NLR 0.91 (95%CI 0.87-0.95) – Figure 4 (SROC, Supplementary Figure 3).

**Figure 4:** Sensitivity and specificity of nasopharyngeal aspirate for active pulmonary tuberculosis when clinical diagnosis used as a reference standard, with random effects meta-analysis.
Oral swab

Thirteen studies, with 39 comparisons of oral swab samplings to microbiological reference standards were identified; of these sixteen (2015-2022) were included in the meta-analysis.

Nine used PCR for analysis, two used culture and one comparison each used Xpert Ultra, Xpert MTB/RIF, TB-LAMP, spogliotyping and auramine smear. Eight comparisons had sputum culture as reference standard, two used sputum culture and Xpert, two used sputum Xpert alone, and four used gastric lavage culture and Xpert (Supplementary Table S8). Six comparisons included children as participants and ten adults. Pooled sensitivity of oral swab samples was 58.5% (95% CI 43.7-71.7%), specificity 85.6% (95% CI 73.9-92.9%), DOR 8.37 (95%CI 3.5-19.9), PLR 4.1 (95% CI 2.1-7.9) and NLR 0.49 (95%CI 0.35-0.68) – Figure 5.

**Figure 5:** Sensitivity and specificity of oral swab for active pulmonary tuberculosis, with random effects meta-analysis
Other sample types

We identified only a small number of studies of alternative sample types, so meta-analysis was not undertaken. Of these three studies compared the accuracy of mouthwash samples, saliva (one study), oral cavity samples (two studies) and nasopharyngeal swabs (one study) - Supplementary Table S9. A further study on nasal swabs and another on saliva did not contain sufficient information to permit summary of results.

Risk of bias assessment

The QUADAS-2 tool was used to assess the risk of bias for each study included (Figure 6, and for individual studies - Supplementary Figures S5 and S6). Overall, studies performed before the 1950s did not report on multiple domains, especially participant selection and use of index tests. In newer studies there was a high risk of bias in Domain 1 (participant selection) due to case control designs and variability in reference standards used, with some reference standards less likely to correctly classify true TB status.

Figure 6: Risk of bias assessment
Diagnosing PTB is challenging, and new, accurate sampling approaches that utilise easier-to-obtain specimens are urgently required. In this systematic review and meta-analysis, which included data on 23767 sample comparisons from studies conducted over an 89-year period between 1933 and 2022, we found that three upper respiratory tract sample types show promise for accurately diagnosing PTB: laryngeal swabs with a pooled sensitivity of 57.8% and a specificity of 93.8%; NPA with a pooled sensitivity and specificity of 65.2% and 97.9%; and oral swabs with 58.5% and 85.6% against microbiological reference standards. These accuracy estimates are similar to currently approved TB diagnostic tools on non-sputum samples such as urine lateral flow lipoarabinomannan assays in those who are HIV-positive (72) and stool in children (3).

Studies that used laryngeal swabs as a sample type were mostly from Europe in the early 20th century and gave a pooled sensitivity of 57.8% against microbiological reference standards (gastric lavage or sputum culture), here there was large variation in swabbing and culture methodologies used over time as new medias and techniques were invented, and all were conducted before the advent of molecular diagnostics. The studies were mostly carried out...
among outpatients being investigated for TB, who were unable to produce sputum, with the
predominant reference standard being culture of gastric lavage. Use of laryngeal swab testing
was discontinued with the advent of chemotherapy and the rapid decline of TB cases in
Northern Europe and the closure of sanatoria (73,74), however the original rationale for
investigating this sample type, needing an outpatient test that was quick and well-tolerated,
reflects similar diagnostic challenges still being faced in high burden countries and the
requirements of the WHO TPP (4). Laryngeal swabs may hold promise as an alternative to
expand access to TB diagnosis if re-evaluated in prospective studies using modern diagnostic
platforms and culture..

The second sample type reviewed was NPA, which were mostly studied in children. NPAs
had a sensitivity 65.5% compared to culture of gastric aspirate or induced sputum samples.
However as microbiological tests of TB in children have a low sensitivity(75) we also
cmpared NPA to a clinical reference standard for TB, where sensitivity fell to 9%. This
reflects the very high proportion of children who were clinically diagnosed with TB in the
absence of any positive microbiological test: a common occurrence in clinical practice which
reflects the urgent need for improved diagnostic strategies in children. NPA and laryngeal
swabs have the benefit of being able to be performed as an outpatient test rather than
requiring admission, which could make them a viable alternative in settings where access to
hospital services and admission are limited due to cost, distance, and availability. Further
research is needed to investigate whether sensitivity can be improved by optimising sample
preparation, collection or analysis, how NPA performs in adults, and how NPA could be used
within diagnostic algorithms to increase microbiological diagnosis (50).

Oral swabs gave a pooled sensitivity of 58.5% against microbiological reference standards.
However, many of the studies that evaluated oral swabs were case control studies with small
numbers which have a high risk of bias and would be expected to over-estimate sensitivity.
There was also a wide variability in the swab types, sampling methodologies, and specimen processing and analysis. Nevertheless, oral swabs do offer an easy option for sample collection and transport, potentially widening access to TB diagnosis. Further research is needed including prospective diagnostic evaluations to find standardised methodologies and assess its accuracy in a wide range of populations, including children.

Limitations for this systematic review and meta-analysis include the variability of the methods of sample collection and analysis. We mitigated against this by stratifying studies into groups of sample types for meta-analysis; however, some within-group variability remains. For some studies from the early 20th century, we were unable to identify the full text despite extensive searches with librarians, and especially if they were published in a language other than English; this may have biased results to studies in more prominent journals that were more likely to be archived in libraries and those in English. Newer studies, particularly those that examined the accuracy of oral swabs, mainly used case control designs, and had small sample sizes.

Although upper respiratory tract sampling seems a promising avenue for improving access to TB diagnostic testing, further prospective studies are needed to optimise sampling and microbiological methodologies to maximise sensitivity. In children, where novel diagnostics are urgently needed NPA, potentially laryngeal swabs, and a combination of sample sites may offer an alternative sampling methodology that can be used in outpatient settings to widen access. Historical methods using laryngeal swabs showed similar sensitivity and higher specificity than modern studies using oral swabbing in much larger numbers of patients, some of whom were unable to expectorate. Updated evaluations of laryngeal swabbing is required to determine whether these findings are replicated in modern diagnostic accuracy evaluations and across a range of populations. Oral swabs are simple to collect and transport, however this review has shown that more prospective study data is needed to
understand whether sensitivity is sufficient for use in clinical practice. If so, this may give a
sample type that could be used in outpatient settings in children and adults who are unable to
expectorate or in an inpatient setting in those too unwell to produce a sample. Overall upper
respiratory tract sampling may offer an alternative non-invasive sample type that can be used
to increase access to microbiological diagnosis for Mtb.

**Contributors:** The study was conceived by HRS and LEC. The study design was developed
by HRS, LEC and PMP. Data extraction was conducted by HRS, MLO, MS and BR. Data
analysis and interpretation were conducted by HRS, PMP, HMR and RMB. The initial
manuscript was prepared by HRS and PMP. All authors edited and approved the final
manuscript (HRS, HMR, RMB, MLO, MS, BR, LEC and PMP).

**Declaration of interests:** We declare no competing interests.

**Data sharing:**
Data of studies and analytical code available at:

https://osf.io/9nuvq/?view_only=a00f5d373c27461c8b19517673c383aa
References


28. Engbaek V HC: Holm, S: Melton. Examination for Tubercle Bacilli by the Gastric
Lavage and Laryngeal Swab Methods. A Comparative Study. Acta Tuberculosea

29. Tonge PG JI: Hughes. A Comparative Study of Laryngeal Swabs and Gastric Lavage, in
the Detection of Tubercle Bacilli. American Review of Tuberculosis and Pulmonary

30. Hsing YT CT: Ma. A Comparative Study of the Efficacy of the Laryngeal Swab,
Bronchial Lavage, Gastric Lavage, and Direct Sputum Examination Methods in detecting
Tubercle Bacilli in a Series of 1,320 Patients. American Review of Respiratory Disease.

Examination of a Sputum Collection and a Pair of Laryngeal Swab Specimens in Patients

32. Kertay N, Marton S. [Comparative bacteriological investigations of the detection of
Mycobacterium tuberculosis, using bronchial secretions obtained with the Marton

33. Hauge J HE: Schaanning. The Value of Repeated Laryngeal Swab Examinations during

34. Lloyd AVC. Bacteriological diagnosis of tuberculosis in children. A comparative study

35. Mankiewi.E I : Shimro. LARYNGEAL SWAB METHOD FOR DETECTION OF
TUBERCLE BACILLI IN PULMONARY TUBERCULOSIS. European Journal of
Clinical and Biological Research. 1970;15(3):328–.

36. Thakur CA A: Coulter, JB: Zutshi, K: Pande, HK: Sharma, M: Banerjee, A: Richardson,

37. Gilje A. Cultural Demonstration of Tubercle Bacilli by Laryngeal Swabbing. / Dyrking

38. Lundar J. [A comparison of laryngeal swabbing and gastric lavage as methods of


40. Roald OK S: Thomassen. Reliability of the Laryngeal Swab Method for Tubercle Bacilli
Cultures. / Larynxkulturers pålitelighet for påvising av tuberkelbasiller. Nordisk Medicin.

41. Smedsrud K. Cultures from Laryngeal Swabs as a Method of assessing the Quantity of
Bacilli in Pulmonary Tuberculosis. / Larynxkultur til kvantitativ vur-dering av
basillutskillelsen ved tuberkulose. Deres betydning i det praktiske tuberkulosear-beid.
42. 1953 Laes et al.pdf.


List of tables

Table 1: Study details included studies.

List of figures

Figure 1: PRISMA 2020 Flow diagram
Figure 2: Sensitivity and specificity of laryngeal sampling for active pulmonary tuberculosis, with random effects meta-analysis

Figure 3: Sensitivity and specificity of nasopharyngeal aspirate for active pulmonary tuberculosis, with random effects meta-analysis

Figure 4: Sensitivity and specificity of nasopharyngeal aspirate for active pulmonary tuberculosis when clinical diagnosis used as a reference standard, with random effects meta-analysis

Figure 5: Sensitivity and specificity of oral swab for active pulmonary tuberculosis, with random effects meta-analysis

Figure 6: A bar chart to show the percentage of studies with High, Low or unclear bias for each domain.

Supplementary Table

S1: Search strategy
S2: Excluded studies
S3: Full study information
S4: Table of data from laryngeal swab reports used in meta-analysis.
S5: Table of data from naso-pharyngeal aspirate reports used in meta-analysis.
S6: Table to show definition of confirmed, probable or not PTB by study.
S7: Table of data from naso-pharyngeal aspirate reports and clinical diagnosis as a reference standard used in the meta-analysis.
S8: Data table to show analysis of reports of oral swabs as sample type.
S9: Data table to show analysis of reports of other sample type data.

Supplementary Figures:

SF1: Sensitivity and specificity of laryngeal sampling for active pulmonary tuberculosis, with random effects meta-analysis
SF2: Sensitivity and specificity of nasopharyngeal aspirate for active pulmonary tuberculosis, with random effects meta-analysis

SF3: Sensitivity and specificity of nasopharyngeal aspirate for active pulmonary tuberculosis when clinical diagnosis used as a reference standard, with random effects meta-analysis

SF4: Sensitivity and specificity of oral swab for active pulmonary tuberculosis, with random effects meta-analysis

SF5: Bias of individual studies presented via each domain question and overall rating.

SF6: Bias of individual studies presented by domain.