



Research paper

Diagnosis of tuberculosis through breath test: A systematic review

Antonia M.I. Saktiawati^{a,b,c}, David Dwi Putera^d, Althaf Setyawan^e,
 Yodi Mahendradhata^{c,f}, Tjip S. van der Werf^{b,g,*}

^a Department of Internal Medicine, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

^b University of Groningen, University Medical Centre Groningen, Department of Pulmonary Diseases and Tuberculosis, Groningen, the Netherlands

^c Center for Tropical Medicine, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

^d Independent Researcher, Yogyakarta, Indonesia

^e Department of Biostatistics, Epidemiology, and Population Health, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

^f Department of Health Policy and Management, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

^g University of Groningen, University Medical Center Groningen, Department of Internal Medicine-Infectious Diseases, Groningen, the Netherlands



ARTICLE INFO

Article history:

Received 14 June 2019

Received in revised form 20 July 2019

Accepted 22 July 2019

Available online 8 August 2019

Keywords:

Breath test

Diagnosis

Tuberculosis

Sensitivity

Specificity

Accuracy

ABSTRACT

Background: Breath tests may diagnose tuberculosis (TB) through detecting specific volatile organic compounds produced by *Mycobacterium tuberculosis* or the infected host.

Methods: To estimate the diagnostic accuracy of breath test with electronic-nose and other devices against culture or other tests for TB, we screened multiple databases until January 6, 2019.

Findings: We included fourteen studies, with 1715 subjects in the analysis. The pooled sensitivity and specificity of electronic-nose were 0.93 (95% CI 0.82–0.97) and 0.93 (95% CI 0.82–0.97), respectively, and no heterogeneity was found. The sensitivity and specificity of other breath test devices ranged from 0.62 to 1.00, and 0.11 to 0.84, respectively.

Interpretation: The low to moderate evidence of these studies shows that breath tests can diagnose TB accurately, however, to give a real-time test result, additional development is needed. Research should also focus on sputum smear negative TB, children, and the positioning of breath testing in the diagnostic work flow.

Funding: The authors received no specific funding for this work.

© 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Tuberculosis (TB) is the leading infectious cause of morbidity and mortality globally [1]. It spreads easily by air, with inhalation of infected droplets exhaled, coughed or sneezed by individuals affected with contagious forms of pulmonary TB. Approximately 23% of the world's population is infected with TB bacteria, and 5–15% will suffer from TB at some point in time [1,2]. Diagnostic delay is a major concern in TB control: there is an average loss of one to three months delay between the first day that patients present to the health care system, and the moment of diagnosis [3,4]. In 2017, there was a gap of 3·6 million between notifications of new cases and the estimated number of incident cases [1], indicating a huge underreporting and under-diagnosis of TB cases. In response, the World Health Organization (WHO) launched an initiative to detect many more people with TB in the next few years [1].

The WHO criteria for pulmonary TB (PTB) diagnosis include clinical symptoms and isolation of *M. tuberculosis* from sputum by culture or

by molecular assays as the line probe assays (LPAs), or detection of acid-fast bacilli by sputum smear microscopy (SSM) if culture or LPAs is unavailable, or – for smear-negative PTB patients – with chest radiography (CXR) showing abnormalities consistent with active PTB [5]. Meanwhile, for extra-pulmonary TB (EPTB), the diagnosis is based on at least one specimen with confirmed *M. tuberculosis* or histological, or clinical evidence consistent with active EPTB, followed by a clinician's decision to treat with TB chemotherapy [5].

CXR implies radiation exposure and has low specificity [6], while SSM lacks sensitivity since the threshold of TB detection is high [7]. Furthermore, SSM can only detect TB in patients in advanced stage; it is less sensitive in Human Immunodeficiency Virus (HIV)-infected individuals, and does not differentiate between MTB and non-tuberculosis mycobacterial (NTM) infection [8,9]. Culture currently is the reference standard [5]. It needs only 10–100 MTB bacilli to establish a diagnosis, however, the laboratory turn-around time is long (two to eight weeks) [1].

Other sputum-dependent tests have been developed more recently, i.e. nucleic acid amplification techniques which allow for fast identification of MTB as well as rapid assessment of rifampicin susceptibility (such as GeneXpert® MTB/RIF (Cepheid, USA) and Truenat MTB assays® (Molbio Diagnostics, India)), LPAs (Hain Lifescience, Germany

* Corresponding author at: Infectious Diseases Service & Tuberculosis Unit, University Medical Center Groningen, University of Groningen, POBox 30 001, 9700 RB Groningen, the Netherlands.

E-mail address: t.s.van.der.werf@umcg.nl (T.S. van der Werf).

Research in context

Evidence before this study

The standard tests to diagnose tuberculosis (TB) such as culture and PCR, have high sensitivity, while sputum smear microscopy has only moderate sensitivity. These tests are not point-of-care, are sputum-dependent, and prone to laboratory fallacy, and need trained laboratory personnel. There is increasing evidence that analysis of exhaled breath could be used to diagnose TB. There are two approaches to analyze breath specimens - chemical and physical techniques. Examples of chemical techniques are gas chromatography combined with mass spectrometry (GC/MS), immunosensor and bio-optical technology, and electronic-nose with sensors that identify chemical interactions between VOCs and the sensor surfaces. The electronic nose has been used for diagnosis of various diseases, *i.e.* asthma, Chronic Obstructive Pulmonary Disease (COPD), urinary tract infection, lung cancer, and brain cancer. There has not been a systematic review regarding the accuracy of electronic-nose equipment in diagnosing TB from patients' breath. We explored the whole array of devices of breath tests, including GC/MS, because there has been a rapid development in this field since the previous review.

Added value of this study

This study provides evidence on the sensitivity and specificity of breath test in diagnosing TB.

Implications of all the available evidence

These low to moderate strength of evidence studies show that an electronic nose had high sensitivity and specificity, while other devices had moderate to high sensitivity and low to moderate specificity. Further improvements are needed before breath test devices can be introduced into the routine TB diagnostic work flow.

and Nipro, Japan) to detect TB and resistance to first and second line anti-TB drugs, and TB LAMP (Eiken, Japan) for detection of TB [1]. However, LPAs and TB LAMP are not universally available or affordable, and even though Xpert MTB/RIF is provided at reduced price for low-middle income countries [1], it is not portable and needs stable electricity supply, limiting its use in highly burdened, remote settings in low- and middle income countries.

One third of TB suspects have difficulty to collect an adequate quality sputum sample [10]. Therefore, a non-sputum based test would be a tremendous asset. Several non-sputum based tests are in development, such as urinary lipoarabinomannan for patients with TB-HIV coinfection, computer-aided detection systems [1], pediatric stool processing prior to Xpert [11], blood host marker, skin patches, and breath tests [12].

Breath tests have several advantages - non-invasive, potentially point-of-care, easy-to-perform, fast, and convenient [13]. Infections change the host metabolism producing distinct volatile organic compounds (VOCs) [14], and *Mycobacterium tuberculosis* (MTB) produces several VOCs [15–17], which can be detected from the breath. There are two approaches to analyze breath specimens - chemical and physical techniques. Examples of chemical techniques are gas chromatography combined with mass spectrometry (GC/MS), immunosensor and bio-optical technology, and electronic-nose with sensors that identify chemical interactions between VOCs and the sensor surfaces [18–23]. Physical techniques measure a physical property of the molecule, such as Field Asymmetric Ion Mobility Spectrometry (FAIMS) that measures movement of ionised molecules

of breath [24]. GC/MS requires complex equipment, operation skills, and a well-conditioned environment, especially to record concentration differences of VOCs specific for TB [15–17,25,26], and different studies report different VOCs [17,25–27]. The electronic-nose has an array of sensors that identifies a pattern of VOCs without considering the specific composition of VOCs.

There has not been a systematic review regarding the accuracy of electronic-nose equipment in diagnosing TB from patients' breath. We also explored other types or devices of breath tests, such as GC/MS, because there has been a rapid development in this field since the previous review [13].

To examine the accuracy of electronic-nose and other devices in diagnosing TB from patients' breath, we reviewed the diagnostic test accuracy (DTA) studies that assessed the sensitivity and specificity of electronic-nose and other devices in diagnosing TB in patients with TB or suspected of having TB.

2. Methods

2.1. Protocol and registration

Methods of the analysis and inclusion criteria were documented in a protocol, and registered in PROSPERO (CRD42019132895).

2.2. Eligibility criteria

We included the DTA studies with participants of any age with TB or suspicion of TB. Diagnosis of TB was based on the WHO guideline [5]. The index test was a breath test, the comparator being sputum smear microscopy, chest radiography, culture, Gene Xpert, pleural biopsy, or a Composite Reference Standard (CRS).

2.3. Information sources

Studies were identified by searching electronic databases, *i.e.* Pubmed (1946–present), Embase (1946–present), Web of Science (1946–present), Medline (1946–present), clinicaltrials.gov, and WHO International Clinical Trials Registry Platform (ICTRP). Cochrane database was also reviewed. No limits were applied for language. The last search was run on 6 January 2019. A.M.S. and D.D.P. conducted the search. We also attempted to acquire any missing information on results from the investigator of articles included in this review, *i.e.* Marcel Bruins [22,23].

2.4. Search strategy and selection criteria

The PRISMA statement was followed for the systematic search [28]. The search strategies used terms such as “tuberculosis”, “diagnosis”, “breath test”, “electronic nose”, and “volatile organic compound”. Appendix 1 shows a detailed description of the search strategies. Studies were included if the subject was human, and were published from 1946 to present (6 January 2019). Abstracts of conferences were included. We excluded unpublished data, duplicate studies, reviews, and case reports. In addition, we hand-searched related articles from the reference lists, and trials registries, *i.e.* clinicaltrials.gov and WHO ICTRP.

2.5. Study selection

Eligibility assessment of titles and abstracts was performed independently by 2 investigators (A.M.S. and D.D.P.) based on the PICOS criteria (Population = patients with TB or suspected with TB; Index test = breath test by electronic nose or other devices; Comparator = other tests for TB; Outcomes = sensitivity, specificity based on culture or WHO guidelines for TB diagnosis if culture result was not available; Study design: diagnostic test accuracy). Disagreements between investigators were resolved by consensus, or involving a third investigator (T.S.W.) when consensus was not reached. Full-text articles were then

reviewed by A.M.S. and D.D.P. with a critical appraisal sheet taken from Joanna Briggs Institute Reviewers' Manual 2015 [29] which was developed based on the Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-2 approach [30]. Resolving disagreements was conducted in the same manner. We then analysed the outcomes from the remaining relevant research articles.

2.6. Data collection process

A data extraction sheet was developed based on Joanna Briggs Institute Reviewers' Manual 2015 [29]. We pilot-tested it on ten randomly-selected included studies, and refined it accordingly. The two investigators (A.M.S. and D.D.P.) extracted the data from included studies. Disagreements were resolved by discussion between the two investigators; if no agreement was reached, a third investigator (T.S.W.) was involved.

2.7. Data items

Information was extracted from each included study on: (1) characteristics of study participants (demographic description, *i.e.* age, sex, inclusion and exclusion criteria; participant groups); (2) type of intervention (breath test with electronic nose or other devices *versus* other tests for TB); (3) type of outcome measures (sensitivity, specificity); (4) funding sources.

2.8. Risk of bias in individual studies

An assessment of risk of bias was conducted with the modified version of the QUADAS tool for Cochrane reviews [31] in a blind manner. Assessments were completed independently by two investigators (A.M.S. and D.D.P.), and disagreements were resolved by consensus, or involving a third investigator (T.S.W.) when consensus was not reached.

2.9. Summary measures and data analysis

The primary outcome measures were sensitivity and specificity of breath test to diagnose TB using culture, or the criteria mentioned in the WHO guideline for TB diagnosis if culture result was not available, as the reference [5]. The secondary outcome measures were Positive Likelihood Ratio (PLR), Negative Likelihood Ratio (NLR), and Diagnostic Odds Ratio (DOR). Sensitivity was defined as probability that a test result would be positive when the disease is present (true positive rate), specificity was the probability that a test result would be negative when the disease was absent (true negative rate). PLR described how many times positive index test results were more likely in the diseased group compared to control, NLR described how many times negative index test results were less likely in the diseased group compared to control, and DOR described how many times the odds were higher of obtaining a test positive result in a diseased compared to control [32]. We calculated pooled sensitivity, specificity, PLR, NLR, and DOR with 95% confidence intervals. We assessed heterogeneity and inconsistency for the pooled estimates by the Cochrane Q and I^2 statistic [33], and if there was heterogeneity, we conducted subgroup analysis to find the sources [32]. We used bivariate hierarchical random effects logistic regression models methods for meta-analysis of DTA to generate the Summary Receiving Operator Characteristics (SROC) curve with point estimates and their 95% confidence region from all study sources. All analyses were performed with STATA (version 15 SE; Stata Corporation, College Station, TX, USA). Stata commands of "metandi" and "metandiplot" are used to facilitate calculating the summary of the fitted point and the curve with 95% prediction region [34].

2.10. Risk of bias across studies

Risk of bias was assessed according to QUADAS items [35]. We had earlier planned to make a funnel plot to assess the publication bias within studies, but we decided not to, as it was not recommended by the Cochrane Handbook considering that it might mislead the interpretation of the publication bias assessment [32]. We also looked for missing data from the included studies that could produce selective reporting bias.

3. Results

3.1. Study selection

A total of 14 studies were identified for inclusion in the review. The search of Pubmed, Embase, Web of Science, Medline, clinicaltrials.gov, ICTRP, and Cochrane databases provided a total of 391 studies. After adjusting for duplicates, 239 remained. Of these, 225 studies were discarded because these papers did not meet the criteria. The full text of the remaining 20 studies was examined in more detail. Fourteen studies met the inclusion criteria and were included in the systematic review. By checking the references, relevant papers and searching for studies that have cited these papers, no additional studies that met the criteria for inclusion were identified.

3.2. Study characteristics

All fourteen studies finally selected for the review were published in English. Table 1 shows characteristics of the studies. Number of subjects analysed ranged from 40 to 251 subjects per study. Only three studies included children (aged ≥ 13 years); the others included only adults. All studies were carried out prospectively, 9 out of 14 studies (64.3%) employed a case-control study design (for studies that consisted of training and validation phases, we only assessed the study design in the validation phase as it reflected the performance of the device). Most of them were conducted in hospitals and lung clinics; six studies were conducted in Africa, two in Asia, two in South America, one in Europe, two in transcontinental countries (Egypt and Russia), and one in both Asia and Europe. (See Fig. 1.)

3.3. Risk of bias

Fig. 2 summarizes the risks of bias and applicability concerns of studies. Only five studies specified that they involved consecutive series of patients; the other studies did not describe the process of sampling and used case-control study design. However, due to the nature of these pilot studies, case-control is commonly used, thus we assessed the risk of bias in terms of 'patient selection' as medium. All studies reported a clear definition of a positive test using a reference diagnostic test, but only six ensured adequate blinding, six did not mention concealment, and the subjects in the validation group in two studies overlapped with the training group (medium to high risk of bias concerning 'index test'). In terms of 'flow and timing', only six out of 14 studies mentioned the exact time interval between the index test and reference test, but the sputum sample could be collected anytime even on the same day of the assessment using the index test (low risk of bias). All of included studies had complete verification for positive reference test, however not all subjects in one study used the same reference test. Eight studies used the gold standard, which is culture, while other 6 studies also used other examinations beside culture (low to medium risk of bias for reference standard). In summary, 12 studies were considered having a medium risk of bias or raised applicability concerns for at least one item of the QUADAS-2 tool.

3.4. Test accuracy

The total number of subjects included in the analysis of sensitivity and specificity of breath tests (validation phase) amounted to

Table 1
Study characteristics and sensitivity-specificity for TB diagnosis by breath test.

Source	Number of patients included in analysis	Type of study subjects	Clinical setting and country	Type of study	Index test	Reference test	Collection of air samples	Type of sensors	Exclusion	Analysis	Funding sources	Rounded sensitivity, % (95% CI)	Rounded specificity, % (95% CI)
Bruins et al. [22]	PoP phase: 30 (15 PTB, 15 healthy); VS phase: 194 (34 TB, 114 non-TB, 46 healthy controls)	PTB patients, PTB suspects, and healthy controls; Children and adults (≥ 15 years)	Hospital and chest clinics, Bangladesh	PoP: case-control, retrospective; VS: cross-sectional, prospective. Consecutive enrolment: unclear.	Electronic nose	Culture	Separate bags	12 metal-oxide sensors	Invalid measurement, outliers data	Artificial Neural Networks (ANN), participants for training and validation were overlap. Leave-one-out cross validation (LOO-CV) approach.	Partially funded by BD	75 (59–89)	87 (81–92)
Nakhleh et al. [19]	198 (64 TB, 67 non-TB, 67 healthy controls)	PTB patients, control: 1) healthy volunteers, and 2) TB-negative subjects; Adults.	Three sites in Cape Town, South Africa	Case-control, prospective. Consecutive enrolment: unclear.	Electronic nose	Culture	Separate tubes	12 nanomaterial-based sensors	Not mentioned	Principal Component Analysis (PCA); independent subjects in the validation phase.	Organics Ltd	90 (68–99)	93 (80–98)
Mohamed et al. [35]	123 (67 PTB, 56 healthy controls)	Newly diagnosed, treatment-naive PTB patients vs healthy controls; Adults (≥ 18 years)	Chest hospitals, Egypt	Case-control, prospective. Consecutive enrolment: yes.	Electronic nose	SSM and liquid culture on modified LJ medium	Separate bag/sacks	10 metal-oxide sensors	Previous history of TB, coexisting medical or surgical illnesses, chronic bronchitis, bronchiectasis, lung abscess, default TB, relapse TB, latent TB	PCA and ANN, different participants for training (60%), testing (25%), and validation (15%).	Science and Technology Development Fund, Egyptian Ministry of Higher Education and Scientific Research, Cairo, Egypt	100 (95–100)	98 (90–100)
Zetola et al. [21]	71 (51 PTB, 20 healthy controls)	Newly diagnosed PTB patients and healthy controls; Adults (≥ 21 years)	Hospital, Botswana	Case-control, prospective. Consecutive enrolment: yes.	Electronic nose	MGIT culture	Separate bag/sacks	8 metallo-porphyrin coated quartz microbalance (QMB) gas sensors	Smokers	PCA and <i>k</i> -Nearest Neighbours, participants for training and validation were overlap. LOO-CV approach.	NIH/NIAID grants	94 (84–99)	90 (68–99)
Teixeira et al. [23]	106 (41 TB, 19 non-TB, 46 healthy controls)	PTB suspects, asthma or COPD patients and healthy controls; Adults (> 18 years)	National Reference Centre for pulmonary diseases, Paraguay	Case-control, prospective. Consecutive enrolment: unclear.	Electronic nose	Culture, or other strong supporting evidence when the culture result is negative	Inside the device	3 metal-oxide sensors	TB suspects not able to expectorate sputum, with respiratory failure or having received TB treatment in the past 6	ANN, participants for training and validation were overlap.	Sonnevanck Suppletie fonds, Harderwijk, The Netherlands	88 (74–96)	92 (83–97)

Table 1 (continued)

Source	Number of patients included in analysis	Type of study subjects	Clinical setting and country	Type of study	Index test	Reference test	Collection of air samples	Type of sensors	Exclusion	Analysis	Funding sources	Rounded sensitivity, % (95% CI)	Rounded specificity, % (95% CI)
Poli et al. [39]	84 (34 TPE/TB Pleural Effusion, 50 non-TPE)	TPE suspects, HIV-negative; Children and adults (≥ 15 years)	Venezuela	Cross-sectional, prospective. Consecutive enrolment: unclear.	Electronic nose	Histopathology of the pleural biopsy	Inside the device	3 metal-oxide sensors	months. Poor general condition, difficulty to breath, no follow-up.	ANN, participants for training and validation were overlap.	Servicio Autónomo Instituto de Biomedicina "Dr. Jacinto Convit", Venezuela	88 (73–97)	94 (83–99)
McNerney et al. [20]	60 (31 PTB, 29 non-TB)	Patients with respiratory complaints; Adults (≥ 17 years)	Outpatient clinic of a hospital, Ethiopia	Cross-sectional, prospective. Consecutive enrolment: unclear.	RBS Breathalyzer (immunosensor and bio-optical technology)	Combination of clinical assessment, SSM, and CXR	Separate plastic tubes	Immunosensor (coated prism with TB antibodies labeled with a fluorescent dye)	Not mentioned	Evanescence wave fluorimetry	TARGETS Communicable Diseases Research Consortium funded by Department for International Development, UK	74 (55–88)	79 (60–92)
van Beek et al. [36]	226 (90 TB, 136 healthy controls)	Culture-confirmed TB patients and healthy controls; Adults (≥ 18 years)	Lung Hospital and construction company, Vietnam	Case-control, prospective. Consecutive enrolment: yes.	eNO (Nitric Oxide) analyser	LJ Culture	Inside the device	N.A.	Negative culture, difficulty to breath, no CXR, HIV(+), current TB treatment.	The concentration of NO was noted in ppb.	Not mentioned	78 (68–86)	62 (47–75)
Phillips et al. [18]	251 (130 PTB, 121 healthy controls)	PTB patients and healthy controls; Children and adults (≥ 13 years)	Hospitals and Health Sciences Institute in Phillipines, United Kingdom, and India	Case-control, prospective. Consecutive enrolment: unclear.	The BreathLink system (portable gas chromatograph coupled to a surface acoustic wave detector)	Culture and/or SSM and/or CXR	Inside the device	N.A.	On anti-TB therapy, received anti-TB therapy in the last six months, technically unsatisfactory breath test samples	Multiple Monte Carlo simulations; Weighted Digital Analysis (WDA) multivariate predictive algorithm.	Menssana Research, Inc., United States Air Force	72 (63–79)	72 (63–80)
Kolk et al. [40]	Training: 100 (50 culture-proven TB, 50 culture-negative non-TB); Validation: 71 (21 PTB, 50 non-TB)	PTB suspects; Adults (> 18 years)	TB clinics, South Africa	Cross-sectional, prospective. Consecutive enrolment: yes.	GC-TOF-MS	MGIT Culture	Separate bags	N.A.	On anti-TB therapy	Support vector machine; LOO-CV approach; independent subjects in the validation phase.	UBS (Union Bank of Switzerland) Optimus Foundation and the Foundation for Innovative New Diagnostics (FIND)	62 (38–82)	84 (71–93)
Maiga et al. [37]	56 (20 PTB, 36 healthy controls)	Newly diagnosed PTB patients and healthy controls; Adults (≥ 18 years)	Health centers and Hospital, Mali	Case-control, prospective. Consecutive enrolment: yes.	BreathTek™ urease breath test	MGIT and 7H11 culture	Inside the device	N.A.	Use of any antibiotic, proton pump inhibitors or bismuth containing preparations < 15 days before study; HIV positivity; pregnancy	Measurement of the ratio of $^{13}\text{CO}_2$ to $^{12}\text{CO}_2$ in a breath sample.	NIH/NIAID grants, the Howard Hughes Medical Institute, the University of Sciences, Techniques, Bamako	70 (46–88)	11 (3–26)

Sahota et al. [24]	40 (21 PTB/EPTB, 2 non-TB, 17 healthy controls)	Suspect PTB and EPTB, TB patients within 1 week of TB treatment; Adults (≥ 18 years)	TB clinics or in-patient admission in a teaching hospital, United Kingdom	Case-control, prospective. Consecutive; unclear.	FAIMS (Field Asymmetric Ion Mobility Spectrometry)	Culture, histology or radiology	Separate bags	N.A.	Not mentioned	Gaussian Process Classifier; LOO-CV method.	Medical Research Council, UK	81 (58–95) (54–94)	79 (36–83)
Beccaria et al. [41]	50 (32 PTB, 18 non-TB)	PTB suspects, including smokers and patients with HIV (+); Adults (≥18 years)	Wits Reproductive Health and HIV Institute, South Africa	Cross-sectional, prospective. Consecutive; unclear.	GC-TOF-MS	Gene Xpert and MGIT culture	Separate bags	N.A.	Not mentioned	Random Forest, linear Support Vector Machines, Partial Least-Squares Discriminant Analysis; Data was divided into training (60%) and validation (40%) phases; LOO-CV method.	Burroughs Wellcome Fund	100 (89–100)	61 (36–83)
Morozov et al. [58]	55 (42 PTB, 13 healthy controls)	Recent and chronic TB; Adults (≥ 18 years)	Russia	Case-control, prospective. Consecutive; unclear.	Rapid ultrasensi-tive immuno-chemistry method	Clinical and CXR, culture and/or PCR or positive responses to anti-TB therapy	Separate bags	N.A.	Not mentioned	Immunoassay	Russian Science Foundation	72 (53–86)	58 (28–85)

N.A.: Not Applicable.

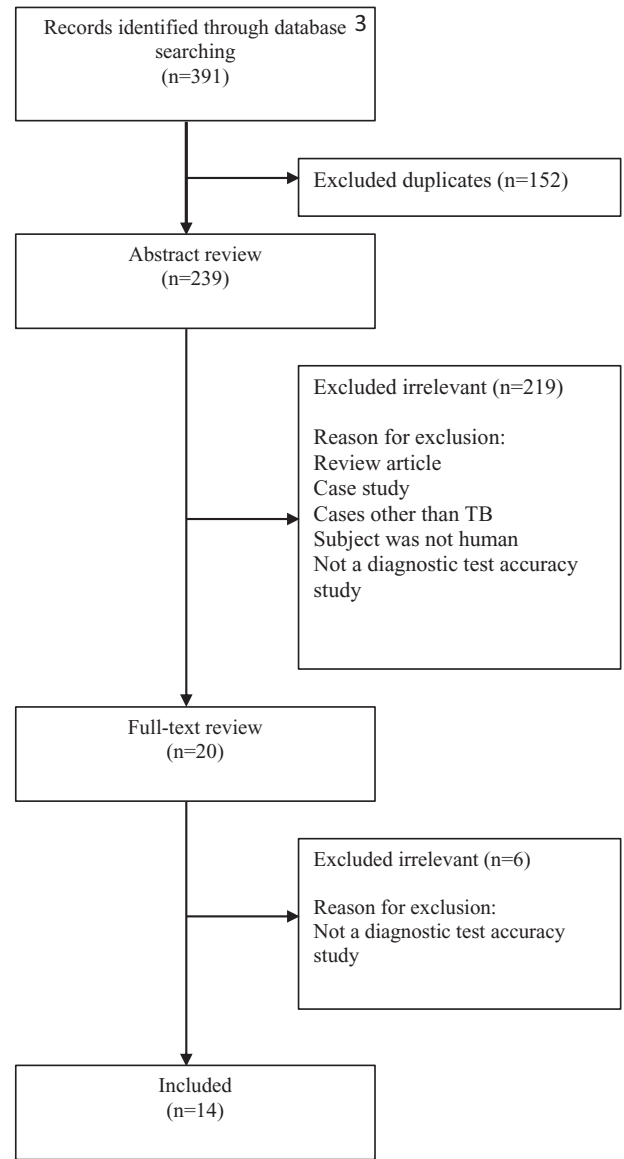
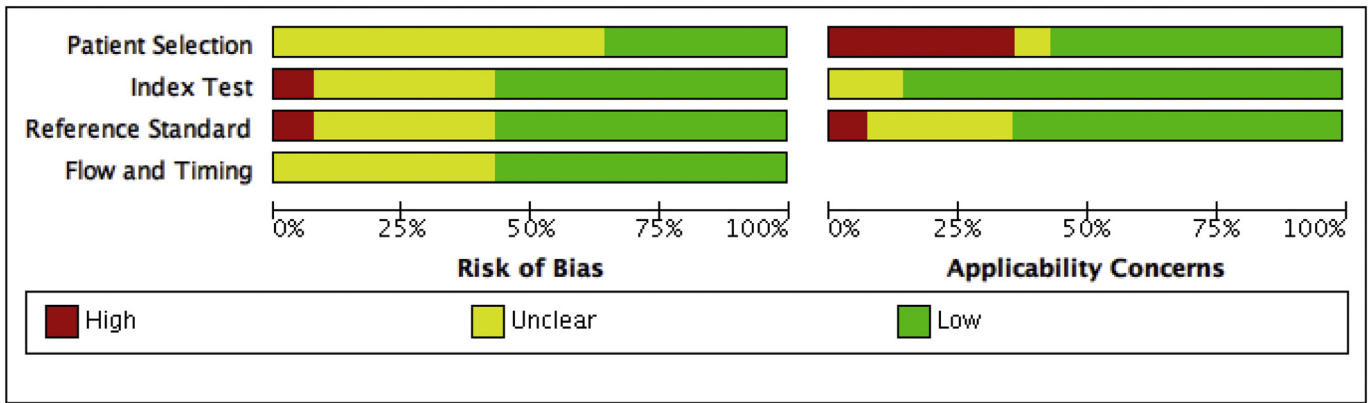


Fig. 1. Process of studies selection throughout the review.

1715 subjects (Table 1). The sensitivity and specificity of each test is shown in Table 1. As the number of breath test devices other than electronic-nose was insufficient for grouping, the statistical analysis was only calculated for the electronic-nose. The Q and I^2 of electronic-nose in diagnosing TB were $p = 0.23$ and 0 (95% uncertainty intervals:0–100), respectively. The pooled sensitivity and specificity of electronic-nose in diagnosing TB were 0.92 (95% CI:0.82–0.97) and 0.93 (95% CI:0.88–0.96), respectively (Fig. 3). The PLR, NLR, and DOR were 13.4 (95% CI:7.2–24.9), 0.08 (95% CI:0.03–0.20), and 162 (95% CI:41–634), respectively. Fig. 4 shows the summary receiving operating characteristics for these studies. The area under the curve (AUC) of electronic-nose was 0.97 (95% CI:0.19–1.00). Meanwhile, the sensitivity and specificity of breath test devices other than electronic nose ranged from 0.62 to 1.00 , and 0.11 to 0.84 , respectively.

4. Discussion

The pooled sensitivity and specificity of electronic-nose in diagnosing TB, either pulmonary or extrapulmonary, were high. There is no



	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Beccaria 2018	+	+	?	?	+	+	+
Bruins 2013	?	+	+	+	+	+	+
Kolk 2012	+	+	+	+	+	+	+
Maiga 2016	+	-	-	+	-	+	-
McNerney 2010	?	?	?	+	?	?	?
Mohammed 2017	?	+	+	+	-	+	+
Morozov 2018	?	+	+	?	+	+	+
Nakhleh 2014	?	+	+	+	+	+	?
Phillips 2012	?	?	?	?	-	+	?
Poli 2018	?	?	+	+	+	?	?
Sahota 2016	?	?	?	?	+	+	+
Teixeira 2017	?	+	+	?	+	+	+
van Beek 2011	+	?	?	?	-	+	+
Zetola 2017	+	+	+	+	-	+	+

Fig. 2. Summary of QUADAS-2 assessments of included studies. Patient selection: describes methods of patient selection; index text: describes the index test and how it was conducted and interpreted; reference standard: describes the reference standard (gold standard test) and how it was conducted and interpreted; flow and timing: describes any patients who did not receive the index tests or reference standard or who were excluded from the 2 × 2 table, and describes the interval and any interventions between index tests and the reference standard [30].

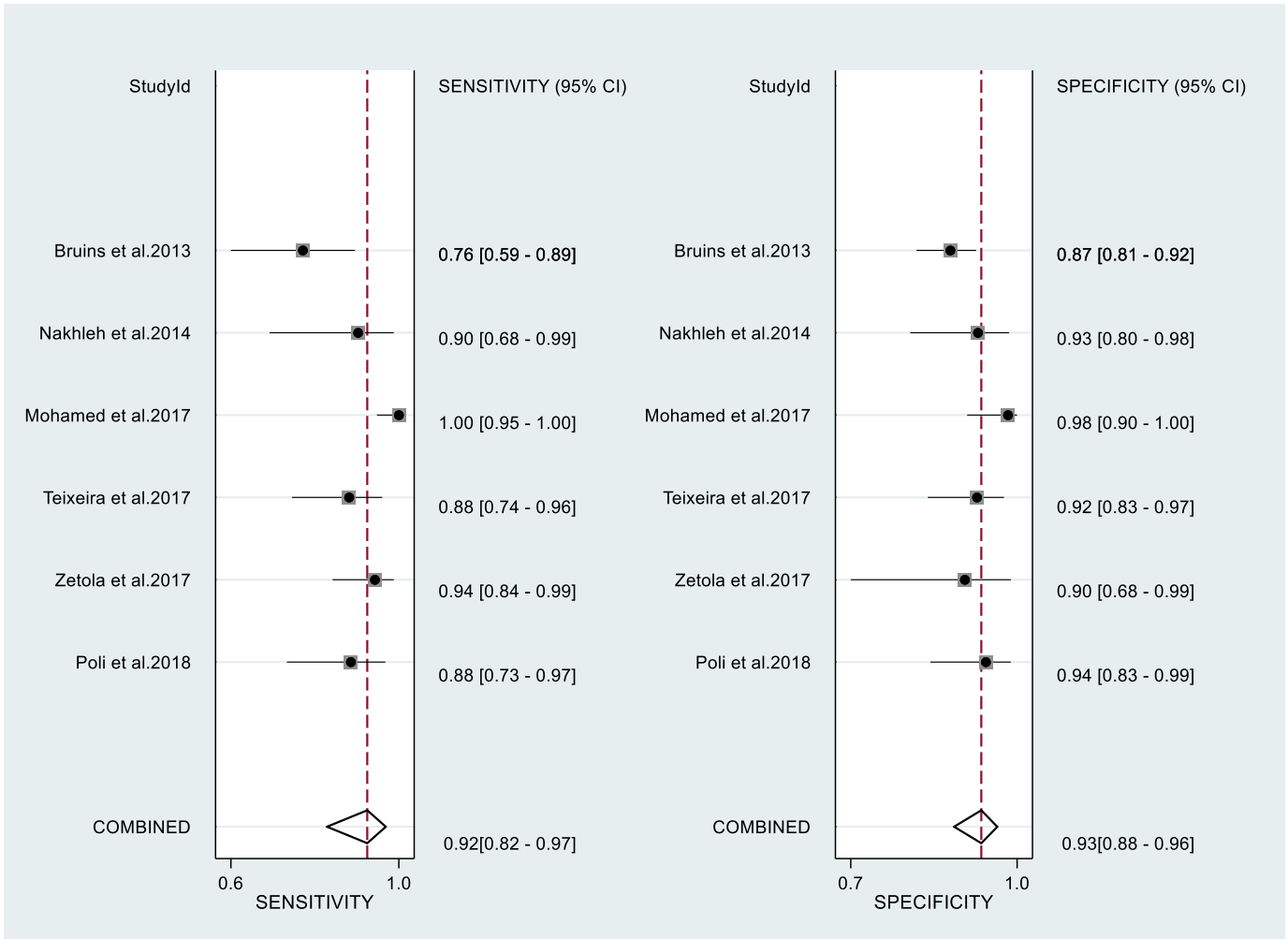


Fig. 3. Paired forest plots of pooled sensitivity and specificity of electronic nose in diagnosing tuberculosis.

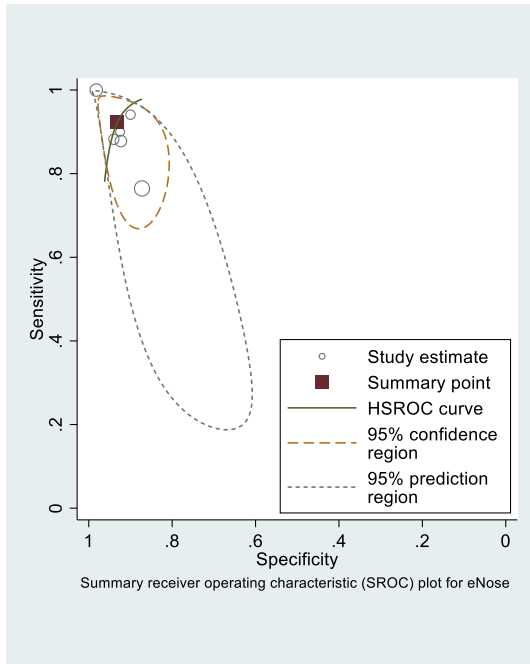


Fig. 4. Summary receiver operating characteristic (SROC) of electronic nose in diagnosing tuberculosis.

heterogeneity, indicating that the variability among studies that diagnose TB with electronic-nose was due to chance. Other devices had moderate to high sensitivity and low to moderate specificity. The strength of evidence within all 14 studies was low to moderate.

Five studies only included subjects representing two extreme sides of the clinical spectrum of TB disease (healthy controls without any TB symptoms, and symptomatic, treatment-naïve, smear positive pulmonary TB cases) [18,21,36–38]. Meanwhile, different stages of TB disease are characterized by dynamic metabolic changes within pathogen and host and their interactions, thus different stages of TB disease may generate different breath test results [21]. In addition, around 20–25% of the world population have latent TB [39], and the event of relapsed TB in the high-TB burden countries is not uncommon. Therefore, these studies had limited generalisability and clinical applicability, and might also produce bias as the results might be over-estimated, thus these assumptions might be not valid for our review. Ideally, besides being able to identify the VOCs produced by either the host or the bacteria, the breath test's sensors should ideally be also able to identify different signals in different stages of TB disease. Only seven studies included other pulmonary infections as controls [19,20,22,24,40–42], while a diagnostic test is typically needed in this population, to identify the TB afflicted among the suspects.

Breath prints are influenced by various factors, which are divided into physiological factors (age, sex, food, beverages), pathological and disease-related conditions (smoking, comorbidities, medication), and sampling-related issues (bias with VOCs in the environment) [43]. Older age changed breath prints in patients with lung cancer [44], males had higher level of isoprene than females [45], several beverages and poultry meat could be

identified with an electronic-nose [46–50], and patients with high body mass index (BMI) had more false-positive test results than patients with normal or low BMI [23]. Smoking increased the levels of several VOCs, *i.e.* benzene and pentane [51], while these two VOC markers were observed in the culture and breath of TB patients [17]. Apart from those mentioned factors, VOCs might be also influenced by genetic profiles [23]. Thus, in the analysis of the breath test, confounding or modifying effects from these factors should always be considered. In only four of the reviewed studies, at least one of these factors were taken into account [18,21,23,37]. Furthermore, there were only three studies including children in their study population, and children were also not exclusively enrolled [18,22,40], while age may influence the breath test result [44].

In several electronic-nose studies, the exhaled breath was collected into a bag/tube with the specimen investigated later [19,21,22,36]. Using bags for collection may be more useful in the flow of healthcare in a clinic, however, it might change the composition of the sample because the VOCs bind with the sample-bag material, and the bag material releases background emission of several pollutants [52]. Samples should be also examined in six hours from the collection time [52], thus it adds a burden and complicates the diagnostic process. As far as the portable sensing technology is not yet specific enough for VOCs associated with TB, we recommend to avoid using bags for sample collection. In addition, although the sensors in an electronic-nose can capture abundant amounts of information, most of them do not have sensitivity as high as other techniques [53], thus to improve the sensitivity, pre-concentration techniques, such as absorption of breath samples onto a solid phase, needs to be performed. Several studies in this review used micro-extraction (SPME) or Tenax as the pre-concentration technique [19,23,40], while other studies did not mention it in their Methods section.

To develop a robust algorithm to enable real-time measurements, an electronic-nose needs a prior knowledge of the data (training phase). Ideally, to avoid a fallacy in breath test analysis due to confounding variables, artefacts, and statistical errors, a training phase should include a large number of study subjects, so that the number of subjects will be bigger than the number of breath test variables/features, and the data can be divided into a training and validation set [54]. However, if the population is small, a leave-one-out cross validation is usually used in the training phase, *i.e.* only one subject is used for validation purposes, while the others are used for training [54]. The device performance can be assessed more objectively if the next stage (validation phase) is carried out in subjects who are not involved in the training phase. In only three out of six studies with electronic-nose, the validation phase used independent subjects from the training phase [19,21,36]. In addition, a large number of subjects in the training phase is of great importance to expose the device with large diversity in breath patterns from true-positive study subjects. Only then, the device can recognize breath samples in the validation phase with high accuracy. For an electronic-nose device, to match a pattern correctly, the pattern must be known before.

The diagnostic accuracy of an electronic-nose can only be determined reliably if the TB status can be classified with high certainty since the analysis of breath test completely depends on the correct classification of the study samples in the training phase. Therefore, a reference test should be as 'gold' as possible, approaching the highest sensitivity and specificity. When the gold standard is imperfect, the positive test results from an investigated diagnostic tool may be considered as false-positives, thus the accuracy of this diagnostic tool will be underestimated. In this review, eight studies used only culture as the reference test for diagnosing pulmonary TB, while the other six studies used culture or other supporting data when the culture result was negative. In our opinion, the use of other supporting data to improve the accuracy of culture is justifiable because culture is also prone to laboratory error. In addition, follow-up should be employed to ensure that the diagnosis made was correct. A previous study has shown that follow-up could help to cover the weakness of culture and improve its accuracy [55].

For other breath test devices, the selected ion flow tube-mass spectrometry (SIFT-MS) becomes the preferable option as it is faster and has

higher sensitivity than GC-MS. [56] Ideally, the chemical platform for chemical breath analysis would have online identification of specific analytes without the need for pre-concentration and should be able to analyze small volatiles (in the ppt-ppq range) [53]. Currently with GC-MS, samples still need to be pre-processed before testing. To validate the biomarkers that were found through GC-MS or other spectrometries, a large cohort study is needed. The use of oral urea administration in diagnosing TB may not be ideal because of low lung concentration, and is confounded by the presence of *H. pylori* in the gut [38]. However, the administration of urea by inhalation for diagnosing TB warrants investigation. Detection of antigen in cough was shown to have moderate sensitivity and specificity that was suspected as a result from the variation of antigen level in cough during the day, thus it was suggested to conduct the breathalyzer test several times in a day [20]. However, repeating a test several times in one day takes time and is therefore impractical. Furthermore, prior to sputum collection, patients needed to have nebulized therapy, thus limiting its use for a point-of-care device in remote areas. This device might also be projected as an add-on test for sputum smear microscopy, which would not provide an advantage over sputum-dependent sputum smear microscopy and culture.

The existing tests have high sensitivity, such as culture and PCR [1], or very high specificity, such as Gene Xpert [57], or moderate sensitivity, such as sputum smear microscopy [1,55]. However, these highly sensitive tests are not point-of-care, are sputum-dependent, prone to laboratory fallacy, and need trained laboratory personnel. With the high sensitivity, easy operation, point-of-care, and hand-held form, the breath test might be potentially used as a screening method rather than a diagnostic tool. Cases that are positive by breath test, could be referred for the tests which are more specific, probably more expensive, and laboratory-based, such as Gene Xpert. To date, there have been no studies that investigate the breath test as a screening tool.

A challenge for the breath test development in the future is to detect TB cases with negative sputum smear. Negative sputum smear was shown to have correlation with a misdiagnosis of TB [55]. Another challenge is to investigate whether the breath test can accelerate the diagnostic process and lower its cost. A point-of-care automated thermal desorption, gas chromatography and mass spectrometry (ATD-GC-SAW) costs \$20,000 [18], while an electronic-nose based on metal-oxide sensors has an advantage of mass production at low cost. Analysis of a breath VOC sample with a point-of-care ATD-GC-SAW takes six min [18], whilst currently the analysis time with electronic-nose still takes several days or weeks as the data needs to be sent to the device producer before obtaining the diagnosis. Once the pattern recognition technique algorithm is fully trained, the results should be generated automatically in seconds. For further optimization of the sensors, specific VOCs in different stages of the disease need to be determined. The use of highly selective sensors that target these VOCs may increase the sensitivity and specificity. Finally, as the group of patients that need non-sputum-based tests most are children, the sensitivity and specificity of breath test exclusively in children needs to be extensively investigated.

Our review has some limitations. Only a limited number of studies concerning breath test using electronic-nose for diagnosing TB were available. Within the identified studies with breath test devices other than electronic-nose, we were not able to conduct meta-analysis by the type of devices because studies were few. Several studies only included treatment-naïve, smear positive pulmonary TB cases, thus hindering conclusions on the sensitivity and specificity of breath test for all stages of TB, and this may introduce overestimation of sensitivity and specificity. Nine out of 14 studies had an unclear risk of bias concerning patient selection, and five had an applicability concern on one or two domains.

5. Conclusions

These low to moderate strength of evidence studies show that breath tests have potential to screen for TB. However, it still needs

further improvement, with a more robust trained pattern recognition technique to give real-time measurements. An ideal breath test should be accurate, fast, easy-to-use, point-of-care, and distributed with low price, so that it can be used not only in clinics, but also for door-to-door screening in remote areas, where TB diagnosis still faces great challenges.

Author contributions

A. M. I. S.: literature search, figures, study design, data collection, data interpretation, writing, D.D.P.: literature search, study design,

data collection, writing, A.S.: data analysis, figures, writing, Y.M.: data interpretation, writing, TSvdW: study design, data interpretation, writing.

Declaration of Competing Interest

D.D.P., A.S., and Y.M. declare no conflict of interest. A.M.I.S. and T.S.v.d.W. report support from the e-Nose company (*i.e.*, support in kind, to conduct a study with the equipment of this company), outside the submitted work.

Appendix A

PUBMED

Period: all years to present (6 Jan 2019)

((("Tuberculosis"[Mesh] OR tuberculosis[tiab] OR mycobacterium[tiab]) OR tb[tiab] AND ("Diagnosis"[Mesh] OR diagnosis[tiab] OR diagnose[tiab] OR diagnostic[tiab]) AND ("Breath Tests"[Mesh] OR breath test[tiab] OR Breathalyzer Tests[tiab] OR breath examination[tiab] OR breath examinations[tiab] OR electronic nose[tiab] OR enose[tiab] OR volatile organic compound[tiab] OR volatil[tiab]))) AND "humans"[Filter])

:107 articles

EMBASE

Period: all years to present (6 Jan 2019)

#1 AND #2 AND #5 AND ([article]/lim OR [article in press]/lim OR [conference paper]/lim OR [erratum]/lim) AND [humans]/lim	#7
#1 AND #2 AND #5	71
#3 OR #4	#6
'electronic nose':ab,ti OR 'enose':ab,ti OR 'e-nose':ab,ti OR 'volatile organic compound':ab,ti OR 'volatile agent':ab,ti	358
('breath analysis':ab,ti OR 'breathing':ab,ti OR 'breath'/exp OR breath) AND examination*:ab,ti	#5
'diagnosis':ab,ti OR diagnose*:ab,ti OR 'diagnostic procedure':ab,ti	18,343
'tuberculosis':ab,ti OR 'mycobacterium':ab,ti OR tb:ab,ti	#4
	2,970
	#3
	15,395
	#2
	2,636,715
	#1
	257,918

:71 articles

WEB OF SCIENCE

Period: all years to present (1945-2019)

TOPIC:((Tuberculosis OR mycobacterium OR tb) AND (Diagnosis OR diagnoses OR diagnostic) AND ((breath test* OR breathalyzer test OR breath examination*) OR (electronic nose* OR bioelectronics nose* OR volatile organic compound* OR volatil*)))
 Refined by: DOCUMENT TYPES: (ARTICLE OR MEETING ABSTRACT OR EDITORIAL MATERIAL OR NOTE OR PROCEEDINGS PAPER)

Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, ESCI.

:136 articles (article=129, meeting abstract=4, proceeding paper=4, editorial material=3, note=1).

COCHRANE LIBRARY

Key words: Tuberculosis

:0 article (no review about breath test to diagnose tuberculosis)

MEDLINE

Database: Ovid MEDLINE(R) <1946 to January Week 1 2019>

Search Strategy:

-
- 1 TUBERCULOSIS/ (101029)
 - 2 Mycobacterium tuberculosis/ (46798)
 - 3 tuberculosis.tw. (159330)
 - 4 TB.tw. (36073)
 - 5 1 or 2 or 3 or 4 (209576)
 - 6 exp Breath Tests/ (14026)
 - 7 breath test\$.tw. (6553)
 - 8 Electronic Nose/ (377)
 - 9 e-nose.tw. (188)
 - 10 Volatile Organic Compounds/ (7460)
 - 11 Volatile Organic Compounds.tw. (5952)
 - 12 VOC.tw. (3022)
 - 13 6 or 7 or 8 or 9 or 10 or 11 or 12 (27537)
 - 14 5 and 13 (96)
 - 15 di.fs. (2407071)
 - 16 an.fs. (1886820)
 - 17 clinical trials as topic.sh. (185645)
 - 18 randomized controlled trial.pt. (473585)
 - 19 Cross-Sectional Studies/ (282312)
 - 20 cohort studies/ (232508)
 - 21 longitudinal studies/ (119723)
 - 22 follow-up studies/ (604739)
 - 23 prospective.ti,ab. (446876)
 - 24 retrospective.ti,ab. (383400)
 - 25 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 (5966282)
 - 26 14 and 25 (77)

:77 articles

References

- [1] World Health Organization. Global tuberculosis report 2018 [internet]. Geneva; [cited 2018 Sep 27]. Available from: <http://apps.who.int/iris/bitstream/handle/10665/274453/9789241565646-eng.pdf?ua=1>
- [2] Vynnycky E, Fine PE. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiol Infect* 1997 Oct; 119(2):183–201.
- [3] Cai J, Wang X, Ma A, Wang Q, Han X, Li Y. Factors associated with patient and provider delays for tuberculosis diagnosis and treatment in Asia: a systematic review and meta-analysis. *PLoS One* 2015 Mar;10(3):e0120088.
- [4] Sreeramareddy CT, Panduru KV, Menten J, Van den Ende J. Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature. *BMC Infect Dis* 2009 Jun;9:91.
- [5] World Health Organization. Treatment of tuberculosis guidelines 4th edition [internet]. [cited 2016 Apr 4]. Available from: <http://www.who.int/tb/publications/2010/9789241547833/en/>; 2010.

- [6] Davies PD, Pai M. The diagnosis and misdiagnosis of tuberculosis. *Int J Tuberc Lung Dis* 2008 Nov;12(11):1226–34.
- [7] Wu ZL, Wang AQ. Diagnostic yield of repeated smear microscopy examinations among patients suspected of pulmonary TB in Shandong province of China. *Int J Tuberc Lung Dis* 2000 Nov;4(11):1086–7.
- [8] Griffith DE, Brown-Elliott BA, Langsjoen B, Zhang Y, Pan X, Girard W, et al. Clinical and molecular analysis of macrolide resistance in *Mycobacterium avium* complex lung disease. *Am J Respir Crit Care Med* 2006 Oct;174(8):928–34.
- [9] Getahun H, Harrington M, O'Brien R, Nunn P. Diagnosis of smear-negative pulmonary tuberculosis in people with HIV infection or AIDS in resource-constrained settings: informing urgent policy changes. *Lancet (London, England)* 2007 Jun;369(9578):2042–9.
- [10] Parsons LM, Somoskövi K, Gutierrez C, Lee E, Paramasivan CN, Abimiku L, et al. Laboratory diagnosis of tuberculosis in resource-poor countries: challenges and opportunities. *Clin Microbiol Rev* 2011;24(2):314–50 Internet. cited 2019 Mar 25. Available from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3122496/pdf/zcm314.pdf>.
- [11] The Union. New TB test announced that could save thousands more children | the union conference [internet]. [cited 2019 Jan 9]. Available from: https://thehague.worldlunghealth.org/2018/10/new-tb-test-announced-that-could-save-millions-more-children/?utm_source=Conference+flash+%2A%2A+Imported+In+Global+List+And+Tagged&utm_campaign=761d81f09d-EMAIL_CAMPAIGN_2018_APRIL10_COPY_01&utm_medium=email&u; 2018.
- [12] England K. Treating patients, not disease: People-centered approach 7th annual TB symposium ministry of health of the Kyrgyz Republic and Médecins Sans Frontières [internet]. [cited 2019 Jan 9]. Available from: http://msf-tb-symposium.org/files/1715/2023/2923/1.1_Kathleen_England_TB_Diagnostics_ENG.pdf.
- [13] Cheepsattayakorn A, Cheepsattayakorn R. Breath tests in diagnosis of pulmonary tuberculosis. *Recent Pat Biotechnol* 2014;8(2):172–5 Available from: <http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1872-2083&volume=8&issue=2&spage=172>.
- [14] Bijland LR, Bomers MK, Smulders YM. Smelling the diagnosis a review on the use of scent in diagnosis disease. *Netherlands J Med* 2013;71:300–7.
- [15] Syhre M, Manning L, Phuanukoonnon S, Harino P, Chambers ST. The scent of *Mycobacterium tuberculosis*—part II breath. *Tuberculosis (Edinb)* 2009 Jul;89(4):263–6.
- [16] Syhre M, Chambers ST. The scent of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* 2008 Jul;88(4):317–23.
- [17] Phillips M, Cataneo RN, Condos R, Erickson GAR, Greenberg J, La Bombardi V, et al. Volatile biomarkers of pulmonary tuberculosis in the breath. *Tuberculosis* 2007 Jan;87(1):44–52.
- [18] Phillips M, Basa-Dalay V, Blais J, Bothamley G, Chaturvedi A, Modi KD, et al. Point-of-care breath test for biomarkers of active pulmonary tuberculosis. *Tuberculosis* 2012 Jul;92(4):314–20.
- [19] Nakhleh M, Jeres R, Gharra A, Binder A, Broza Y, Pascoe M, et al. Detecting active pulmonary tuberculosis with a breath test using nanomaterial-based sensors. *Eur Respir J* 2014;43(5):1522–5.
- [20] McNerney R, Wondafraash BA, Amena K, Tesfaye A, McCash EM, Murray NJ. Field test of a novel detection device for *Mycobacterium tuberculosis* antigen in cough. *BMC Infect Dis* 2010;10.
- [21] Zetola NM, Modongo C, Matsiri O, Tamuhla T, Mbongwe B, Matlhagela K, et al. Diagnosis of pulmonary tuberculosis and assessment of treatment response through analyses of volatile compound patterns in exhaled breath samples. *J Infect* 2017;74(4):367–76.
- [22] Bruins M, Rahim Z, Bos A, van de Sande WW, Endtz HP, van Belkum A. Diagnosis of active tuberculosis by e-nose analysis of exhaled air. *Tuberculosis* 2013 Mar;93(2):232–8.
- [23] Coronel Teixeira R, Rodríguez M, Jiménez de Romero N, Bruins M, Gómez R, Yntema JB, et al. The potential of a portable, point-of-care electronic nose to diagnose tuberculosis. *J Infect* 2017;75(5):441–7.
- [24] Sahota AS, Gowda R, Arasaradnam RP, Daulton E, Savage RS, Skinner JR, et al. A simple breath test for tuberculosis using ion mobility: a pilot study. *Tuberculosis* 2016;99:143–6 Internet. Available from <https://doi.org/10.1016/j.tube.2016.05.005>.
- [25] Mgode GF, Weetjens BJ, Nawrath T, Lazar D, Cox C, Jubitana M, et al. *Mycobacterium tuberculosis* volatiles for diagnosis of tuberculosis by *Cricetomys* rats. *Tuberculosis* 2012;92(6):535–42 Internet. Available from <https://doi.org/10.1016/j.tube.2012.07.006>.
- [26] Phillips M, Basa-Dalay V, Bothamley G, Cataneo RN, Lam PK, Natividad MP, et al. Breath biomarkers of active pulmonary tuberculosis. *Tuberculosis (Edinb)* 2010 Mar;90(2):145–51.
- [27] Beccaria M, Mellors TR, Petion JS, Rees CA, Nasir M, Systrom HK, et al. Preliminary investigation of human exhaled breath for tuberculosis diagnosis by multidimensional gas chromatography – Time of flight mass spectrometry and machine learning. *J Chromatogr B Anal Technol Biomed Life Sci* 2018;1074–5 Internet. (December 2017):46–50. Available from: <https://doi.org/10.1016/j.jchromb.2018.01.004>.
- [28] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, John PA. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions explanation and elaboration – Liberati et al_ 339 b2700 – BMJ.PDF. *Br Med J* 2009;b2700:339.
- [29] Campbell JM, Klugar M, Ding S, Carmody DP, Hakonsen SJ, Jadotte YT, et al. The systematic review of studies of diagnostic test accuracy [internet]. The Joanna Briggs institute reviewer's manual; 2015 [cited 2019 Jan 10]. Available from www.joannabriggs.org.
- [30] Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. Qadast-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011 Oct 18;155(8):529–36 Internet. cited 2019 Jan 10. Available from <http://annals.org/article.aspx?doi=https://doi.org/10.7326/0003-4819-155-8-201110180-00009>.
- [31] Reitsma H, Rutjes A, Whiting P, Vlassov V, Deeks J. Assessing methodological quality. In: Deeks JJ, Bossuyt PMGC, editors. *Cochrane handbook for systematic reviews of diagnostic test accuracy version 100* [internet]. The Cochrane Collaboration; 2009 cited 2019 Jan 10]. p. 0–27. Available from: https://methods.cochrane.org/sites/methods.cochrane.org.sdt/files/public/uploads/ch09_Oct09.pdf.
- [32] Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Chapter 10: analysing and presenting results. In: Deeks JJ, Bossuyt PMGC, editors. *Cochrane handbook for systematic reviews of diagnostic test accuracy version 10* [internet]. The Cochrane Collaboration; 2010 cited 2019 Jan 10]. p. 1–44. Available from: <http://srdta.cochrane.org/>.
- [33] Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005;58(9):882–93.
- [34] Harbord RM, Whiting P. Metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata J* 2009;9(2):211–29.
- [35] Reitsma J, Rutjes A, Whiting P, Vlassov V, Leeflang M, Deeks J. Chapter 9: assessing methodological quality. In: Deeks JJ, Bossuyt P, Gatsonis C, editors. *Cochrane handbook for systematic reviews of diagnostic test accuracy version 100* [internet]. The Cochrane Collaboration; 2009. p. 2009 [cited 2018 Dec 7]. Available from: <http://srdta.cochrane.org/>.
- [36] Mohamed EI, Mohamed MA, Moustafa MH, Abdel-Mageed SM, Moro AM, Baess AI, et al. Qualitative analysis of biological tuberculosis samples by an electronic nose-based artificial neural network. *Int J Tuberc Lung Dis* 2017;21(7):810–7.
- [37] Van Beek SC, Nhung NV, Sy DN, Sterk PJ, Tiemersma EW, Cobelens FGJ. Measurement of exhaled nitric oxide as a potential screening tool for pulmonary tuberculosis. *Int J Tuberc Lung Dis* 2011;15(2):185–91.
- [38] Maiga M, Cohen K, Baya B, Srikrishna G, Siddiqui S, Sanogo M, et al. Stool microbiome reveals diverse bacterial ureases as confounders of oral urea breath testing for helicobacter pylori and *Mycobacterium tuberculosis* in Bamako, Mali. *J Breath Res* 2016;10(3):036012.
- [39] Houben RMGJ, Dodd PJ. The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. Metcalfe JZ, editor. *PLOS Med* [Internet] 2016. <https://doi.org/10.1371/journal.pmed.1002152> Oct 25 [cited 2019 Mar 7];13(10):e1002152. Available from: .
- [40] Poli F, Quesada L, Poli S, Garcia D, Ippoliti G, Salas E, et al. Use of an electronic nose as a novel screening method for the diagnosis of tuberculous pleural effusion; 2018; 5543.
- [41] Kolk AHJ, Van Berkel JBN, Claessens MM, Walters E, Kuijper S, Dallinga JW, et al. Breath analysis as a potential diagnostic tool for tuberculosis. *Int J Tuberc Lung Dis* 2012;16(6):777–82.
- [42] Beccaria M, Bobak C, Maitshotlo B, Mellors TR, Purcaro G, Franchina FA, et al. Exhaled human breath analysis in active pulmonary tuberculosis diagnostics by comprehensive gas chromatography-mass spectrometry and chemometric techniques. *J Breath Res* 2018;13(1):016005.
- [43] Bikov A, Lázár Z, Horvath I. Established methodological issues in electronic nose research: how far are we from using these instruments in clinical settings of breath analysis? *J Breath Res* 2015;9(3):034001.
- [44] Bikov A, Hermadi M, Korosi B, Kunos L, Zsomboki G, Sutto Z, et al. Expiratory flow rate, breath hold and anatomic dead space influence electronic nose ability to detect lung cancer. *BMC Pulm Med* 2014;14(1):202.
- [45] Lechner M, Moser B, Niederseer D, Karlseder A, Holzknicht B, Fuchs M, et al. Gender and age specific differences in exhaled isoprene levels. *Respir Physiol Neurobiol* 2006;154(3):478–83.
- [46] Buratti S, Sinelli N, Bertone E, Venturello A, Casiraghi E, Geobaldo F. Discrimination between washed Arabica, natural Arabica and Robusta coffees by using near infrared spectroscopy, electronic nose and electronic tongue analysis. *J Sci Food Agric* 2015;95(11):2192–200.
- [47] Chen Q, Liu A, Zhao J, Ouyang Q. Classification of tea category using a portable electronic nose based on an odor imaging sensor array. *J Pharm Biomed Anal* 2013;84:77–83.
- [48] Dutta R, Kashwan KR, Bhuyan M, Hines EL, Gardner JW. Electronic nose based tea quality standardization. *Neural Netw* 2003;16(5–6):847–53.
- [49] Yang CJ, Ding W, Ma LJ, Jia R. Discrimination and characterization of different intensities of goaty flavor in goat milk by means of an electronic nose. *J Dairy Sci* 2015;98(1):55–67.
- [50] Wojnowski W, Majchrzak T, Dymerski T, Gębicki J, Namieśnik J. Portable electronic nose based on electrochemical sensors for food quality assessment. *Sensors* 2017;17(12):2715.
- [51] Kushch I, Schwarz K, Schwentner L, Baumann B, Dzien A, Schmid A, et al. Compounds enhanced in a mass spectrometric profile of smokers' exhaled breath versus non-smokers as determined in a pilot study using PTR-MS. *J Breath Res* 2008;2(2):026002.
- [52] Mochalski P, King J, Unterkofler K, Amann A. Stability of selected volatile breath constituents in Tedlar, Kynar and Flexfilm sampling bags. *Analyst* 2013 Mar 7;138(5):1405–18 Internet. cited 2019 Apr 8. Available from <http://www.ncbi.nlm.nih.gov/pubmed/23232621>.
- [53] Chambers ST, Scott-Thomas A, Epton M. Developments in novel breath tests for bacterial and fungal pulmonary infection. *Curr Opin Pulm Med* 2012 May;18(3):228–32.

- [54] Bruins M. Transferable odor differentiation models for infectious disease diagnostics. Rotterdam: Erasmus MC University Medical Center; 2014.
- [55] Saktiawati AMI, Subronto YW, Stienstra Y, Sumardi Supit F, Werf TS. Sensitivity and specificity of routine diagnostic work-up for tuberculosis in lung clinics in Yogyakarta, Indonesia: a cohort study. *BMC Public Health* [Internet] 2019 Apr 2;19(1):363. <https://doi.org/10.1186/s12889-019-6658-8> Available from <http://www.ncbi.nlm.nih.gov/pubmed/30940123>. (PubMed PMID: 30940123; PubMed Central PMCID: PMC6444523).
- [56] Španěl P, Smith D. Progress in SIFT-MS: breath analysis and other applications. *Mass Spectrom Rev* 2011;30(2):236–67.
- [57] Li S, Liu B, Peng M, Chen M, Yin W, Tang H, et al. Diagnostic accuracy of Xpert MTB/RIF for tuberculosis detection in different regions with different endemic burden: a systematic review and meta-analysis. *PLoS One* 2017;12(7):1–13.