

Article

Accuracy of the Electronic Nose Breath Tests in Clinical Application: A Systematic Review and Meta-Analysis

Hsiao-Yu Yang^{1,2,3,*} , Wan-Chin Chen^{1,4} and Rodger-Chen Tsai¹

¹ Institute of Environmental and Occupational Health Sciences, National Taiwan University, Taipei 10055, Taiwan; 117152@cch.org.tw (W.-C.C.); ricxeor@gmail.com (R.-C.T.)

² Department of Public Health, National Taiwan University College of Public Health, Taipei 10055, Taiwan

³ Department of Environmental and Occupational Medicine, National Taiwan University Hospital, Taipei 10002, Taiwan

⁴ Department of Family Medicine, Changhua Christian Hospital, Changhua 50006, Taiwan

* Correspondence: hyang@ntu.edu.tw; Tel.: +886-2-3366-8102

Abstract: (1) Background: An electronic nose applies a sensor array to detect volatile biomarkers in exhaled breath to diagnose diseases. The overall diagnostic accuracy remains unknown. The objective of this review was to provide an estimate of the diagnostic accuracy of sensor-based breath tests for the diagnosis of diseases. (2) Methods: We searched the PubMed and Web of Science databases for studies published between 1 January 2010 and 14 October 2021. The search was limited to human studies published in the English language. Clinical trials were not included in this review. (3) Results: Of the 2418 records identified, 44 publications were eligible, and 5728 patients were included in the final analyses. The pooled sensitivity was 90.0% (95% CI, 86.3–92.8%, $I^2 = 47.7\%$), the specificity was 88.4% (95% CI, 87.1–89.5%, $I^2 = 81.4\%$), and the pooled area under the curve was 0.93 (95% CI 0.91–0.95). (4) Conclusion: The findings of our review suggest that a standardized report of diagnostic accuracy and a report of the accuracy in a test set are needed. Sensor array systems of electronic noses have the potential for noninvasiveness at the point-of-care in hospitals. Nevertheless, the procedure for reporting the accuracy of a diagnostic test must be standardized.

Keywords: volatile organic compound; electronic nose; sensors; breath test; breathomics



Citation: Yang, H.-Y.; Chen, W.-C.; Tsai, R.-C. Accuracy of the Electronic Nose Breath Tests in Clinical Application: A Systematic Review and Meta-Analysis. *Biosensors* **2021**, *11*, 469. <https://doi.org/10.3390/bios11110469>

Received: 6 November 2021

Accepted: 20 November 2021

Published: 22 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Human exhaled breath contains numerous volatile metabolites produced during diseases' physiological and pathological processes that can be used as volatile biomarkers for diagnosis [1]. Breathomics is an emerging science to diagnose diseases by analyzing volatile organic compounds (VOCs) produced by changes in metabolic processes caused by disease [1]. The electronic nose (E-nose) and gas chromatography-mass spectrometry (GC-MS) are two methods to analyze these VOCs. In contrast to the E-nose, GC-MS allows us to explore possible biological pathways and identify specific VOCs associated with the pathological changes of the diseases. The E-nose aims to develop point-of-care diagnostic breath tests [2]. The E-nose uses a nonselective sensor array to identify the pattern of VOCs in exhaled breath. When VOCs from a breath sample is presented to the sensor array, the E-nose system processes the response signals of sensor arrays and uses machine learning techniques to discriminate the VOCs of patients from healthy subjects and provides the most likely diagnosis based on smell [3]. Currently, the E-nose has been implemented in the diagnosis of lung cancer [4], breast cancer [5], colorectal cancer [6], ovarian cancer [7], gastric cancer [8], head-and-neck cancer [9], chronic obstructive lung disease (COPD) [10], interstitial lung disease [11], liver cirrhosis [12], ventilator-associated pneumonia [13], and Coronavirus Disease 2019 (COVID-19) [14]. In artificial intelligence (AI), the development of electronic nose systems is an emerging science that can provide real-time analysis

and assist clinical decisions. There are two major types of sensors: (1) nanomaterial-based sensors, including single-walled carbon nanotubes (CNTs), monolayer capped metal nanoparticle (MCNP) films and metal oxide (MO) sensors, and (2) electroacoustic sensors that include quartz microbalance (QMB) and surface acoustic wave (SAW) sensors [15].

The current knowledge gap on the application of E-noses to clinical diagnosis remains uncertain. Due to the advancement of material sciences, many types of E-nose sensors have been developed in recent years [16]. Although many types of sensors have been designed to detect more diseases in recent years, E-noses have not yet been applied in clinical practice. An updated systemic review and meta-analysis are necessary to provide quantitative and qualitative estimates of the accuracy of the E-nose in actual patients.

The specific aims of this review were to (1) summarize the diagnostic accuracy of sensor-based exhaled breath tests for clinical diagnoses and (2) compare the accuracy of different types of sensors. To achieve these aims, we performed a systematic review of the published evidence regarding the use of the sensors in breath tests for clinical diagnosis.

2. Materials and Methods

This meta-analysis was conducted following the PRISMA 2020 guidelines for reporting systematic reviews [17]. The review included only studies that analyzed VOCs in the exhaled breath of human subjects in hospitals. Studies that involved cell lines or animal studies were excluded. All analyses were based on previously published studies, and thus, no ethical approval or patient consent was required.

2.1. Eligibility Criteria

Studies were included if they met the following criteria: (1) the study analyzed VOCs within exhaled breath; (2) the study was an observational study, with a cross-sectional, case-control, or prospective design; and (3) the study's population consisted of patients or healthy controls enrolled from hospitals. The exclusion criteria were as follows: (1) *in vitro* experiments; (2) animal studies; (3) studies in which VOCs were analyzed not in exhaled breath but in breath condensate or tissue, including urine, blood, stool, or other biofluids; (4) reports not published in the English language; (5) studies of laboratory testing of sensor prototypes that were not applied in a clinical setting; (6) duplicate publications; (7) letters or review articles; and (8) studies that did not provide sufficient information on case number, control number, sensitivity, and specificity to construct the 2×2 contingency table.

2.2. Information Sources

We selected related studies published between 1 January 2000 and 14 October 2021 by searching PubMed and Web of Science. We also searched documents that cited any of the initially included studies as well as the references of the initially included studies.

2.3. Search Strategy

We used the following combined text in Web of Science: (ALL = (breath analysis OR breath test)) AND ALL = (sensor). The complete search used for PubMed was: Search: (breath analysis OR breath test) AND (sensor) Filters: from 2010–2022 ((“breath tests”[MeSH Terms] OR (“breath”[All Fields] AND “tests”[All Fields]) OR “breath tests”[All Fields] OR (“breath”[All Fields] AND “analysis”[All Fields]) OR “breath analysis”[All Fields] OR (“breath tests”[MeSH Terms] OR (“breath”[All Fields] AND “tests”[All Fields]) OR “breath tests”[All Fields] OR (“breath”[All Fields] AND “test”[All Fields]) OR “breath test”[All Fields])) AND (“sensor”[All Fields] OR “sensors”[All Fields] OR “sensoric”[All Fields] OR “sensorics”[All Fields] OR “sensing”[All Fields] OR “sensorization”[All Fields] OR “sensorized”[All Fields] OR “sensors”[All Fields])) AND (2010:2022[pdat]). Furthermore, the reference lists of relevant articles were manually examined to determine additional potentially related studies. The searches were carried out independently by two investigators (H.-Y.Y. and R.C.T.). Later, we searched documents that cited any of

the initially included studies as well as the references of the initially included studies. However, no extra articles that fulfilled the inclusion criteria were found in these searches. Full details of the search strategy are provided in Figure 1.

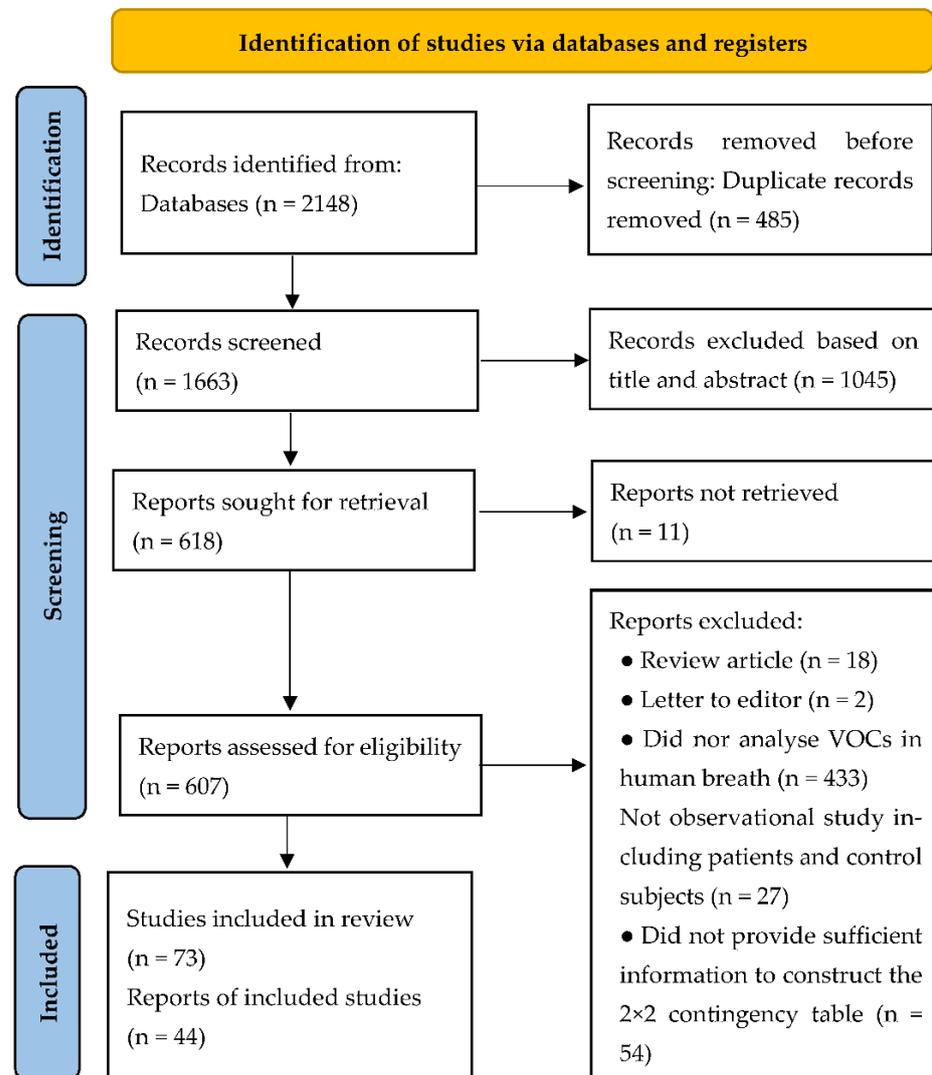


Figure 1. PRISMA flow chart of literature search.

2.4. Selection Process

Three investigators (R.-C.T., W.-C.C.) independently reviewed the study titles and abstracts at first and discussed the inconsistencies until consensus was obtained. Disagreements were resolved by consensus and discussion with the corresponding author (H.-Y.Y.). We contacted the corresponding author if further information was needed. If no response was received, the study was excluded from the meta-analysis.

2.5. Data Items

We extracted the following study characteristics from each eligible study: the name of the first author, publication year, country, disease, number of participants, and type of sensor. Each investigator also recorded or calculated the number of false positives (FPs), true positives (TPs), false negatives (FNs), and true negatives (TNs). For studies that reported the results of different machine learning algorithms, we selected the best results for the meta-analysis. For studies with multiple comparison groups (i.e., cancer, benign disease, and healthy controls), we derived data from the primary disease and healthy controls.

2.6. Quality Assessment

We used a modified Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool to assess the quality of the included studies. QUADAS-2 consists of four domains, including patient selection, index test, reference standard, and flow of patients through the study [18]. Two reviewers (H.-Y.Y. and W.-C.C.) independently rated all included studies. We used the QUADAS-2 sheet built in the RevMan 5.3 software of Cochrane to provide a methodological quality summary [19].

2.7. Statistical Analysis

We obtained the numbers of FPs, TPs, FNs, and TNs to calculate the pooled point estimates of sensitivity, specificity, and summary ROC curve of breath tests [20]. ROC values of 0.7–0.8, 0.8–0.9, and 0.9–1 are regarded as good, very good, and excellent diagnostic accuracy, respectively [21]. Statistical heterogeneity caused by non-threshold effects was tested by the Q test and I^2 test. An I^2 value greater than 50% was considered to indicate significant heterogeneity [22]. If considerable heterogeneity could not be eliminated, a random-effects model was used [23]. We generated funnel plots to evaluate small study effects and applied Egger's test to assess funnel plot asymmetry [24]. Because simple pooling of sensitivity and specificity is usually inappropriate, as this approach ignores threshold differences, we also calculated the diagnostic odds ratio (DOR):

$$\text{DOR} = \frac{\text{TP}}{\text{FP}} \bigg/ \frac{\text{FN}}{\text{TN}} \quad (1)$$

The DOR is not prevalence dependent and offers considerable advantages in a meta-analysis of diagnostic studies with increased precision. The value of a DOR ranges from 0 to infinity, with higher values indicating better discriminatory test performance [25]. Because the accuracy in the test set is usually lower than that in the overall dataset, we compared the accuracy from the overall dataset and test set.

2.8. Sensitivity Analysis

We conducted sensitivity analyses to determine if there was an undue influence of publication bias on the pooled estimates of accuracy. We deleted studies when the number of FNs or FPs was less than or equal to zero or one, which resulted in a DOR higher than 300. We also restricted the studies to low risk of bias, in which there was less than or equal to one high-risk QUADAS-2 domain, to see the influence of bias from patient selection, index test, reference standard, and flow of patients through the study.

2.9. Subgroup Analysis

Subgroup analysis was performed to explore the sources of heterogeneity according to the characteristics of the included articles. We conducted a subgroup analysis to compare the DOR of different types of sensors.

We performed the meta-analysis with the R packages of MADA, META, and Metafore and Review Manager 5 software. A two-tailed p -value less than 0.05 was considered statistically significant.

3. Results

The review identified 44 relevant publications and 5728 subjects (Figure 1). The number of subjects included in those studies ranged from 25 to 1167 (median 82). Within included studies, aneurysm ($n = 1$), appendicitis ($n = 1$), asthma ($n = 1$), breast cancer ($n = 2$), bronchial and laryngeal cancer ($n = 1$), colorectal cancer ($n = 1$), chronic obstructive pulmonary disease (COPD) ($n = 1$), COPD and lung cancer ($n = 1$), cutaneous leishmaniasis ($n = 1$), echinococcosis ($n = 1$), epilepsy ($n = 1$), gastric cancer ($n = 3$), hemodialysis ($n = 1$), head and neck cancer ($n = 2$), head-and-neck cancer and lung cancer ($n = 1$), heart failure ($n = 1$), interstitial lung disease ($n = 1$), liver cirrhosis ($n = 2$), lung cancer ($n = 7$), multiple sclerosis ($n = 2$), ovarian cancer ($n = 2$), Parkinson's disease ($n = 1$), pneumoconiosis ($n = 1$),

preeclampsia (n = 1), rhinosinusitis (n = 1), COVID-19 (n = 3), small airway dysfunction (n = 1), and ventilator-associated pneumonia (n = 2) were studied. The most common type of sensor was metal oxide (n = 11), followed by carbon nanotubes (n = 9), gold nanoparticles and carbon nanotubes (n = 7), gold nanoparticles (n = 6), quartz microbalances (n = 5), metal nanoparticles (n = 2), organic polymers (n = 1), polycyclic aromatic hydrocarbons and single-wall carbon nanotubes (n = 1), and WO₃ nanowires (n = 1) (Table 1).

Table 1. The table displays for each included study.

Study	Disease	Sensor	Case	Control	Sensitivity	Specificity
Mommers [26], 2020	Aneurysm and recurrent hernia	Metal-oxide	64 ^a	74	0.81	0.73
Wong [27], 2019	Appendicitis	Conductive polymer	5	45	0.8	0.8
Montuschi [28], 2010	Asthma	Quartz microbalances	30	21	0.87	0.95
Barash [29], 2015	Breast cancer	Gold nanoparticles and carbon nanotubes	169	82	0.88	0.83
Yang [5], 2021	Breast cancer	Carbon nanotubes	70 ^b	18 ^a	0.86 ^b	0.97 ^b
Fielding [30], 2020	Bronchial and laryngeal cancer	Carbon nanotubes	42	13	0.95	0.69
Amal [6], 2016	Colorectal cancer	Gold nanoparticles and carbon nanotubes	20 ^b	36 ^b	0.94 ^b	0.91 ^b
Shafiek [31], 2015	COPD	Carbon nanotubes	124	30	0.69	0.75
Binson [10], 2021	COPD and lung cancer	Metal-oxide	70	144	0.81	0.94
Welearegay [32], 2018	Cutaneous leishmaniasis	Metal nanoparticles	28 ^a	28	0.96	1
Welearegay [33], 2019	Echinococcosis	Metal nanoparticles	36	40	0.97	0.98
van Dartel [34], 2020	Epilepsy	Metal-oxide	74	110	0.76	0.67
Broza [8], 2019	Gastric cancer	Gold nanoparticles	102	1065	0.82	0.79
Xu [35], 2013	Gastric cancer	Gold nanoparticles and carbon nanotubes	37	93	0.89	0.9
Leja [36], 2021	Gastric cancer	Gold nanoparticles	47	105	0.92	0.86
Umapathy [37], 2019	Haemodialysis	Metal-oxide	21	17	0.86	0.29
Gruber [38], 2014	Head and neck cancer	Nanomaterial-based sensor	22	19	0.77	0.9
Leunis [39], 2014	Head and neck cancer	Metal-oxide	36	23	0.9	0.8
Hakim [9], 2011	Head-and-neck cancer and lung cancer	Gold nanoparticles	36 ^a	52	1	0.92
Finamore [40], 2018	Heart failure	Quartz microbalances	30 ^b	39 ^b	0.8 ^b	0.82 ^b
Moor [11], 2021	Interstitial lung disease	Metal-oxide	322	48	1	1
De Vincentis [12], 2016	Liver cirrhosis	Quartz microbalances	58	56	1	0.98
Zaim [41], 2021	Liver cirrhosis	WO ₃ nanowires	22	32	0.97	1
Gasparri [42], 2016	Lung cancer	Quartz microbalances	72	74	0.88	1
Huang [4], 2018	Lung cancer	Carbon nanotubes	56	188	0.92	0.93
Hubers [43], 2014	Lung cancer	Carbon nanotubes	38	39	0.87	0.43
Kononov [44], 2020	Lung cancer	Metal-oxide	19 ^b	17 ^b	0.95 ^b	1 ^b
Rocco [45], 2016	Lung cancer	Quartz microbalances	23	77	0.86	0.95
Shlomi [46], 2017	Lung cancer	Gold nanoparticles and carbon nanotubes	16	30	0.75	0.93
Tan [47], 2016	Lung cancer	Metal-oxide	12	13	0.83	0.88
Broza et al. [48], 2017	Multiple sclerosis	Gold nanoparticles	128	58	0.76	0.81
Nakhleh et al. [49], 2015	Parkinson's disease	Gold nanoparticles and carbon nanotubes	16	37	0.81	0.76
Ionescu et al. [50], 2011	Multiple sclerosis	Polycyclic aromatic hydrocarbons and single-wall carbon nanotubes	34	17	0.85	0.71

Table 1. Cont.

Study	Disease	Sensor	Case	Control	Sensitivity	Specificity
Amal et al. [51], 2015	Ovarian cancer	Gold nanoparticles and carbon nanotubes	48	48	0.85	0.65
Raspagliesi et al. [7], 2020	Ovarian cancer	Metal-oxide	86	114	0.98	0.95
Yang et al. [52], 2018	Pneumoconiosis	Carbon nanotubes	34	64	0.68	0.84
Nakhleh et al. [53], 2016	Preeclampsia	Gold nanoparticles	31	31	0.92	0.91
Broza et al. [54], 2018	Rhinosinusitis	Gold nanoparticles and carbon nanotubes	17	30	0.76	0.8
Zamora-Mendoza et al. [55], 2022	SARS-CoV-2	Carbon nanotubes	42	30	0.97	1
Shan et al. [14], 2020	SARS-CoV-2	Gold nanoparticles	41	57	1	0.81
Wintjens et al. [56], 2020	SARS-CoV-2	Metal-oxide	57	162	0.86	0.54
Tsai et al. [57], 2021	Small airway dysfunction	Carbon nanotubes	12	60	0.92	0.95
Chen et al. [13], 2020	Ventilator-associated pneumonia	Carbon nanotubes	33	26	0.72	0.77
Schnabel et al. [58], 2015	Ventilator-associated pneumonia	Metal-oxide	33	53	0.88	0.66

^a Included data from the model for two disease outcomes. ^b Data derived from a test database.

3.1. Pooled Sensitivity, Specificity, ROC and DOR

The sensitivity of breath tests by sensor arrays ranged from 67.6% to 100%, whereas the specificity ranged from 29.4% to 100%. The pooled sensitivity was 90.0% (95% CI, 86.3–92.8%, $I^2 = 47.7%$), the specificity was 88.4% (95% CI, 87.1–89.5%, $I^2 = 81.4%$), the pooled area under the curve of 0.93 (95% CI 0.91–0.95) (Figure 2), and the pooled DOR was 40.7 (95% CI 24.2–68.5, $I^2 = 77.0%$) (Figure 3). The funnel plot asymmetry and linear regression test (p value < 0.05) suggested potential publication bias (Figure 4).

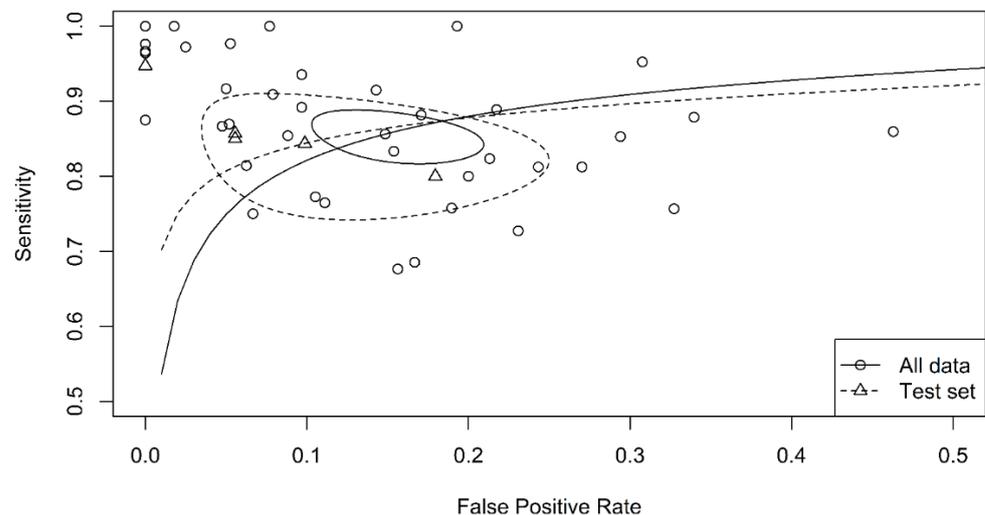


Figure 2. Summary receiver operating characteristic curve graph of the included studies. The accuracy using all data was higher than that of the test set.

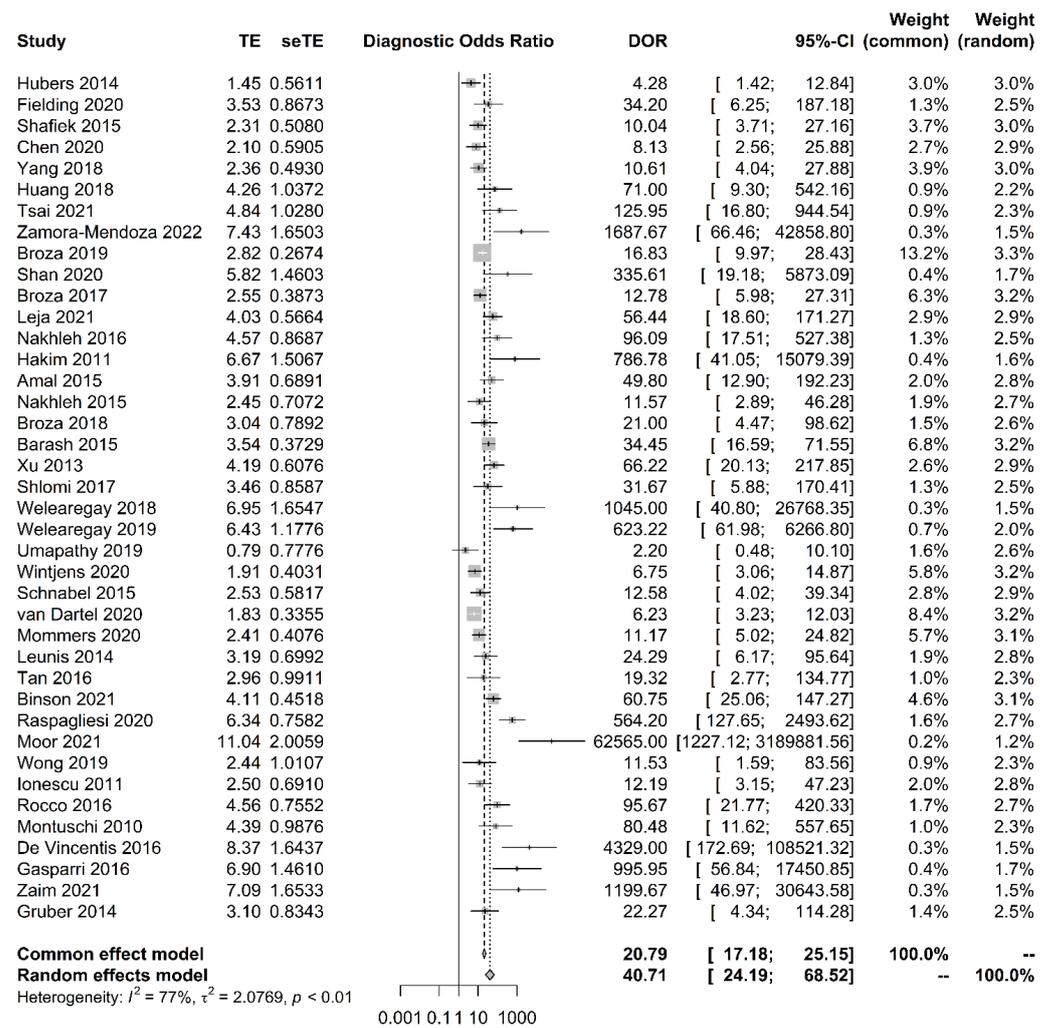


Figure 3. Forest plot and pooled diagnostic odds ratio analysis. Vertical dashed lines indicate 95% CI for the pooled effect. The size of the data markers reflects the weight. Error bars indicate 95% CI.

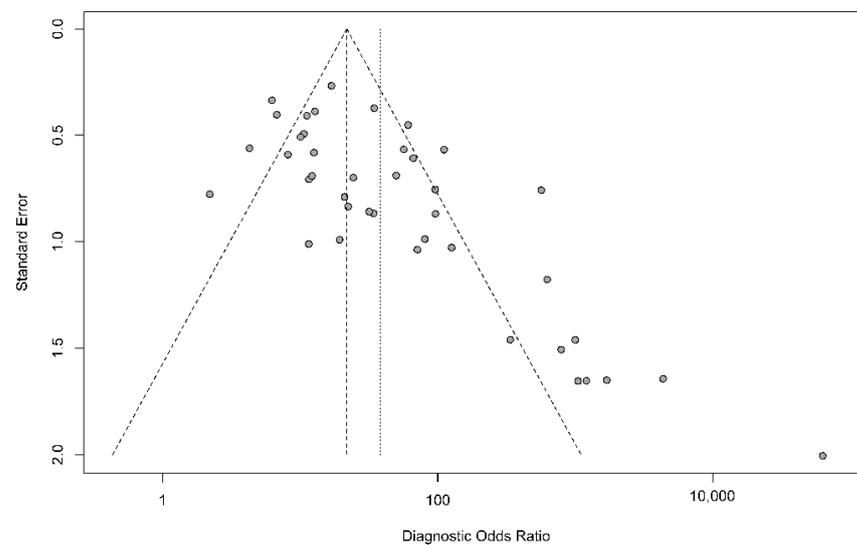


Figure 4. Funnel plot of the diagnostic odds ratio. A skewed asymmetrical funnel plot shows that there is publication bias. In the right lower corner, the small sample size studies (therefore large standard error) are more prone to publication bias than large studies.

3.2. Quality Assessment

The assessment of biases and applicability to outcomes utilizing QUADAS-2 are detailed in Figure 5. Major sources of bias were patient selection, followed by failing to report the reference standard, and flow and timing. The patient selection also became the major applicability concern for the E-nose test.

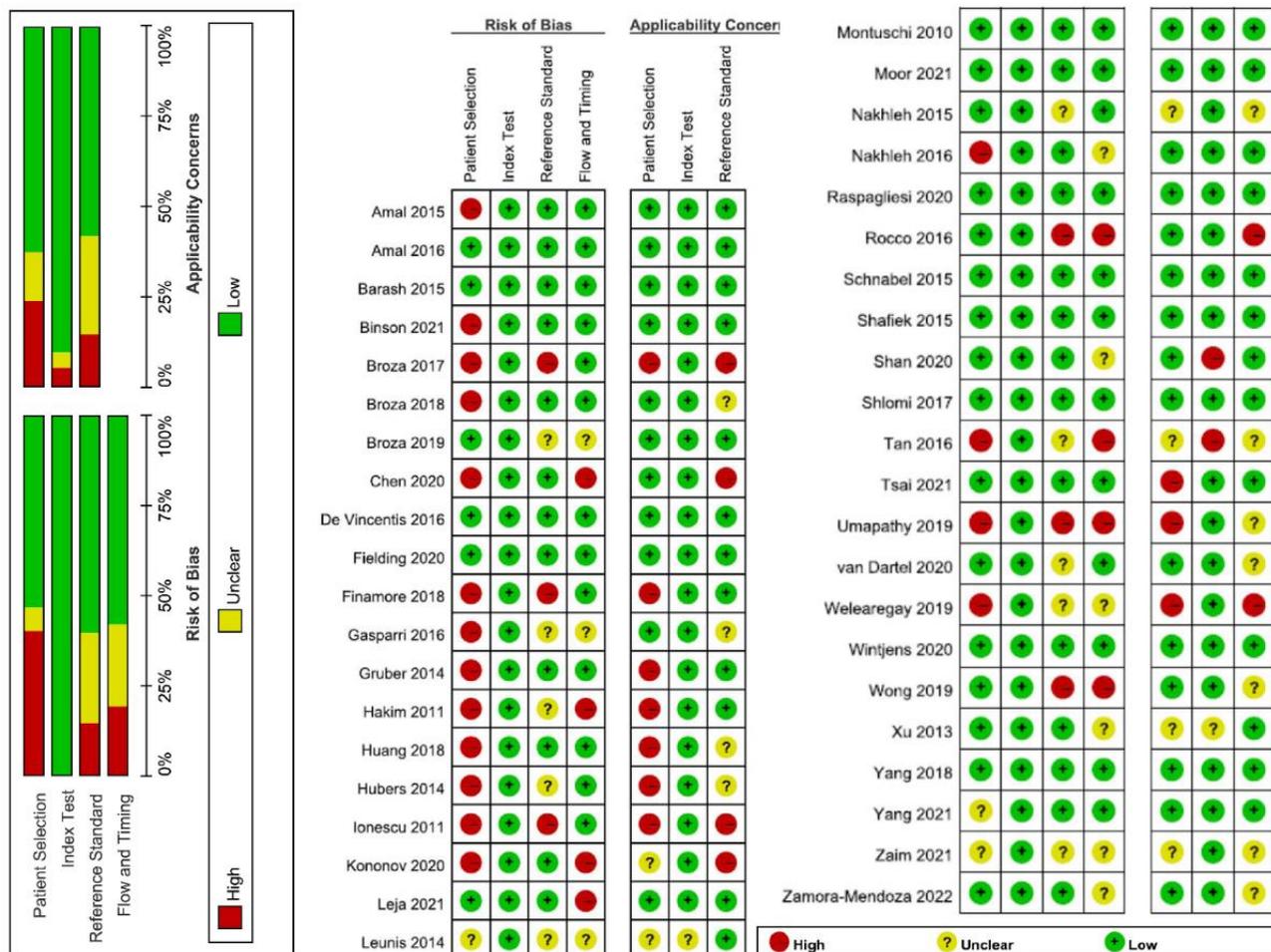


Figure 5. Quality assessment of included studies by the QUADAS-2 tool. This figure shows the proportion of studies with low (green colour), unclear (yellow), and high risk/concern (red). In terms of the overall risk of bias, there were concerns about the risk of bias for 26.5% of the studies (13/44), with two of these assessed as at high risk of bias.

3.3. Sensitivity Analysis

After excluding 14 studies that reported that the number of false negatives or false positives was less than or equal to zero or one, the funnel plot was symmetric without publication bias (linear regression test of funnel plot asymmetry, $p = 0.07$), suggesting that the influence of these studies on the pooled results was acceptable, and the pooled results were robust to some extent. The pooled sensitivity was 83.4% (95% CI 80.3–86.1%, $I^2 = 38.7%$), pooled specificity was 82.3% (95% CI 76.8–86.7%, $I^2 = 83.4%$), pooled DOR was 19.3 (95% CI 13.6–27.4, $I^2 = 61.7%$), and pooled area under the curve was 0.91 (95% CI 0.89–0.93). When restricting the analysis to studies considered to be at low risk of bias ($n = 27$), the pooled sensitivity was 90.8% (95% CI 86.0–94.0%, $I^2 = 56.5%$), pooled specificity was 89.4% (95% CI 87.9–90.8%, $I^2 = 81.4%$), and pooled area under the curve was 0.93 (95% CI 0.91–0.96).

3.4. Subgroup Analysis

We compared the accuracy of different types of sensors. Metal nanoparticle sensors had the highest DOR, sensitivity, and specificity (Table 2) (Figure 6).

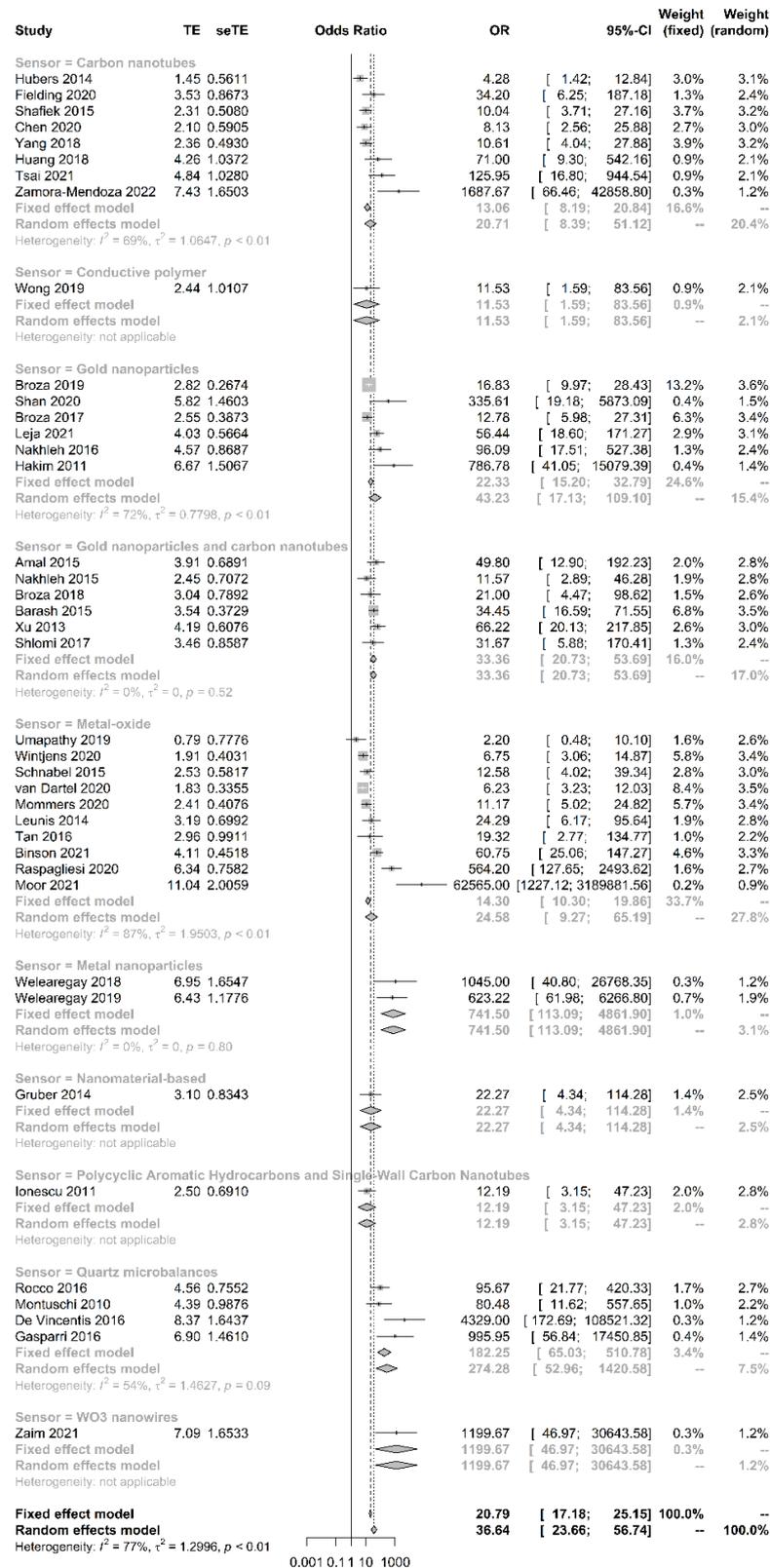


Figure 6. Subgroup analysis for pooled diagnostic odds ratio based on the type of sensor. The type of sensor is based on the classification provided in the literature.

Table 2. Subgroup analysis based on the type of sensor.

Type ¹	Sensitivity (95% CI)	I ²	Specificity (95% CI)	I ²
Carbon nanotube (n = 8)	0.86 (0.75, 0.93)	69.4%	0.86 (0.71, 0.94)	82.1%
Conductive polymer (n = 1)	0.80 (0.31, 0.97)	NA	0.80 (0.66, 0.89)	NA
Gold nanoparticles (n = 6)	0.94 (0.80, 0.98)	39.8%	0.83 (0.78, 0.88)	48.5%
Gold nanoparticles and carbon nanotube (n = 6)	0.86 (0.82, 0.90)	0.0%	0.87 (0.82, 0.91)	32.5%
Metal-oxide (n = 10)	0.91 (0.81, 0.96)	35.2%	0.81 (0.63, 0.91)	89.5%
Metal nanoparticles (n = 2)	0.97 (0.88, 0.99)	0.0%	0.99 (0.90, 1.00)	0.0%
Nanomaterial-based (n = 1)	0.77 (0.56, 0.90)	NA	0.89 (0.66, 0.97)	NA
Polycyclic aromatic hydrocarbons and single wall carbon nanotubes (n = 1)	0.85 (0.69, 0.94)	NA	0.71 (0.46, 0.87)	NA
Quartz microbalances (n = 4)	0.93 (0.81, 0.97)	0.0%	0.98 (0.93, 0.99)	0.0%
WO ₃ nanowires (n = 1)	0.97 (0.80, 1.00)	NA	1.00 (0.00–1.00)	NA

¹ The type of sensor is based on the classification provided in the literature.

4. Discussion

4.1. Summary of Main Results

This study provided evidence that the electronic nose analysis of exhaled breath has high accuracy in detecting diseases in actual patients. To the best of our limited knowledge, this is the first study to provide an overall estimate of the accuracy of the E-nose in clinical practice.

4.2. Strengths of the Review

This is the first study to provide a comprehensive review and pooled estimates of the diagnostic accuracy of E-noses in a clinical setting. This review provides quantitative estimates of the accuracy of different sensors, which will provide a basis for future researchers to choose suitable sensors and estimate the required sample size. There was moderate heterogeneity in the pooled sensitivity ($I^2 = 47.7\%$, $p < 0.05$) and high heterogeneity in specificity ($I^2 = 81.4\%$, $p < 0.05$). After we ruled out studies that reported extraordinarily high-accuracy studies to obtain a more reliable accuracy in clinical breath tests, the heterogeneity was improved and showed no publication bias. Because the preparation of breath tests, collection of breath, storage of breath samples, and preprocessing sensor array data will influence the results of breath tests [16], current studies lack the standardization procedures of breath collection and machine learning analysis, which might cause heterogeneity from unknown sources. We suggested that a depository of analytical procedures before the implementation of statistical modeling might be essential to prevent heterogeneity in the diagnostic accuracy of breath tests.

4.3. Applicability of Findings to the Review Question

The current state of knowledge on the application of E-noses to clinical diagnosis remains uncertain. Although many types of sensors have been developed for many diseases, the accuracy of the E-nose is unclear because of the study design, patient selection, and lack of a standardized way to report the diagnostic accuracy. Ideally, a study should enroll a consecutive or random sample of eligible patients with the suspected disease to prevent the potential for bias [18]. However, in the clinical setting, when the prevalence of the disease is not high, researchers usually prefer to conduct a case-control study that enrolled participants with known disease and a control group without the condition that may increase the overall diagnostic accuracy [59]. If the disease prevalence of the research subjects included in the study is different from the target population, this will affect the applicability of the E-nose in the target population [18].

4.4. Limitations

Different studies used different machine learning algorithms, and the optimization procedure was not reported in most studies. The influence of accuracy might be affected

by the type of sensor and statistical analyses. Furthermore, the limited sample size of the test sets may decrease the reported accuracy. Few studies had an independent test set to validate the test. This systematic review does not include studies with external validation tests; the pooled estimates of diagnostic accuracy from these studies cannot be generalized to other populations. We suggest that multicenter clinical studies among target populations with appropriate sample sizes and an independent validation set in different hospitals are crucial before an E-nose can be used in clinical applications.

4.5. Future Direction

There are several timings of breath tests in clinical practice. First, for patients who present with common nonspecific symptoms that could be an early indication of cancer, an exhaled breath test could act as a screening test. Second, for patients with suspected symptoms of diseases requiring further investigation, an E-nose can become a noninvasive point of care method before specialized investigations. Third, patients can receive therapy to detect disease recurrence. Current studies are focused on the first application. We suggest future studies for the second and third timings. Longitudinal studies are needed in the future to determine whether the electronic nose can be used to detect the recurrence of diseases.

The reproducibility of the results and reliability of instruments are future directions. Because E-nose studies are from diverse research fields, many researchers do not know how to provide essential items for reporting diagnostic accuracy studies. We recommend that future studies include clinical epidemiologists before implementing new breath tests to strengthen the study design, minimize the risk of bias and make the results more reliable. An E-nose is not capable of independently making a clinical diagnosis at this time. Physicians' clinical diagnosis based on clinical symptoms, signs, laboratory tests, and pathological reports remain an essential requirement before starting therapy or surgery in the current stage.

A standard breath test must control the flow rate and humidity and collect alveolar air that contains the metabolites from the alveolar-capillary membrane and released into the alveolar space [52]. To improve the efficiency of the electronic nose breath test, we suggest that future research can continue to optimize the breath collection device, which can automatically control the flow rate and humidity and monitor the CO₂ concentration to collect alveolar air containing volatile biomarkers.

5. Conclusions

Based on our meta-analysis, metal oxide sensors have good accuracy and may become important chip materials for electronic nose systems in the future. We encourage researchers currently using metal oxide sensors to conduct clinical trials to verify accuracy. In a breathomic study, case-control studies are suitable for exploring volatile biomarkers. However, in E-nose studies for diagnostic or screening purposes, studies that make inappropriate exclusions may result in an overestimation of diagnostic accuracy. The study ideally should enroll a consecutive or random sample of eligible patients with the suspected disease to prevent the potential for bias [18]. Machine learning techniques have gradually been applied in the medical field to establish a prediction model. However, many machine learning studies reported only the best accuracy value without showing details for readers to evaluate the reliability of test results. We suggest that studies should report the accuracy of the test set or an independent validation set. Researchers should not only show the best results with the highest accuracy; instead, a study should clearly explain all the procedures and conservatively estimate the accuracy for physicians in making clinical decisions [13].

Author Contributions: Conceptualization, H.-Y.Y.; methodology, H.-Y.Y.; formal analysis, H.-Y.Y.; investigation, W.-C.C., R.-C.T.; data curation, W.-C.C., R.-C.T.; writing—original draft preparation, H.-Y.Y.; writing—review and editing, H.-Y.Y.; funding acquisition, H.-Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Ministry of Science and Technology, Taiwan, grant numbers [MOST 109-2314-B-002-166-MY3] and was financially supported by the National Taiwan University (NTU) within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan [NTU-110L881004]. The funding bodies played no role in the design or execution of these studies beyond financial support.

Institutional Review Board Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Van der Schee, M.P.; Paff, T.; Brinkman, P.; van Aalderen, W.M.; Haarman, E.G.; Sterk, P.J. Breathomics in lung disease. *Chest* **2015**, *147*, 224–231. [[CrossRef](#)]
2. Hakim, M.; Broza, Y.Y.; Barash, O.; Peled, N.; Phillips, M.; Amann, A.; Haick, H. Volatile organic compounds of lung cancer and possible biochemical pathways. *Chem. Rev.* **2012**, *112*, 5949–5966. [[CrossRef](#)] [[PubMed](#)]
3. Queralto, N.; Berliner, A.N.; Goldsmith, B.; Martino, R.; Rhodes, P.; Lim, S.H. Detecting cancer by breath volatile organic compound analysis: A review of array-based sensors. *J. Breath Res.* **2014**, *8*, 027112. [[CrossRef](#)] [[PubMed](#)]
4. Huang, C.H.; Zeng, C.; Wang, Y.C.; Peng, H.Y.; Lin, C.S.; Chang, C.J.; Yang, H.Y. A Study of diagnostic accuracy using a chemical sensor array and a machine learning technique to detect lung cancer. *Sensors* **2018**, *18*, 2845. [[CrossRef](#)] [[PubMed](#)]
5. Yang, H.Y.; Wang, Y.C.; Peng, H.Y.; Huang, C.H. Breath biopsy of breast cancer using sensor array signals and machine learning analysis. *Sci. Rep.* **2021**, *11*, 9. [[CrossRef](#)]
6. Amal, H.; Leja, M.; Funka, K.; Lasina, I.; Skapars, R.; Sivins, A.; Ancans, G.; Kikuste, I.; Vanags, A.; Tolmanis, I.; et al. Breath testing as potential colorectal cancer screening tool. *Int. J. Cancer* **2016**, *138*, 229–236. [[CrossRef](#)]
7. Raspagliesi, F.; Bogani, G.; Benedetti, S.; Grassi, S.; Ferla, S.; Buratti, S. Detection of ovarian cancer through exhaled breath by electronic nose: A prospective study. *Cancers* **2020**, *12*, 2408. [[CrossRef](#)] [[PubMed](#)]
8. Broza, Y.Y.; Khatib, S.; Gharra, A.; Krilaviciute, A.; Amal, H.; Polaka, I.; Parshutin, S.; Kikuste, I.; Gasenko, E.; Skapars, R.; et al. Screening for gastric cancer using exhaled breath samples. *Br. J. Surg.* **2019**, *106*, 1122–1125. [[CrossRef](#)] [[PubMed](#)]
9. Hakim, M.; Billan, S.; Tisch, U.; Peng, G.; Dvorkind, I.; Marom, O.; Abdah-Bortnyak, R.; Kuten, A.; Haick, H. Diagnosis of head-and-neck cancer from exhaled breath. *Br. J. Cancer* **2011**, *104*, 1649–1655. [[CrossRef](#)] [[PubMed](#)]
10. Binson, V.A.; Subramoniam, M.; Mathew, L. Discrimination of COPD and lung cancer from controls through breath analysis using a self-developed e-nose. *J. Breath Res.* **2021**, *15*, 10. [[CrossRef](#)]
11. Moor, C.C.; Oppenheimer, J.C.; Nakshbandi, G.; Aerts, J.; Brinkman, P.; Maitland-van der Zee, A.H.; Wijzenbeek, M.S. Exhaled breath analysis by use of eNose technology: A novel diagnostic tool for interstitial Lung disease. *Eur. Resp. J.* **2021**, *57*, 8. [[CrossRef](#)] [[PubMed](#)]
12. De Vincentis, A.; Pennazza, G.; Santonico, M.; Vespasiani-Gentilucci, U.; Galati, G.; Gallo, P.; Vernile, C.; Pedone, C.; Incalzi, R.A.; Picardi, A. Breath-print analysis by e-nose for classifying and monitoring chronic liver disease: A proof-of-concept study. *Sci. Rep.* **2016**, *6*, 9. [[CrossRef](#)] [[PubMed](#)]
13. Chen, C.Y.; Lin, W.C.; Yang, H.Y. Diagnosis of ventilator-associated pneumonia using electronic nose sensor array signals: Solutions to improve the application of machine learning in respiratory research. *Respir. Res.* **2020**, *21*, 12. [[CrossRef](#)]
14. Shan, B.J.; Broza, Y.Y.; Li, W.J.; Wang, Y.; Wu, S.H.; Liu, Z.Z.; Wang, J.; Gui, S.Y.; Wang, L.; Zhang, Z.H.; et al. Multiplexed nanomaterial-based sensor array for detection of COVID-19 in exhaled breath. *ACS Nano* **2020**, *14*, 12125–12132. [[CrossRef](#)] [[PubMed](#)]
15. Haick, H.; Broza, Y.Y.; Mochalski, P.; Ruzsanyi, V.; Amann, A. Assessment, origin, and implementation of breath volatile cancer markers. *Chem. Soc. Rev.* **2014**, *43*, 1423–1449. [[CrossRef](#)] [[PubMed](#)]
16. Van der Sar, I.G.; Wijbenga, N.; Nakshbandi, G.; Aerts, J.; Manintveld, O.C.; Wijzenbeek, M.S.; Hellemons, M.E.; Moor, C.C. The smell of lung disease: A review of the current status of electronic nose technology. *Respir. Res.* **2021**, *22*, 246. [[CrossRef](#)] [[PubMed](#)]
17. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [[CrossRef](#)] [[PubMed](#)]
18. Whiting, P.F.; Rutjes, A.W.; Westwood, M.E.; Mallett, S.; Deeks, J.J.; Reitsma, J.B.; Leeflang, M.M.; Sterne, J.A.; Bossuyt, P.M.; Group, Q. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.* **2011**, *155*, 529–536. [[CrossRef](#)]
19. Review Manager (RevMan) [Computer Program]. Version 5.4, The Cochrane Collaboration, 2020. Available online: <https://training.cochrane.org/> (accessed on 19 November 2021).
20. Harbord, R.M.; Whiting, P.; Sterne, J.A.; Egger, M.; Deeks, J.J.; Shang, A.; Bachmann, L.M. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. *J. Clin. Epidemiol.* **2008**, *61*, 1095–1103. [[CrossRef](#)]
21. Simundic, A.M. Measures of diagnostic accuracy: Basic definitions. *EJIFCC* **2009**, *19*, 203–211. [[PubMed](#)]
22. Higgins, J.P.; Thompson, S.G.; Deeks, J.J.; Altman, D.G. Measuring inconsistency in meta-analyses. *BMJ* **2003**, *327*, 557–560. [[CrossRef](#)]

23. Jackson, D.; White, I.R.; Thompson, S.G. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Stat. Med.* **2010**, *29*, 1282–1297. [[CrossRef](#)]
24. Peters, J.L.; Sutton, A.J.; Jones, D.R.; Abrams, K.R.; Rushton, L. Comparison of two methods to detect publication bias in meta-analysis. *JAMA* **2006**, *295*, 676–680. [[CrossRef](#)] [[PubMed](#)]
25. Glas, A.S.; Lijmer, J.G.; Prins, M.H.; Bossel, G.J.; Bossuyt, P.M.M. The diagnostic odds ratio: A single indicator of test performance. *J. Clin. Epidemiol.* **2003**, *56*, 1129–1135. [[CrossRef](#)]
26. Mommers, E.H.H.; van Kooten, L.; Nienhuijs, S.W.; de Vries Reilingh, T.S.; Lubbers, T.; Mees, B.M.E.; Schurink, G.W.H.; Bouvy, N.D. Can electric nose breath analysis identify abdominal wall hernia recurrence and aortic aneurysms? A proof-of-concept study. *Surg. Innov.* **2020**, *27*, 366–372. [[CrossRef](#)]
27. Wong, D.C.; Relton, S.D.; Lane, V.; Ismail, M.; Goss, V.; Bytheway, J.; West, R.M.; Deuchars, J.; Sutcliffe, J. Bedside breath tests in children with abdominal pain: A prospective pilot feasibility study. *Pilot Feasibility Stud.* **2019**, *5*, 121. [[CrossRef](#)]
28. Montuschi, P.; Santonico, M.; Mondino, C.; Pennazza, G.; Mantini, G.; Martinelli, E.; Capuano, R.; Ciabattini, G.; Paolesse, R.; Di Natale, C.; et al. Diagnostic performance of an electronic nose, fractional exhaled nitric oxide, and lung function testing in asthma. *Chest* **2010**, *137*, 790–796. [[CrossRef](#)]
29. Barash, O.; Zhang, W.; Halpern, J.M.; Hua, Q.L.; Pan, Y.Y.; Kayal, H.; Khoury, K.; Liu, H.; Davies, M.P.; Haick, H. Differentiation between genetic mutations of breast cancer by breath volatolomics. *Oncotarget* **2015**, *6*, 44864–44876. [[CrossRef](#)]
30. Fielding, D.; Hartel, G.; Pass, D.; Davis, M.; Brown, M.; Dent, A.; Agnew, J.; Dickie, G.; Ware, R.S.; Hodge, R. Volatile organic compound breath testing detects in-situ squamous cell carcinoma of bronchial and laryngeal regions and shows distinct profiles of each tumour. *J. Breath Res.* **2020**, *14*, 13. [[CrossRef](#)]
31. Shafiek, H.; Fiorentino, F.; Merino, J.L.; López, C.; Oliver, A.; Segura, J.; de Paul, I.; Sibila, O.; Agustí, A.; Cosío, B.G. Using the electronic nose to identify airway infection during COPD exacerbations. *PLoS ONE* **2015**, *10*, e0135199. [[CrossRef](#)]
32. Welearegay, T.G.; Diouani, M.F.; Osterlund, L.; Ionescu, F.; Belgacem, K.; Smadhi, H.; Khaled, S.; Kidar, A.; Cindemir, U.; Laouini, D.; et al. Ligand-Capped Ultrapure Metal nanoparticle sensors for the detection of cutaneous Leishmaniasis disease in exhaled breath. *ACS Sens.* **2018**, *3*, 2532–2540. [[CrossRef](#)]
33. Welearegay, T.G.; Diouani, M.F.; Osterlund, L.; Borys, S.; Khaled, S.; Smadhi, H.; Ionescu, F.; Bouchekoua, M.; Aloui, D.; Laouini, D.; et al. Diagnosis of human echinococcosis via exhaled breath analysis: A promise for rapid diagnosis of infectious diseases caused by Helminths. *J. Infect. Dis.* **2019**, *219*, 101–109. [[CrossRef](#)]
34. Van Dartel, D.; Schelhaas, H.J.; Colon, A.J.; Kho, K.H.; de Vos, C.C. Breath analysis in detecting epilepsy. *J. Breath Res.* **2020**, *14*, 7. [[CrossRef](#)] [[PubMed](#)]
35. Xu, Z.Q.; Broza, Y.Y.; Ionsecu, R.; Tisch, U.; Ding, L.; Liu, H.; Song, Q.; Pan, Y.Y.; Xiong, F.X.; Gu, K.S.; et al. A nanomaterial-based breath test for distinguishing gastric cancer from benign gastric conditions. *Br. J. Cancer* **2013**, *108*, 941–950. [[CrossRef](#)]
36. Leja, M.; Kortelainen, J.M.; Polaka, I.; Turppa, E.; Mitrovics, J.; Padilla, M.; Mochalski, P.; Shuster, G.; Pohle, R.; Kashanin, D.; et al. Sensing gastric cancer via point-of-care sensor breath analyzer. *Cancer* **2021**, *127*, 1286–1292. [[CrossRef](#)] [[PubMed](#)]
37. Umopathy, S.; Nasimsha, N.; Kumar, M.; Kalidoss, R.; Thomas, A.C.; Lakshmi, M.; Gafoor, E.R. Design and development of portable prototype for human breath analysis: A comparative study between haemodialysis patients and healthy subjects. *Biomed. Phys. Eng. Express* **2019**, *5*, 9. [[CrossRef](#)]
38. Gruber, M.; Tisch, U.; Jeries, R.; Amal, H.; Hakim, M.; Ronen, O.; Marshak, T.; Zimmerman, D.; Israel, O.; Amiga, E.; et al. Analysis of exhaled breath for diagnosing head and neck squamous cell carcinoma: A feasibility study. *Br. J. Cancer* **2014**, *111*, 790–798. [[CrossRef](#)] [[PubMed](#)]
39. Leunis, N.; Boumans, M.L.; Kremer, B.; Din, S.; Stobberingh, E.; Kessels, A.G.H.; Kross, K.W. Application of an electronic nose in the diagnosis of head and neck cancer. *Laryngoscope* **2014**, *124*, 1377–1381. [[CrossRef](#)]
40. Finamore, P.; Pedone, C.; Lelli, D.; Costanzo, L.; Bartoli, I.R.; De Vincentis, A.; Grasso, S.; Parente, F.R.; Pennazza, G.; Santonico, M.; et al. Analysis of volatile organic compounds: An innovative approach to heart failure characterization in older patients. *J. Breath Res.* **2018**, *12*, 8. [[CrossRef](#)]
41. Zaim, O.; Diouf, A.; El Bari, N.; Lagdali, N.; Benelbarhdadi, I.; Ajana, F.Z.; Llobet, E.; Bouchikhi, B. Comparative analysis of volatile organic compounds of breath and urine for distinguishing patients with liver cirrhosis from healthy controls by using electronic nose and voltammetric electronic tongue. *Anal. Chim. Acta* **2021**, *1184*, 339028. [[CrossRef](#)]
42. Gasparri, R.; Santonico, M.; Valentini, C.; Sedda, G.; Borri, A.; Petrella, F.; Maisonnette, P.; Pennazza, G.; D'Amico, A.; Di Natale, C.; et al. Volatile signature for the early diagnosis of lung cancer. *J. Breath Res.* **2016**, *10*, 7. [[CrossRef](#)] [[PubMed](#)]
43. Hubers, A.J.; Brinkman, P.; Boksem, R.J.; Rhodius, R.J.; Witte, B.I.; Zwinderman, A.H.; Heideman, D.A.M.; Duin, S.; Koning, R.; Steenbergen, R.D.M.; et al. Combined sputum hypermethylation and eNose analysis for lung cancer diagnosis. *J. Clin. Pathol.* **2014**, *67*, 707–711. [[CrossRef](#)] [[PubMed](#)]
44. Kononov, A.; Korotetsky, B.; Jahatspanian, I.; Gubal, A.; Vasiliev, A.; Arsenjev, A.; Nefedov, A.; Barchuk, A.; Gorbunov, I.; Kozyrev, K.; et al. Online breath analysis using metal oxide semiconductor sensors (electronic nose) for diagnosis of lung cancer. *J. Breath Res.* **2020**, *14*, 14. [[CrossRef](#)] [[PubMed](#)]
45. Rocco, R.; Incalzi, R.A.; Pennazza, G.; Santonico, M.; Pedone, C.; Bartoli, I.R.; Vernile, C.; Mangiameli, G.; La Rocca, A.; De Luca, G.; et al. BIONOTE e-nose technology may reduce false positives in lung cancer screening programmes. *Eur. J. Cardio-Thorac. Surg.* **2016**, *49*, 1112–1117; discussion 1117. [[CrossRef](#)]

46. Shlomi, D.; Abud, M.; Liran, O.; Bar, J.; Gai-Mor, N.; Ilouze, M.; Onn, A.; Ben-Nun, A.; Haick, H.; Peled, N. Detection of Lung Cancer and EGFR Mutation by Electronic Nose System. *J. Thorac. Oncol.* **2017**, *12*, 1544–1551. [[CrossRef](#)]
47. Tan, J.L.; Yong, Z.X.; Liam, C.K. Using a chemiresistor-based alkane sensor to distinguish exhaled breaths of lung cancer patients from subjects with no lung cancer. *J. Thorac. Dis.* **2016**, *8*, 2772–2783. [[CrossRef](#)]
48. Broza, Y.Y.; Har-Shai, L.; Jeries, R.; Cancilla, J.C.; Glass-Marmor, L.; Lejbkowitz, I.; Torrecilla, J.S.; Yao, X.L.; Feng, X.L.; Narita, A.; et al. Exhaled Breath Markers for Nonimaging and Noninvasive Measures for Detection of Multiple Sclerosis. *ACS Chem. Neurosci.* **2017**, *8*, 2402–2413. [[CrossRef](#)]
49. Nakhleh, M.K.; Badarny, S.; Winer, R.; Jeries, R.; Finberg, J.; Haick, H. Distinguishing idiopathic Parkinson’s disease from other parkinsonian syndromes by breath test. *Parkinsonism Relat. Disord.* **2015**, *21*, 150–153. [[CrossRef](#)]
50. Ionescu, R.; Broza, Y.; Shaltieli, H.; Sadeh, D.; Zilberman, Y.; Feng, X.; Glass-Marmor, L.; Lejbkowitz, I.; Müllen, K.; Miller, A.; et al. Detection of multiple sclerosis from exhaled breath using bilayers of polycyclic aromatic hydrocarbons and single-wall carbon nanotubes. *ACS Chem. Neurosci.* **2011**, *2*, 687–693. [[CrossRef](#)]
51. Amal, H.; Shi, D.Y.; Ionescu, R.; Zhang, W.; Hua, Q.L.; Pan, Y.Y.; Tao, L.; Liu, H.; Haick, H. Assessment of ovarian cancer conditions from exhaled breath. *Int. J. Cancer* **2015**, *136*, E614–E622. [[CrossRef](#)]
52. Yang, H.Y.; Peng, H.Y.; Chang, C.J.; Chen, P.C. Diagnostic accuracy of breath tests for pneumoconiosis using an electronic nose. *J. Breath Res.* **2018**, *12*, 8. [[CrossRef](#)] [[PubMed](#)]
53. Nakhleh, M.K.; Baram, S.; Jeries, R.; Salim, R.; Haick, H.; Hakim, M. Artificially intelligent nanoarray for the detection of preeclampsia under real-world clinical conditions. *Adv. Mater. Technol.* **2016**, *1*, 1600132. [[CrossRef](#)]
54. Broza, Y.Y.; Braverman, I.; Haick, H. Breath volatolomics for diagnosing chronic rhinosinusitis. *Int. J. Nanomed.* **2018**, *13*, 4661–4670. [[CrossRef](#)] [[PubMed](#)]
55. Zamora-Mendoza, B.N.; de Leon-Martinez, L.D.; Rodriguez-Aguilar, M.; Mizaikoff, B.; Flores-Ramirez, R. Chemometric analysis of the global pattern of volatile organic compounds in the exhaled breath of patients with COVID-19, post-COVID and healthy subjects. Proof of concept for post-COVID assessment. *Talanta* **2022**, *236*, 9. [[CrossRef](#)] [[PubMed](#)]
56. Wintjens, A.; Hintzen, K.F.H.; Engelen, S.M.E.; Lubbers, T.; Savelkoul, P.H.M.; Wesseling, G.; van der Palen, J.A.M.; Bouvy, N.D. Applying the electronic nose for pre-operative SARS-CoV-2 screening. *Surg. Endosc.* **2020**, *8*. [[CrossRef](#)]
57. Tsai, Y.G.; Shie, R.H.; Huang, C.H.; Chen, C.D.; Lin, W.C.; Yang, H.Y. Use of the electronic nose to screen for small airway dysfunction in schoolchildren. *Sens. Actuator B Chem.* **2021**, *345*, 8. [[CrossRef](#)]
58. Schnabel, R.M.; Boumans, M.L.L.; Smolinska, A.; Stobberingh, E.E.; Kaufmann, R.; Roekaerts, P.; Bergmans, D. Electronic nose analysis of exhaled breath to diagnose ventilator-associated pneumonia. *Respir. Med.* **2015**, *109*, 1454–1459. [[CrossRef](#)]
59. Whiting, P.; Rutjes, A.W.; Reitsma, J.B.; Glas, A.S.; Bossuyt, P.M.; Kleijnen, J. Sources of variation and bias in studies of diagnostic accuracy: A systematic review. *Ann. Int. Med.* **2004**, *140*, 189–202. [[CrossRef](#)]