

## Review article:

# EXHALED VOLATILE ORGANIC COMPOUNDS IN THE DETECTION OF COLORECTAL CANCER: A SYSTEMATIC REVIEW AND META-ANALYSIS

Daniah Alsaadi<sup>a,b</sup>, Nicolle Clements<sup>b</sup>, Natiya Gabuniya<sup>a,c</sup>, Nader Francis<sup>a,d</sup>,  
Manish Chand<sup>a</sup>

- <sup>a</sup> The Division of Surgery and Interventional Science, University College London, London, United Kingdom  
<sup>b</sup> Clinical Research Facility Galway, Galway University Hospital, National University of Ireland, Galway, Republic of Ireland  
<sup>c</sup> Plastic and Reconstructive Surgery Department, Guy's and St Thomas' Hospital, London, United Kingdom  
<sup>d</sup> Department of General Surgery, Yeovil District Hospital NHS Foundation Trust, Yeovil, United Kingdom

\* **Corresponding author:** Daniah Alsaadi, The Division of Surgery and Interventional Science, University College London, London, United Kingdom.  
E-mail: [daniah.alsaadi@outlook.com](mailto:daniah.alsaadi@outlook.com)

<https://dx.doi.org/10.17179/excli2024-7042>

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>).

## ABSTRACT<sup>1</sup>

There is an apparent need for novel non-invasive colorectal cancer (CRC) screening tests that are more acceptable to patients and can reliably detect CRC or reduce the number of unnecessary colonoscopies performed in cancer-free patients. An emerging number of studies demonstrate the potential value of exhaled volatile organic compounds (VOCs) as a diagnostic and triaging test for CRC. A systematic appraisal and meta-analysis of the published evidence was done to determine whether exhaled VOCs can be used in the detection and screening of CRC. Nine electronic databases were searched from inception of the databases until August 2020. Quantitative and descriptive data of CRC patients and healthy control (HC) participants who underwent VOCs breath analysis was extracted. In addition, where possible, sampling methods, analytical platforms, processors, and specific breath biomarkers found in each study were recorded. Fourteen articles were included in the systematic review with 491 colorectal patients and 754 HC participants (n=1245). Sub-group meta-analysis was conducted on nine of those articles and the pooled sensitivity was estimated to be 0.89 (95 % CI = 0.80-0.99) whereas specificity was 0.83 (95 % CI = 0.74-0.92). Heterogeneity of pooled sensitivity and specificity was estimated as  $I^2=11.11$  %. Although this study was limited by small sample size and different analytical platforms, the proposed future framework resolves such limitations and standardizes future research. It is reasonable to deduce that VOCs breath analysis is certainly a field of research that can progress to replace traditional methods within the framework of CRC screening and diagnosis.

**Keywords:** Colorectal cancer, volatile organic compounds, exhaled, breath analysis

<sup>1</sup> We declare that our abstract was presented at a conference as an e-poster and was subsequently published on BJS as seen below:

EP-575. Alsaadi D, Clements N, Gabuniya N, Chand M, Francis N. Exhaled volatile organic compounds in the detection of colorectal cancer: A systematic review and meta-analysis. *Brit J Surg* 2002;109 (Suppl 5):znac245.139. [doi: 10.1093/bjs/znac245.139](https://doi.org/10.1093/bjs/znac245.139). Published: 09 August 2022.

## INTRODUCTION

Colorectal cancer (CRC) is reported as the third most commonly diagnosed cancer and the second leading cause of cancer-related deaths worldwide (Rawla et al., 2019). Despite the introduction of colorectal screening programs, patient compliance has remained low worldwide (Wools et al., 2016). Colonoscopy remains the gold-standard investigative approach for the screening and diagnosing of CRC (CDC, 2020b). Complications, such as bleeding and perforation, as well as the necessity for bowel preparation are commonly weighed against the result, i.e. primarily reassurance (Rees et al., 2016). In 2018, The United States (US) Behavioural Risk Factor Surveillance System stated that approximately 21.7 million adults aged 50–75 years have never been screened for CRC (CDC, 2020a). Some screening programs employ the use of faeces-based non-invasive tests prior to conducting a colonoscopy. The current fast-track colonoscopy referral pathway in the UK puts forward approximately 300,000 suspected patients for screening annually (Thornton et al., 2016). Due to the non-specific nature of CRC symptoms, the majority of these patients are screened via colonoscopy within two weeks and only 3 %–10 % are diagnosed with CRC. Consequently, there is a need for novel non-invasive CRC screening tests that are more acceptable to patients and that can reliably detect CRC or reduce the number of unnecessary colonoscopies performed in cancer-free patients.

Theoretically, different diseases can be categorized by unique metabolomic profiles (Hanna et al., 2019). Volatile organic compounds (VOCs; carbon-based compounds that mainly comprise alkanes, benzenes and aldehydes) are a subtype of metabolites that are linked to oxidative stress and cell-membrane peroxidation. Reference to VOCs in exhaled breath was first made by Pauling and his colleagues in 1971 (Pauling et al., 1971). They can be detected in several types of biological samples such as faeces, urine, serum, sputum and in the breath. These carbon compounds can be affected by both internal and

external factors such as smoking (Reade, 2016). Cancer is a metabolic disease that uses cellular metabolic changes to maintain a high rate of proliferation. Gene mutations and protein changes in the tumor cells lead to an oxygenation of the polyunsaturated fatty acids in the cell membranes and thus to a change in the VOCs in cancer patients (Altomare et al., 2016). Although the detection of disease through breath analysis is an emergent field of research, it can be traced back to approximately 44 BC. Hippocrates taught his students to smell the breath of their patients and to pour human saliva on hot coals to identify their ailments (Reade, 2016). In the 1980s, urea breath tests began to be used clinically in the diagnosis of *Helicobacter pylori*-related gastritis (Graham and Miftahussurur, 2018).

The current paper considers breath analysis for the detection of CRC, as its painless and non-invasive nature makes it a favorable method that will increase patient compliance in the future. A review of the current literature indicated that the bulk of the available literature instead considered VOCs without a focus on a specific cancer subtype. Studies that did report specific data for CRC and exhaled VOCs were small in size. Recognizing that these studies may individually represent limited data for impacting clinical practice on their own and considering that no diagnostic meta-analysis or systematic review had previously been conducted to integrate these data and derive relevant conclusions, a systematic appraisal and meta-analysis of the published evidence was conducted in the current paper to determine whether the analysis of VOCs in the breath could enable the detection of CRC and whether it held value as a non-invasive screening method.

## MATERIALS AND METHODS

### *Guidelines*

The Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines were followed while conducting this study and while reporting this systematic review. Concurrently, the Meta-analysis of Observational Studies in Epi-

demology (MOOSE) guidelines were followed to ensure the highest possible meta-analysis quality. The PRISMA and MOOSE checklists are included in the supplementary information, Appendix A. Obtaining patient consent and approval from a research ethics committee was not required to conduct this study because this systematic review and meta-analysis were based on previously published studies. Nonetheless, established ethical standards were followed. This manuscript received no funding, has no conflict of interests to disclose and was not registered.

### ***Search strategy***

The search strategy was validated by a librarian (JS) at University College London (UCL). The search was performed independently by two investigators (DA and NC). Nine electronic databases were searched starting at their inception date, i.e. Proquest, Pubplus, Cochrane, Google Scholar, Ovid Embase Classic + Embase, Ovid MEDLINE and In-Process and Other Non-Indexed Citations, Ovid MEDLINE, the US National Library of Medicine database (Pubmed) and Scirus. Furthermore, the search was supplemented using the UCL explore library (which included 719 databases). The initial search was conducted on 15 July 2020 and repeated on 15 August 2020 to identify additional literature published since the first search. The full list of search strings utilized is reported in the supplementary information, Appendix B.

### ***Eligibility criteria***

All relevant articles were retrieved without geographic, size, date, or outcome limitations. Eligibility criteria included: (1) articles that included segregated quantitative data on patients with CRC; (2) studies that included analyzed VOCs, particularly within the exhaled breath of participants for diagnosing CRC; (3) studies were written in the English language; (4) peer-reviewed journal articles, conference abstracts and posters with sufficient segregated data; (5) studies must have reported at least one or more of the following types of quantitative data: sensitivity,

specificity, accuracy, recognition, true positive (TP), false negative (FN), false positive (FP), true negative (TN), number of CRC and healthy control (HC) participants and characteristic and demographic factors including age, smoking and gender. Papers were excluded if: (1) studies were performed on non-human subjects; (2) the patients studied had secondary CRC or surveillance; (3) studies that failed to provide any quantitative data. Meta-analysis criteria differed slightly in that both sensitivity and specificity had to have been reported in addition to the number of CRC patients and those in the HC group.

### ***Study screening***

Mendeley was used for screening. The process of screening was conducted independently by the same investigators previously noted (DA and NC). An initial article title screen was conducted within all databases during the electronic search to eliminate obviously irrelevant articles. They were then imported into a joint Mendeley account by both assessors. Next, the two reviewers independently conducted a secondary title screen, which was followed by an abstract screen and a final full-text screen as per the previously determined eligibility criteria. Screening decisions were then cross-matched, and any discrepancies were resolved during a panel discussion with a third and fourth reviewer (SN and CP) and a consensus was achieved.

### ***Descriptive and methodologic analysis***

The extracted data were categorized according to factors that could potentially influence exhaled VOC levels. A list of all data extracted can be found in the supplementary information, Appendix C. These aspects were classified into the following domains: (1) patient selection criteria for the CRC and HC group; (2) patient characteristics and demographic factors including gender, smoking and CRC stage; (3) the sampling and collection methods; (4) the analytical platforms and processes used; (5) other specific variability sources.

### **Quality analysis**

The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool was used to assess the risk of bias and applicability and quality of the articles included in this review. Additionally, the Standards for Reporting of Diagnostic Accuracy Studies (STARD) checklist was used to ensure the full transparency and completeness of the studies.

### **Statistical analysis**

The purpose of conducting a meta-analysis was to assess the strength of the evidence found in these articles; this was done by combining their results to derive a pooled weighted estimate with a higher statistical power. Articles that provided both sensitivity and specificity were analyzed using two mathematical methods. The first approach calculated the overall mean of the sensitivities and specificities stated in the respective studies. The second approach involved the retrograde mathematical calculation of the given sensitivities and specificities to obtain the TP, TN, FP and FN of each study and the resulting pooled sensitivity and specificity. The Positive Predictive Value (PPV) and Negative Predictive Value (NPV) of each study were also determined using the TP, TN, FP and FN obtained from the retrograde calculations. The confidence interval (CI) for the TP, TN, FP, FN, PPV and the NPV of each study were obtained using the simple asymptotic with continuity correction formula. In addition, the percentage of the variability that was due to heterogeneity was calculated using the quantifying inconsistency formula  $I^2 = 100 \% \times (Q - df)/Q$ , where Q and df are the Cochran chi-squared statistic and degrees of freedom, respectively. A forest plot of the sensitivity and specificity was created to illustrate the results obtained in this study.

## **RESULTS**

### **Literature search**

Following the initial title screening of the nine databases and the UCL explore library database, 208 articles were imported into Mendeley; 197 articles were assessed in the

secondary title screening after 11 duplicates were removed. A total of 121 articles were excluded after the secondary title screening for reasons including the following: irrelevance (102 articles); the type of paper (12 articles); language (one article); animal study (two articles); other types of VOCs (four articles). Subsequently, the abstract screening eliminated an additional 11 papers for reasons including the following: irrelevance (six articles); type of paper (three articles); CRC recurrence (one article); no quantitative data (one article). The remaining 65 articles were screened for eligibility as per the inclusion and exclusion criteria, leading to the exclusion of 51 articles for reasons including the following: irrelevance (36 articles); included no quantitative data (eight articles); other types of VOCs (two articles); type of study (two articles); animal study (one article); CRC recurrence and follow up (two articles). Hence, this systematic review included a total of 14 articles that passed our eligibility criteria in the final full-text screening (Altomare et al., 2013, 2016; Amal et al., 2016; Depalma et al., 2014; De Vietro et al., 2020; Di Lena et al., 2012; Leja et al., 2015; Markar et al., 2019; Nakhleh et al., 2017; Peng et al., 2010; Sonoda et al., 2011; Van Keulen et al., 2020; Wang et al., 2014; Zambrana et al., 2012). Among the included articles, nine were included in the meta-analysis because they provided both sensitivity and specificity values (Altomare et al., 2013, 2016; Amal et al., 2016; Di Lena et al., 2012; Leja et al., 2015; Markar et al., 2019; Sonoda et al., 2011; Van Keulen et al., 2020; Zambrana et al., 2012). A PRISMA flow diagram summarizing the systematic search results and the screening process is illustrated in Figure 1.

### **The overall characteristics of the selected studies and quality assessment**

As shown in Table 1, the articles were published between the years of 2010 and 2020, 9 were based in Europe, while 5 were carried out in West and East Asia. In total, 1,416 participants were assessed, 552 of whom were CRC patients and 864 represented



### PRISMA 2009 Flow Diagram

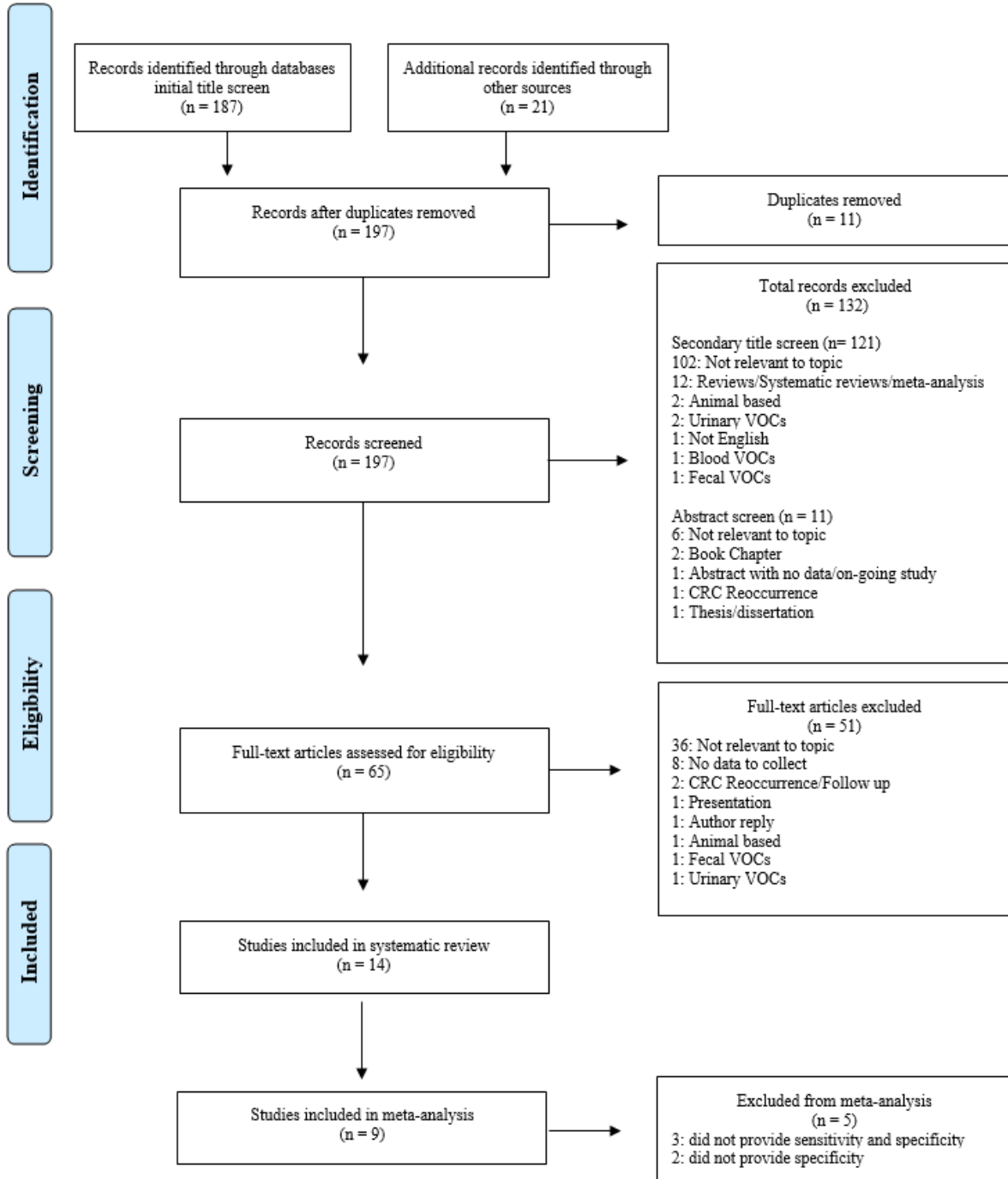


Figure 1: PRISMA flow diagram

**Table 1:** Overall characteristics, data, and STARD scores

Ar- ti- cle	First author	Year	Loca- tion	Number of participants		VOCs participants		Mean age (yr)		CRC Gender		HC Gender		Number of smokers		Number of non-smokers		CRC Stage	Sensitiv- ity	Specif- icity	Analytical platform	Collection method	Publica- tion type	STARD	SR	MA	
				CRC	HC	CRC	HC	CRC	HC	Mal es	Fem- al es	Mal es	Fem- al es	CRC	HC	CRC	HC										
1	Di Lena	2012	Italy	34	36	34	36	NR	NR	NR	NR	NR	NR	0	0	34	36	NR	0.8	0.9	TD-GCMS	Inert Bag (Tedlar)	Abstract	-	Y	Y	
2	Zam- brana	2012	Spain	38	43	38	43	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	0-IV	0.9	0.8	SESI-MS	Direct into system	Poster	-	Y	Y	
3	Amal	2016	Lativa Israel	65	122	20	36	66	60	41	24	31	91	5	20	60	102	0-IV	0.9	0.9	GC-MS	GaSampler collection bags	Case control	20	Y	Y	
4	Leja	2015	Lativa Israel	71	131	63	131	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	0.9	0.9	GC-MS + GNP	NR	Abstract	-	Y	Y	
5	Sonoda	2011	Japan	33	132	33	132	71	65	15	18	64	68	3	28	30	104	0-IV	0.9	1	Canine scent	Breath sampling bag	Case c ontrol	18	Y	Y	
6	Wang	2014	China	20	20	20	20	58	50	13	7	8	12	5	8	15	12	I-IV	NR	NR	SPME- GC-MS	Intert Bag (Tedlar)	Case control	13	Y	N	
7	Depalma	2014	Italy	15	15	15	15	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	1	NR	TD-GCMS	Intert Bag (Tedlar)	Abstract	-	Y	N	
8	Van Keulen	2020	Neth- er- lands	70	128	62	104	68	62	43	27	66	62	41	78	29	50	I-IV	1	0.6	E-Nose	Direct into system	Cross-sec- tional study	24	Y	Y	
9	De Vietro	2020	Italy	7	20	7	20	65	60	4	3	8	12	NR	NR	NR	NR	III	NR	NR	GC-MS	direct into ReCIVA breath sam- pler	Case control	19	Y	N	
10	Altomare	2016	Italy	15	15	15	15	68	62	8	7	9	6	NR	NR	NR	NR	I-III	0.9	0.1	E-Nose	Inert Bag (Tedlar)	Case control	10	Y	Y	
11	Altomare	2013	Italy	37	41	37	41	63	47	20	17	13	28	NR	NR	NR	NR	I-IV	0.9	0.8	TD-GCMS	Inert Bag (Tedlar)	Case control	20	Y	Y	
12	Peng	2010	Israel	26	22	26	22	NR	NR	NR	NR	NR	NR	15	4	11	18	I-IV	<0.3	NR	SPME- GC-MS / GNP	Mylar sam- pling bag	Case control	14	Y	N	
13	Markar	2019	UK	50	50	50	50	NR	NR	30	20	22	28	15	18	35	32	0-IV	1	0.8	SIFT-MS	Nalophan bags	Case control	23	Y	Y	
14	Nakhleh	2017	Multi- center EU	71	89	71	89	66	60	42	29	67	22	9	9	62	80	NR	NR	NR	GC-MS	Mylar sam- pling bag	Cross-sec- tional study	16	Y	N	
	<b>Total</b>	-	-	<b>552</b>	<b>864</b>	<b>491</b>	<b>754</b>	-	-	<b>216</b>	<b>152</b>	<b>288</b>	<b>329</b>	<b>93</b>	<b>165</b>	<b>276</b>	<b>434</b>	-	-	-	-	-	-	-	<b>14</b>	<b>9</b>	
	<b>Mean (Stand- ard de- viation)</b>	-	-	-	-	-	-	<b>65.6 ± 3.9</b>	<b>58.3 ± 6.3</b>	-	-	-	-	-	-	-	-	-	<b>0.9 ± 0.1</b>	<b>0.8 ± 0.3</b>	-	-	-	-	<b>17.7 ± 4.5</b>		

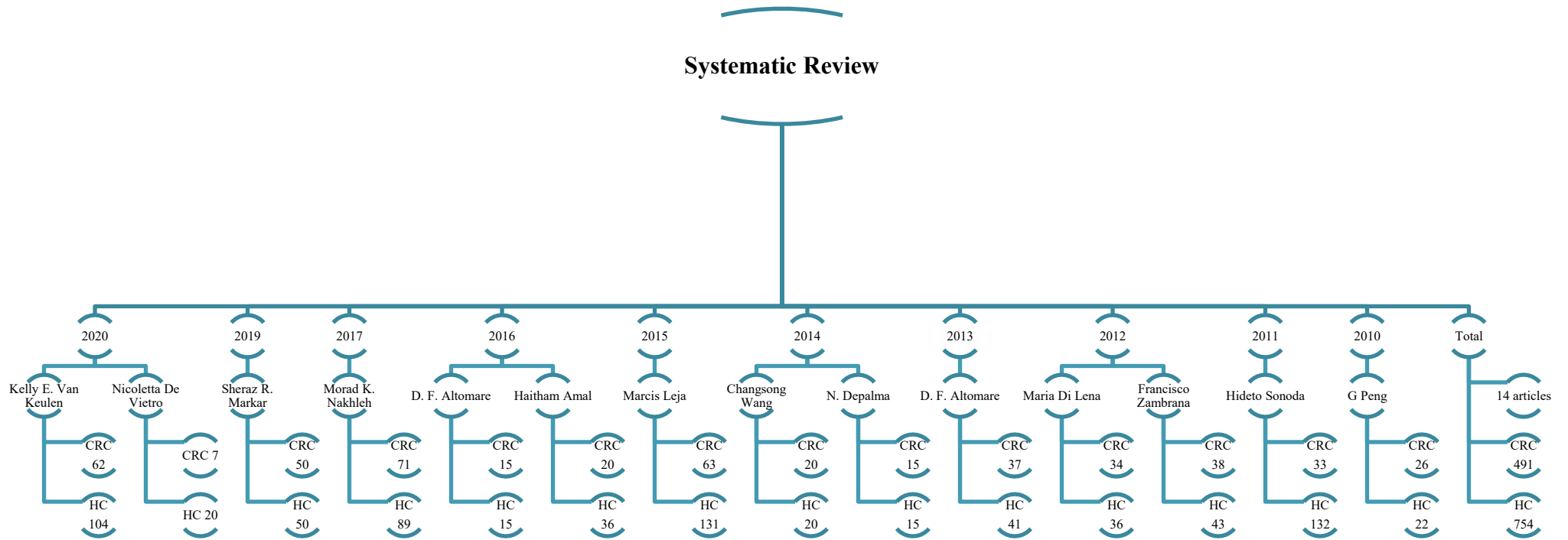
CRC: colorectal cancer; HC: healthy control; STARD: Standards for Reporting of Diagnostic Accuracy Studies; SR: Systematic Review; MA: Meta-Analysis

the HC group. However, further inspection of the studies revealed that only 491 CRC patients and 754 HC participants had been sampled for VOC testing. A diagram of the included articles and the number of CRC and HC participants is shown in Figure 2. The mean age of the CRC subjects was 65.62 (58–71) years, although six of the included articles failed to provide participant ages (Depalma et al., 2014; Di Lena et al., 2012; Leja et al., 2015; Markar et al., 2019; Peng et al., 2010; Zambrana et al., 2012). Among the CRC cohort, the number of male patients was higher ( $n = 216$ ) compared with the female group ( $n = 152$ ); however, the gender of 184 CRC patients was unreported. Four studies used the American Joint Committee on Cancer's 0–IV full staging system (Amal et al., 2016; Markar et al., 2019; Sonoda et al., 2011; Zambrana et al., 2012) whereas stage 0 (carcinoma in situ) was not included in five of the other studies (Altomare et al., 2013, 2016; Peng et al., 2010; Van Keulen et al., 2020; Wang et al., 2014). The remaining five studies did not provide staging details but did confirm that the patients had CRC, either by colonoscopy or histology (Depalma et al., 2014; Di Lena et al., 2012; Leja et al., 2015; Nakhleh et al., 2017). Only 93 CRC patients were declared as current or ex-smokers and 276 participants had never smoked; six articles did not disclose smoking status ( $n = 183$ ) (Altomare et al., 2013; Depalma et al., 2014; De Vietro et al., 2020; Leja et al., 2015; Van Keulen et al., 2020; Zambrana et al., 2012). The HC participants included 288 males and 329 females; the gender of 247 participants was not stated. The mean age of the HC cohort was 58.25 (47–65) years. Among the HC participants, 165 were put forth as smokers while 434 were classified as non-smokers with the remainder ( $n = 265$ ) unreported. The overall characteristics and data of the included studies are illustrated in Table 1.

Ten of the above studies stated their recruitment criteria for the control group. All studies provided assurances that these participants had received colonoscopies; however, some studies stated that these had been

performed within the last two years while others did not report a time range (Nakhleh et al., 2017). Six studies excluded patients with inflammatory bowel diseases, malignancies and those who had undergone recent gastrointestinal surgery or had a history of chemo or radiotherapy (Altomare et al., 2013, 2016; Amal et al., 2016; Nakhleh et al., 2017; Sonoda et al., 2011; Van Keulen et al., 2020). Conversely, one study was more specific and excluded those who had an active infection and liver disease (Markar et al., 2019), while another study excluded those who had mental health or who were pregnant (Wang et al., 2014).

Concerning breath sampling and collection methods, most studies collected these samples indirectly through collection bags that were subsequently transferred into vacuums, syringes, or glass vials except for three studies that had patients exhale directly into analyser systems (De Vietro et al., 2020; Van Keulen et al., 2020; Zambrana et al., 2012). However, breath samples were collected under different specified conditions. Eight of these studies specified that participants were required to fast prior to sampling, ranging from 2–12 h (Altomare et al., 2013, 2016; Amal et al., 2016; De Vietro et al., 2020; Markar et al., 2019; Nakhleh et al., 2017; Wang et al., 2014; Zambrana et al., 2012). Two studies prohibited smoking 2 h before sampling (Amal et al., 2016; Nakhleh et al., 2017). Another study required 12 h of withholding the ingestion of alcohol and coffee (Peng et al., 2010). Three studies did not report such conditions. Moreover, one study had a protocol stated for excluding inadequate breath samples while others did not (Van Keulen et al., 2020). Another condition to consider was the environment in which the breath sample was taken. Five of these studies disclosed that breath samples had been taken in the same area (Altomare et al., 2013; De Vietro et al., 2020; Markar et al., 2019; Peng et al., 2010; Van Keulen et al., 2020), two of which kept participants in the room for at least 10 to 20 min before breath collection (Altomare et al., 2013; Markar et al., 2019) and



**Figure 2:** Included articles and number of CRC and HC participants

CRC:Colorectal cancer; HC: Healthy Control



one study had the participants' breath filtered by medical air (De Vietro et al., 2020). Other factors included breath sample filtration and the quality of the exhaled breath. Four studies filtered residual environmental contaminants from the breath samples via flushing with nitrogen gas (Wang et al., 2014) and by installing a filter (charcoal, carbon and bacterial) cartridge (Nakhleh et al., 2017; Peng et al., 2010; Van Keulen et al., 2020). Concerning the quality of the breath samples, two studies had participants repeatedly inhaling for 3–5 min to reach total lung capacity (Amal et al., 2016; Peng et al., 2010). Other studies instructed participants to simply take a single deep nasal inhalation ( $n = 1$ ) and to take five consecutive breaths ( $n = 1$ ) (Markar et al., 2019; Van Keulen et al., 2020). One study collected breath samples specifically when the carbon dioxide level exceeded 3 % (Altomare et al., 2016). Finally, only one study declared that breath samples were taken from CRC patients prior to surgery (Markar et al., 2019).

In relation to the analytical platforms, eight studies used Gas chromatography–mass spectrometry (GC-MS); three of these studies coupled GC-MS with thermal desorption (TD-GC-MS) (Altomare et al., 2013; Depalma et al., 2014; Di Lena et al., 2012), while two studies integrated solid-phase micro-extraction (SPME-GC-MS) (Peng et al., 2010; Wang et al., 2014). The remaining studies utilized other analytical platforms including electronic nose ( $n = 2$ ) (Altomare et al., 2016; Van Keulen et al., 2020), canine scent ( $n = 1$ ) (Sonoda et al., 2011), secondary electrospray ionisation (SESI-MS) ( $n = 1$ ) (Zambrana et al., 2012), selected ion flow tube (SIFT-MS) ( $n = 1$ ) (Markar et al., 2019) and a cross-reactive nanomaterial-based sensor with gold nanoparticles (GNP) ( $n = 1$ ) (Leja et al., 2015). Additionally, one study compared the performance of SPME-GC-MS against functionalized GNPs (Peng et al., 2010).

Other sources of variability included the different methods and protocols that the studies employed to classify and analyze VOCs in the breath samples. The leave-out method was

employed by eight studies and these were further classified by either the support vector machine method or linear discriminant analysis (Altomare et al., 2013, 2016; Amal et al., 2016; Depalma et al., 2014; Di Lena et al., 2012; Nakhleh et al., 2017; Van Keulen et al., 2020; Zambrana et al., 2012). The remaining studies used different methods as follows: pattern recognition methods ( $n = 3$ ) (Amal et al., 2016; Leja et al., 2015; Nakhleh et al., 2017), principal component analysis ( $n = 3$ ) (Altomare et al., 2016; Peng et al., 2010; Wang et al., 2014), a probabilistic neural network ( $n = 2$ ) (Altomare et al., 2013, 2016) and external VOC databases ( $n = 3$ ) (De Vietro et al., 2020; Wang et al., 2014; Zambrana et al., 2012). Moreover, some studies utilized artificial intelligence to run a predictive algorithm module that selected a small set of VOCs. Another factor that contributed to inconsistency was the length of time it took to analyze the breath samples from the time of breath collection. One study stated that samples were analyzed within 1 h ( $n = 1$ ) (Markar et al., 2019) after collection, whereas another study analyzed the samples four days after collection (Peng et al., 2010). The remainder of the studies did not disclose these details.

Quality assessment tests were performed on 10 of the included studies, four studies were excluded due to publication type (i.e. abstracts/posters) (Depalma et al., 2014; Di Lena et al., 2012; Leja et al., 2015; Zambrana et al., 2012). The STARD scores for the included articles ranged from 10 to 24 with an average of 17.7. The final scores of the STARD for each study are shown in Table 1 and the full details of the scoring process are provided in the supplementary information, Appendix D. The QUADAS-2 test revealed that patient selection may have been a major source of bias because all studies were indicated as having had a high or unclear risk; this was due to the failure of providing recruitment criteria and due to the absence of a positive control group. Likewise, the absence of a validation set for the index test in selected studies, as well as the overall lack of clarity regarding the flow and timing also suggested

a risk of bias. Consequently, the overall applicability quality of the included studies was high; however, the studies were at high risk of including bias. The QUADAS-2 tool results are summarized in Table 2.

### Meta-analysis

Subgroup meta-analysis was performed using both statistical methods that were previously described and the results are shown in Table 3. Using the first approach, the pooled mean sensitivity and specificity of nine studies was calculated as 0.89 and 0.76, respectively (Altomare et al., 2013, 2016; Amal et al., 2016; Di Lena et al., 2012; Leja et al.,

2015; Markar et al., 2019; Sonoda et al., 2011; Van Keulen et al., 2020; Zambrana et al., 2012). Using the second approach, the estimated diagnostic accuracy was measured using sensitivity, specificity, PPV, NPV, positive likelihood ratio (PLR), negative likelihood ratio (NLR) and the diagnostic odds ratio (DOR). Pooled sensitivity was 0.89 (95 % CI = 0.80–0.99) and specificity was 0.83 (95 % CI = 0.74–0.92). Heterogeneity in the pooled sensitivity was low at  $I^2 = 11.11\%$ ,  $P = 0.243$ , while the heterogeneity of the pooled results for specificity was also low at  $I^2 = 11.11\%$ ,  $P = 0.230$ . Mean PPV, mean NPV,

**Table 2:** The QUADAS-2 tool

Study	Risk of Bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Di Lena, 2012							
Zambrana, 2012							
Amal, 2016	●	●	●	?	●	●	●
Leja, 2015							
Sonoda, 2011	?	●	●	?	●	●	●
Wang, 2014	●	?	●	●	●	●	●
Depalma, 2014							
van Keulen, 2020	?	●	●	●	●	●	●
De Vietro, 2020	●	?	●	●	●	●	●
Altomare, 2016	●	●	●	?	●	●	●
Altomare, 2013	●	●	●	?	●	●	●
Peng, 2010	?	?	●	●	●	●	●
Markar, 2019	●	●	●	?	●	●	●
Nakhleh, 2017	?	●	●	?	●	●	●

PLR, NLR and DOR were 0.76 (95 % CI = 0.63–0.88), 0.93 (95 % CI = 0.86–0.99), 5.22 (95 % CI = 3.03–8.99), 0.13 (95 % CI = 0.05–0.32) and 42.95 (95 % CI = 12.67–145.62), respectively. A forest plot of the sensitivities and specificities calculated using the second approach was created to illustrate the results obtained in this study (see Figure 3).

### ***The performance of individual volatile organic compounds***

The individual VOCs detected in the breath of CRC patients were reported in nine studies with a combined total of 66. A quantitative summary of the VOCs that were found in at least three articles or more is illustrated in Table 4. The total number of CRC patients tested from the relevant articles was calculated to weigh the compound’s significance. Ethanol had the highest significance as it was reported by four studies with a total of 204 CRC breath samples (Amal et al., 2016; Leja et al., 2015; Markar et al., 2019; Nakhleh et al., 2017). Acetone was detected in the breath of 161 samples from four studies in total (Amal et al., 2016; De Vietro et al., 2020; Leja et al., 2015; Nakhleh et al., 2017). Three studies with a combined 154 participants found ethyl acetate (Amal et al., 2016; Leja et al., 2015; Nakhleh et al., 2017). Likewise, 4-methyloctane and nonanal were reported in three studies with 120 (Altomare et al., 2013; Amal et al., 2016; Leja et al., 2015) and 115 (Altomare et al., 2013; De Vietro et al., 2020; Nakhleh et al., 2017) participants, respectively. An additional comparison was made to

discover how many of the most frequently reported compounds were noted in each study. The studies conducted by Amal (2016) and Leja (2015) reported ethanol, acetone, ethyl acetate and 4-methyloctane (Amal et al., 2016; Leja et al., 2015). The study conducted by Nakhleh (et al., 2017) reported ethanol, acetone, ethyl acetate and nonanal (Nakhleh et al., 2017). These three studies employed GC-MS. A complete and detailed table of all the reported VOCs is provided in the supplementary information, Appendix E.

## **DISCUSSION**

An emerging number of studies have demonstrated the potential value of exhaled VOCs as a diagnostic and triaging test for CRC; however, their use in clinical practice is yet to be observed (Altomare et al., 2013, 2016; Amal et al., 2016; De Vietro et al., 2020; De Palma et al., 2014; Di Lena et al., 2012; Leja et al., 2015; Markar et al., 2019; Nakhleh et al., 2017; Peng et al., 2010; Sonoda et al., 2011; Van Keulen et al., 2020; Wang et al., 2014; Zambrana et al., 2012). A study conducted by Altomare showed that the pattern of VOCs changed following the removal of CRC, thereby confirming a close relationship between tumor metabolism and exhaled VOCs (Altomare et al., 2013). Similarly, a study conducted by Wang was able to detect nine VOCs with increased concentration in breath samples of CRC patients (Wang et al., 2014). Additional research was able to identify higher levels of acetone and ethyl ace-

**Table 3:** Subgroup analysis on exhaled VOCs based on two statistical approaches

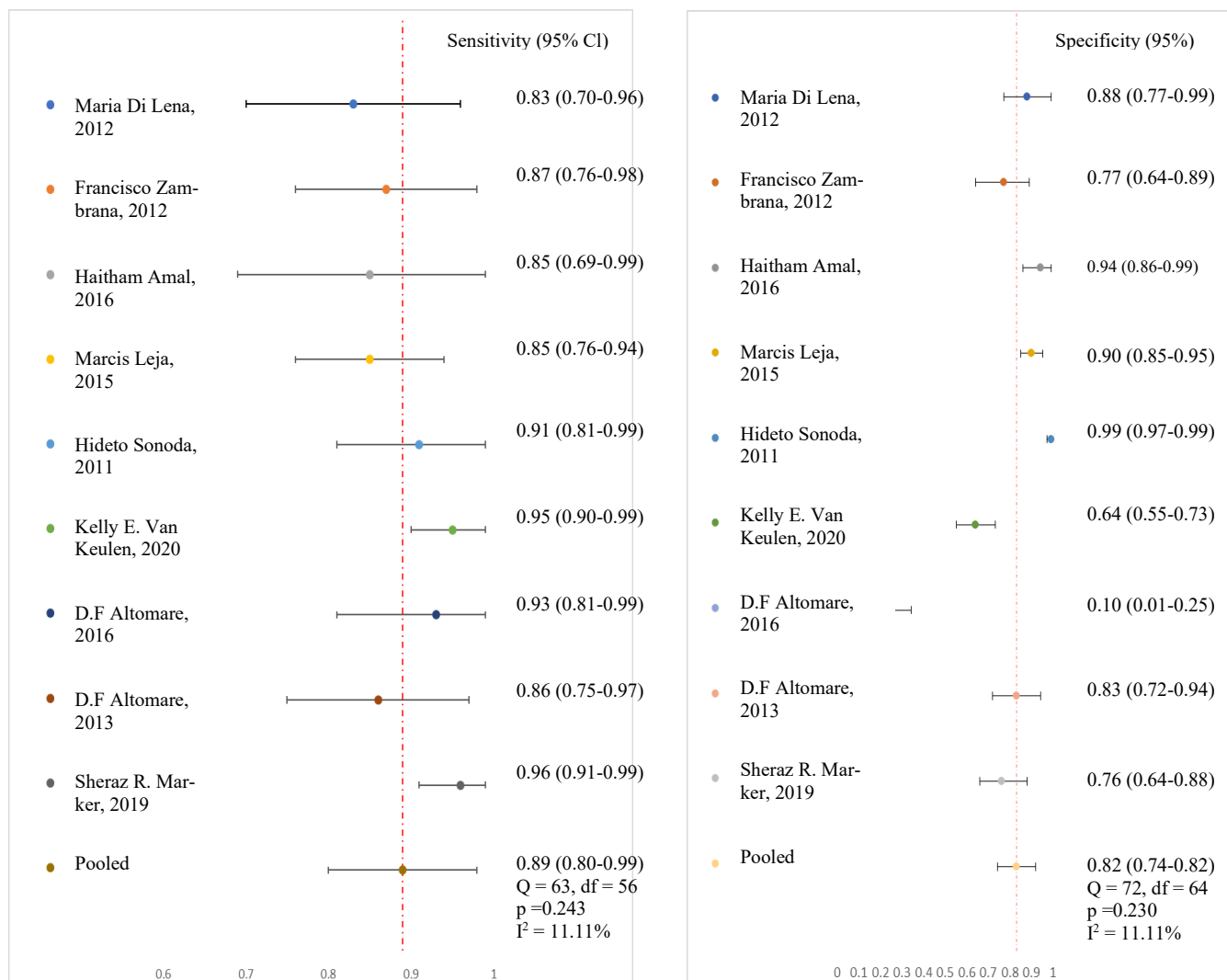
#### *First Approach*

Datasets, n	CRC, n	HC, n	Sensitivity	Range of mean	Specificity	Range of mean
<b>9</b>	413	698	0.89	0.83-0.96	0.76	0.1-0.99

#### *Second Approach*

Da-tasets, n	CRC n	HC n	Mean TP	Mean FN	Mean TN	Mean FP	Mean sensi-tivity	Mean speci-ficity	Mean PPV	Mean NPV	Mean DOR
<b>9</b>	413	698	34.95	4.16	54.14	11.20	0.89	0.83	0.76	0.93	42.95

N= Total Number, CRC= Colorectal Cancer, HC= Healthy Controls, TP= True Positive, FN= False Negative, TN= True Negative, FP= False Positive, PPV= Positive Predictive Value, NPV= Negative Predictive Value, DOR= diagnostic Odds Ratio



**Figure 3:** A forest plot of the sensitivities and specificities calculated using the second approach  
 CI= Confidence Interval, Q= Cochran Chi-squared statistic, df= Degrees of freedom , and I<sup>2</sup>= Inconsistency calculated by 100 % × (Q - df)/Q

**Table 4:** A quantitative summary of the commonly reported VOCs compounds

	Zambrana, 2012	Amal, 2016	Leja, 2015	Wang, 2014	De Vietro, 2020	Altomare, 2013	Peng, 2010	Markar, 2019	Nakhleh, 2017	Number of articles	Number of patients
<b>Analysis platform</b>	SESI-MS	GC-MS	GC-MS + GNP	SPME-GC-MS	GC-MS	TD-GCMS	SPME-GC-MS/	SIFT-MS	GC-MS		
<b>CRC participants</b>	38	20	63	20	7	37	26	50	71		
<b>Ethanol</b>		+	+					+	+	4	204
<b>Acetone</b>		+	+		+				+	4	161
<b>Ethyl acetate</b>		+	+						+	3	154
<b>4-methyloctane</b>		+	+			+				3	120
<b>Nonanal</b>					+	+			+	3	115
<b>Total</b>		4	4		2	2		1	4		

tate in CRC patients (Amal et al., 2016). More recently, in a study conducted by Sonoda, dogs were trained to distinguish CRC from healthy controls using exhaled breath samples with 91 % sensitivity and 95 % specificity results (Sonoda et al., 2011). Multiple analytical platforms are used when attempting to analyze VOCs. GC-MS allows for volatile molecules to be physically separated and identified and for the exact composition of the sample to be determined. SPME can be utilised for sample pre-concentration when using GC-MS (Reade, 2016). In this process, polymer-coated fiber is used to absorb compounds from patient samples. Other methods of extraction that utilize mass spectrometry include SESI-MS and SIFT-MS, and cross-reactive nanomaterial-based sensor with GNPs (Reade, 2016).

This paper provides the first comprehensive and up-to-date review of the performance of exhaled VOCs in CRC detection. The given amalgamated results of the meta-analysis support the general hypothesis with a pooled sensitivity of 0.89 (95 % CI = 0.80–0.99) and specificity of 0.83 (95 % CI = 0.74–0.92). In addition, mean PPV and mean NPV were 0.76 (95 % CI = 0.63–0.88) and 0.93 (95 % CI = 0.86–0.99), respectively. These results imply that VOCs can be used to differentiate between patients with CRC relatively accurately, or at the very least, reliably differentiate healthy patients from those with a gastrointestinal pathology. The DOR is a single indicator of test accuracy; in this review, its result was 42.95 (95 % CI = 12.67–145.62) implying a good discriminatory test performance. The PLR was 5.5 (95 % CI = 3.03–8.99) indicating that a CRC patient is 5.5 times more likely to have a positive result when compared with a healthy patient. Similarly, the NLR was 0.13 (95 % CI = 0.05–0.32) implying that a patient with a negative result had a 9.7-fold decrease in the odds of having CRC. The heterogeneity of the included articles was  $I^2 = 11.11\%$  for pooled sensitivity and specificity, which suggested a low level of heterogeneity. Selected interpretations can be made from these results. That

is, VOC analysis can be used to screen for CRC but it is not yet sufficiently conclusive to diagnose the condition on its own. Its use prior to invasive colonoscopies may, however, reduce the number of false positives in screening programs.

### **Limitations**

There were limited published studies on the subject and those that were included were either case-control or cross-sectional studies that included small participant samples ranging from 27 to 194 which reduces the statistical impact of the results. The selection criteria for the healthy controls' cohort differed between studies; some included only those with a completely negative colonoscopy while others included those with adenomas, bowel diseases or did not disclose such information. Similarly, these studies included CRC patients in different stages of the disease. Furthermore, the impact of factors such as age, gender and smoking status remained undetermined because a separate analysis in this regard was not reported by the studies. Studies varied significantly in terms of the collecting, handling, storing and analysis of their samples. Additionally, the instructions and protocols provided before collecting a sample varied between studies. Although the same breath collection bags were used in ten of these studies their storing time varied. Three studies that used the GC-MS method detected the four most common VOCs (Amal et al., 2016; Leja et al., 2015; Nakhleh et al., 2017). The types of VOCs detected could be dependent on the analytical methods and processing techniques used, which dictates that the findings are partial to that.

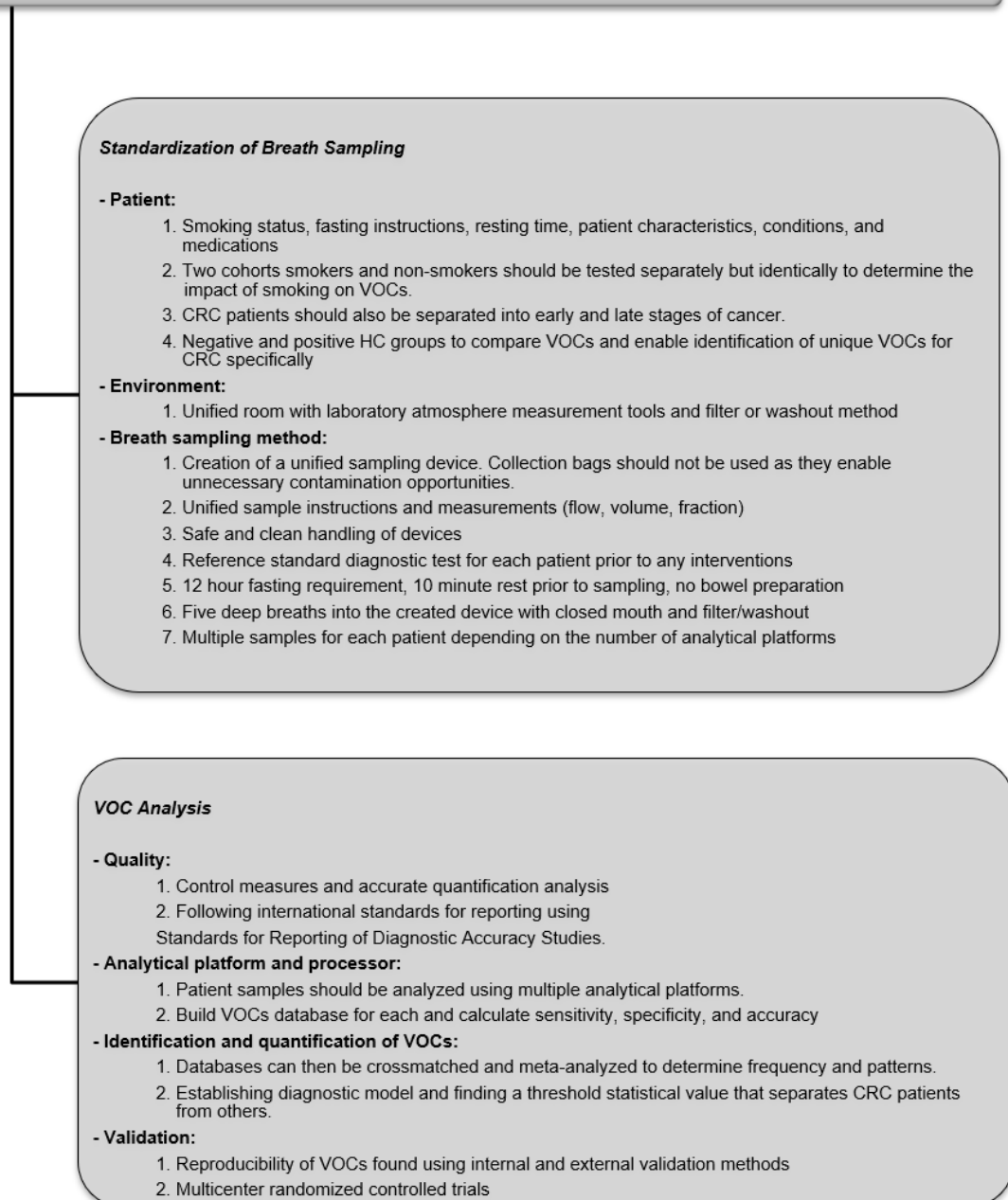
### **Future development**

The standardization of future research is warranted for ensuring its progression and advancement into clinical practice. To achieve this, extensive and well-controlled comparison studies are required to determine the superior analytical platform, processor and collection method. Confounding heterogeneity factors also require further study, particularly

smoker vs non-smoker breath samples. Future studies must use larger sample sizes to confirm reliability. A proposed framework for conducting and reporting future studies on VOC breath testing in CRC patients was

constructed, based on existing studies and the findings of this review in conjunction with STARD and QUADAS-2 checklists. The framework is illustrated in Figure 4.

**Figure 4:** Proposed framework for conducting and reporting future studies evaluating the role of VOCs in CRC detection



## CONCLUSION

Breath analysis is a time-efficient, inexpensive, safe, painless, and non-invasive method that is likely to have high patient compliance rates. However, the small sample size, the inconsistency of data and different analytical platforms currently limit the advancement of this field. It is predicted that, with research progression and standardization, breath analysis can eventually be used as an alternative and non-invasive mass screening tool prior to conducting colonoscopies, thereby reducing the number of patients undergoing colonoscopies and lowering FP rates accordingly.

### *Author contributions*

Daniah Alsaadi - conceptualization, data collection and analysis, writing and reviewing.

Nicolle Clements; Natiya Gabuniya - data collection and analysis, reviewing.

Nader Francis; Manish Chand – Supervision and reviewing.

### *Acknowledgments*

Not applicable.

### *Conflict of interest*

The authors have no conflict of interest to disclose.

### *Funding*

The authors have no relevant financial or non-financial interests to disclose.

### *Ethical approval and consent to participate*

Ethical approval from the National Office for Research Ethics Committees and the University College London Guidance was not required.

## REFERENCES

Altomare DF, Di Lena M, Porcelli F, Trizio L, Travaglio E, Tutino M, et al. Exhaled volatile organic compounds identify patients with colorectal cancer. *Br J Surg*. 2013;100(1):144–50.

Altomare DF, Porcelli F, Picciariello A, Pinto M, Di Lena M, Caputi Iambrenghi O, et al. The use of the PEN3 e-nose in the screening of colorectal cancer and polyps. *Tech Coloproctol*. 2016;20:405–9.

Amal H, Leja M, Funka K, Lasina I, Skapars R, Sivins A, et al. Breath testing as potential colorectal cancer screening tool. *Int J Cancer*. 2016;138:229–36.

CDC, Centers for Disease Control and Prevention. Use of colorectal cancer screening tests among adults aged 50–75, by race/ethnicity and state, behavioral risk factor surveillance system, 2018. [Internet]. Atlanta: Centers for Disease Control and Prevention; [2020]. Available from: <https://www.cdc.gov/cancer/colorectal/statistics/use-screening-tests-BRFSS.htm>.

CDC, Centers for Disease Control and Prevention. Colorectal Cancer Screening. [Internet]. Atlanta: Centers for Disease Control and Prevention; [2020]. Available from: [https://www.cdc.gov/cancer/colorectal/basic\\_info/screening/index.htm](https://www.cdc.gov/cancer/colorectal/basic_info/screening/index.htm).

Depalma N, Di Lena M, Porcelli F, Travaglio E, Longobardi F, Demarinis Loiotile A, et al. Detection of colorectal polyps by exhaled VOCs. Preliminary data. *Tech Coloproctol*. 2014;18(1):92-3.

De Vietro N, Aresta A, Rotelli MT, Zambonin C, Lippolis C, Picciariello A, et al. Relationship between cancer tissue derived and exhaled volatile organic compound from colorectal cancer patients. Preliminary results. *J Pharm Biomed Anal*. 2020;180:113055.

Di Lena M, Porcelli F, Trizio L, Giuratrabocchetta S, Travaglio E, Tutino M, et al. Colorectal cancer screening by breath analysis: A specific pattern of volatile organic compounds (VOCs) can discriminate between patients and healthy controls. *Gastroenterology*. 2012;142(5):S-528.

Graham DY, Miftahussurur M. Helicobacter pylori urease for diagnosis of Helicobacter pylori infection: A mini review. *J Adv Res*. 2018;13:51–7.

Hanna GB, Boshier PR, Markar SR, Romano A. Accuracy and methodologic challenges of volatile organic compound-based exhaled breath tests for cancer diagnosis: A systematic review and meta-analysis. *JAMA Oncol*. 2019;5(1):e182815.

Leja M, Amal H, Funka K, Vanags A, Kikuste I, Sivins A, et al. Nanoarray sensor technology-based volatile marker tests to detect colorectal cancer and colonic adenomas. *Gastroenterology*. 2015;148(4):S17.

- Markar SR, Chin S-T, Romano A, Wiggins T, Antonowicz S, Paraskeva P, et al. Breath volatile organic compound profiling of colorectal cancer using selected ion flow-tube mass spectrometry. *Ann Surg.* 2019;269:903–10.
- Nakhleh MK, Amal H, Jeries R, Broza YY, Aboud M, Gharra A, et al. Diagnosis and classification of 17 diseases from 1404 subjects via pattern analysis of exhaled molecules. *ACS Nano.* 2017;11(1):112–25.
- Pauling L, Robinson AB, Teranishi R, Cary P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proc Natl Acad Sci U S A.* 1971;68:2374–6.
- Peng G, Hakim M, Broza YY, Billan S, Abdah-Bortnyak R, Kuten A, et al. Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. *Br J Cancer.* 2010;103:542–51.
- Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors. *Prz Gastroenterol.* 2019;14(2):89–103.
- Reade SP. Volatile organic compound profiling of mouse models of intestinal inflammation and colon carcinogenesis. PhD thesis. Liverpool, UK: University of Liverpool, 2016. <https://livrepository.liverpool.ac.uk/3001721/>.
- Rees CJ, Bevan R, Zimmermann-Fraedrich K, Rutter MD, Rex D, Dekker E, et al. Expert opinions and scientific evidence for Sonoda colonoscopy key performance indicators. *Gut.* 2016;65:2045–60.
- Sonoda H, Kohnoe S, Yamazato T, Satoh Y, Morizono G, Shikata K, et al. Colorectal cancer screening with odour material by canine scent detection. *Gut.* 2011;60:814–9.
- Thornton L, Reader H, Stojkovic S, Allgar V, Woodcock N. Has the 'Fast-Track' referral system affected the route of presentation and/or clinical outcomes in patients with colorectal cancer? *World J Surg Oncol.* 2016;14(1):158.
- Van Keulen KE, Jansen ME, Schrauwen RWM, Kolkman JJ, Siersema PD. Volatile organic compounds in breath can serve as a non-invasive diagnostic biomarker for the detection of advanced adenomas and colorectal cancer. *Aliment Pharmacol Ther.* 2020;51:334–46.
- Wang C, Ke C, Wang X, Chi C, Guo L, Luo S, et al. Noninvasive detection of colorectal cancer by analysis of exhaled breath. *Anal Bioanal Chem.* 2014;406:4757–63.
- Wools A, Dapper EA, Leeuw JRJD. Colorectal cancer screening participation: A systematic review. *Eur J Public Health.* 2016;26:158–68.
- Zambrana F, Herrero A, Vidal-de-Miguel G, Bailador G, Criado E, Sánchez-Ávila C, et al. On-line breath analysis of volatile organic compounds as a method for colorectal cancer detection. *J Clin Oncol.* 2012;30(15 Suppl):1570.