Deutsches Krebsforschungszentrum, Heidelberg

Wissenschaftlicher Vorstand: Prof. Dr. Michael Baumann

Abteilung Klinische Epidemiologie und Alternsforschung

Leiter: Prof. Dr. med. Hermann Brenner

Potential use of analysis of volatile organic compounds in exhaled breath for cancer screening

Inauguraldissertation

zur Erlangung des Doctor scientiarum humanarum (Dr. sc. hum.)
an der Medizinischen Fakultät Heidelberg
der Ruprecht-Karls-Universität

vorgelegt von:
Agnė Krilavičiūtė
aus Vilnius, Litauen

2017

`

Dekan: Prof. Dr. med. Wolfgang Herzog

Doktorvater: Prof. Dr. med. Hermann Brenner

`

Table of Contents

`

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIX

ABBREVIATIONS

AA – advanced adenoma

AUC – area under the receiver operating characteristic curve

BMI – body mass index

CI – confidence intervals

CRC – colorectal cancer

E-nose - “electronic nose”

FIT – fecal immunochemical tests for hemoglobin

GC – gastric cancer

GC-MS – gas chromatography – mass spectrometry

H. Pylori – Helicobacter Pylori

IQR – interquartile range

IUPAC –International Union of Pure and Applied Chemistry

M/Z – mass to charge ratio

N – number

NNS – number needed to screen

NPV – negative predictive value

PPV – positive predictive value

NSAIDs – non-steroidal anti-inflammatory drugs

RT – retention time

S/N – signal to noise ratio

TD – thermal desorption

UGE – Upper gastrointestinal endoscopy

VOC – volatile organic compound

Dissertation outlook

In this dissertation, the potential of exhaled breath analysis as non-invasive screening tool for cancer detection will be examined, focusing mostly on gastrointestinal cancers (colorectal and gastric cancer). The structure deviates from the usual structure of just one iteration of Methods, Results and Discussion, as the topic is relatively new with little standard guidelines making analyzed questions arise from each other rather than be focused on one specific research question. First, I briefly introduce currently employed cancer screening methods and challenges in their implementations, as well as exhaled breath analysis and its potential for cancer detection. Second, I systematically summarize available data on exhaled breath analysis for cancer detection and introduce the challenges exhaled breath analysis faces in this area. Next, I examine one of the challenges, i.e., the role of critical covariates in breath analysis based on breath samples of more than 1,400 healthy individuals. Third, focusing on colorectal and gastric cancers, I describe how the prevalence of preclinical (undetected) disease in screening populations affects expected true positive and true negative results from the screening tests and screening programs in general; and I examine the potential of breath testing to be applied as a screening tool in various populations. Finally, I summarize the topics above and provide suggestions for possible next steps for moving towards the use of breath analysis to improve the performance of cancer screening.

1. Background and introduction

1.1. Cancer screening

Besides better treatment, early detection is an essential component for reducing cancer mortality. However, differences in the nature of disease (e.g. lethality, sojourn time between preclinical and clinical stages (development speed)), availability of effective treatment, as well as invasiveness, cost, and logistics of potential screening tools all influence how screening may contribute to mortality reduction, and what potential harms such programs may be associated with.

On the population level, individual cancers are rare diseases, which usually makes the direct unselected screening of the entire asymptomatic population suboptimal not only due to unnecessary harms arising from the invasiveness of the screening procedures, but also due to high resource utilization and high cost. The screening of selected high-risk populations, or omitting from screening those who do not have the disease are promising approaches to improve the efficiency of a screening program, through improving uptake among those who should be screened, i.e. screening participation, and reducing resources needed as well as potential harms.

Today, there are only few cancers for which well-established screening methods are available. While screening for colorectal (CRC), breast and cervical cancer is widely accepted and established internationally, population-based screening for prostate cancer and gastric cancer (GC) is limited to only a few countries worldwide (78, 168, 172). For other malignancies, including one of the most frequent, lung cancer, currently no well-established, proven screening options exist.

Both CRC and GC mortality remain high in general – GC is the second and CRC the forth leading cancer cause of death worldwide (57), meaning that there remains ample potential to improve current screening programs and further reduce cancer mortality.

1.1.1 Screening for CRC and preselection of participants in CRC screening

CRC is suitable target for effective screening, due to its slow progress from pre-cancerous adenoma to clinically manifest disease (32), and the high diagnostic accuracy of colonoscopy, the gold standard for CRC screening (21). During colonoscopy, a flexible scope with a camera

is inserted through the anus to inspect the rectum and the colon, with the possibility to obtain biopsies of suspicious lesions and to remove detected precancerous conditions, such as polyps, adenomas and advanced adenomas (AA). Removal of AA substantially contributes to CRC prevention through reducing incidence and consequently to reducing mortality (34, 128). However, the diagnostic efficiency of colonoscopy greatly depends on the completeness of bowel preparation which involves diet restrictions days before colonoscopy and induced diarrhea to clear bowels. As these steps can be inconvenient for participants, the instructions might not be carried out completely, which may reduce the overall effectiveness of screening colonoscopy. The side effects of colonoscopy include pain and inconvenience, bleeding, infections and bowel perforation (106). These factors, together with the fear of colonoscopy itself, embarrassment and fear of positive tests result were reported to create a barrier for participation in screening programs (22).

The direct use of colonoscopy is the preferred screening tool in some countries (e.g., Germany, the U.S. and Austria), but the majority of CRC screening programs rely on a two-step approach, where only a preselected population is considered for screening by colonoscopy (163). Preselection is achieved by prior testing, most commonly using fecal immunochemical tests (FITs) for human blood (hemoglobin) in stool samples. FITs can detect approximately 79% of CRC and correctly identify 94% of cancer-free individuals (102). Recently, two new preselection tests were approved for average-risk population screening – the Cologuard test which combines FIT with the molecular assays for aberrantly methylated specific DNA markers (NDRG4 and BMP3 genes), β-actin and analysis of KRAS mutation in stool (82), and the Epi proColon test which measures elevated methylated Septin9 DNA levels in plasma (43). Major limitations of these tests are their high costs. Data on their performance, specifically sensitivity, specificity and participation rate derived from studies conducted in screening setting are still very limited. The research on non-invasive highly acceptable and economical testing to preselect populations for colonoscopy screening is still ongoing and might be highly relevant for improving CRC screening programs in the future.

1.1.2 Screening for GC and preselection of participants in GC screening

Due to the lack of specific symptoms, GC is usually diagnosed in advanced stage with poor prognosis (5-year survival rates of only 10-30%) (85). Upper gastrointestinal endoscopy (UGE) is the gold standard for GC detection, which uses a thin scope with a camera to

examine changes in esophagus, stomach and duodenum, and allows obtaining multiple biopsies from various places in the stomach for further assessment. The efficiency of the method strongly depends on the experience of the endoscopist. UGE is an invasive method, with bleeding and perforation being the most frequent complications (106).

Currently, population-based GC screening programs have been implemented only in Korea (168) and Japan (78), which are among the countries with the highest GC incidence worldwide (57). In both countries men and women aged 40 years and more are entitled to be screened biennially with UGE and photofluorography, respectively. Screening for GC is generally not recommended elsewhere (41, 104, 206).

In theory, the preselection of high-risk subgroups for further invasive testing could enhance GC screening effectiveness, but currently available methods remain imperfect. Testing for the best-established GC risk factor, *Helicobacter Pylori (H. Pylori)* infection, is of limited use for risk stratification due to high prevalence of the infection (59, 90), with the majority of affected people still having low GC risk. Serum pepsinogens are often discussed markers for GC risk stratification (1, 60). Unfortunately, they are limited for the identification of atrophic gastritis rather than cancer, and only exhibit a moderate diagnostic value for GC identification (80). Both *H. Pylori* status and serum pepsinogen levels are associated with increased risk of non-cardia GC (cancer located in a body of the stomach), which comprises of 73% of all GC worldwide (45), but have limited predictive value for cardia GC (cancers located in the upper part of the stomach). Cardia GC was shown to be the dominant cancer subtype in some countries (e.g., Australia, the U.S. and the United Kingdom) (45), and this subtype might require different screening strategies and tests.

1.1.3 Gastrointestinal cancer screening – the need for novel screening tools

For both CRC and GC, current gold standard screening technologies (colonoscopy and UGE) are invasive interventions creating a barrier to participation. A screening approach with preselection of the target population by non-invasive or minimally invasive methods is associated with better screening uptake (127), yet, there is a need for broadly acceptable tools for preselection of screening populations in gastrointestinal cancer screening.

Among novel non-invasive cancer detection tests, exhaled breath analysis (breath testing) may hold promising potential for cancer screening. Breath testing in general is harmless to

participants, fast and easy to perform, and a likely cheap method, making it potentially suitable for mass screening, either as a replacement of current screening tests or in a combination with existing cancer screening technologies.

1.2. Background and introduction of breath analysis

1.2.1 Origin of breath testing

Infections, inflammations and chronic diseases, as well as external factors such as diet, medication and smoking induce changes in bodily metabolism, which may alter or manifest in bodily secretions (10, 70, 113). Vapors from urine, skin and blood, as well as composition of exhaled breath have been investigated for candidate markers to detect various diseases. Among all these approaches, analysis of exhaled breath is the least invasive method and therefore very attractive for both researchers and patients.

The air we breathe consists of nitrogen, oxygen and carbon dioxide, together with tiny amounts of water vapor and other molecules, including volatile organic compounds (VOCs). In exhaled breath analysis, VOCs are targeted components, as metabolic alterations in the body are thought to be reflected in the absence or presence of specific VOCs, as well as changes in regular VOCs concentrations. The “breath-print”, the total composition of all the elements in exhaled breath is unique for a given person at the time.

In 1971, Nobel Prize winner Linus Pauling reported that urine vapor and exhaled breath of healthy individuals contain more than 200 different VOCs (131). This led to the outbreak in research for exhaled breath analysis to be used as harmless tool for disease detection. By now, several breath tests have been developed in medicine and are broadly administered, including C13-urea breath test for *H. Pylori* detection (86), hydrogen breath test for small bowel bacterial overgrowth (87) and nitric oxide measurements in breath to monitor treatment in asthma (88).

1.2.2 Overview of breath analysis methods

The general idea of exhaled breath analysis is to detect “the smell of the disease”. Since the days of Hippocrates physicians were using their nose to detect various illnesses, as specific odors might be linked with certain diseases (Table 1). However, human nose is comparatively

insensitive to be used as a diagnostic tool. Other, more sensitive methods will be briefly described in the following subsections.

Table 1. Odours in breath associated with specific diseases (from Wilson et al. (200))

Condition	Description	Odour
Empyema	Bacterial-caused accumulation of pus in pleural cavity	Foul, putrid
Esophageal diverticulum	Condition, where oesopahgeal wall forms a sac or pouch	Feculent, foul
Fetor hepaticus	Condition when liver secretions passes to lungs	Newly-mown clover, sweet
Hypermethioninemia	Disorder of excess amino acid methinine in blood	Sweet, fruity, fishy, boiled cabbage, rancid butter
Intestinal obstruction	A blokage in bowels	Feculent, foul
Isovaleric acidemia	Isovaleric acid deficiency	Cheesy
Ketoacidosis	High levels of acids in blood	Sweet, fruity, acetone-like
Liver failure	Medical condition of impaired liver	Musty fish, raw liver, feculent, coal tar-like, rotting meat-like
Lung abscess	Necrosis of lung tisue	Foul, putrid
Pneumonia	Bacterial or viral lung infection	Putrid
Renal failure	Medical condition of impaired kindneys	Stale urine
Trench mouth	Infection of the gums	“Bad breath”
Uremia	Condition described by urea in blood	Fishy, ammonia, urine-like

1.2.2.1 Canine scent detection

Canine olfaction is 1,000-10,000,000 times more sensitive compared to the olfaction of humans, and dogs are broadly trained and used by the police and military for detection of hazardous substances, such as explosives for example. In medicine, canine ability to smell cancer was first reported in 1989, when a case report was published describing a patient who sought medical care after unusual behavior of the family dog, which kept sniffing one mole of the owner while completely ignoring the rest of them. That specific mole was later diagnosed as melanoma of the skin (199). More recently promising results with trained sniffer dogs were reported for lung, bladder and CRC detection (35).

However, the practical implementation of canine detection in any screening program is obviously challenging due to characteristics of the animal itself (rather short lifespan, long training process, variability of the results between measurements, etc.) and undoubtedly potential logistic issues. Nevertheless, canine ability "to smell" malignancies demonstrates that cancer indeed can be described by a specific odor, and given the right tools, one may indeed detect cancer through its "smell".

1.2.2.2 Analytical methods using compound identification

Powerful chemical-analytical methods exist to separate, identify and quantify the components comprising a sample. Gas chromatography–mass spectrometry (GC-MS) is the gold standard of the analytical methods. However, while GC-MS allows the identification of specific individual components in the sample (which can lead to the discovery of cancer-specific biomarkers in exhaled breath), the method is complex to apply as it requires storing of the samples and multiple sample preparation steps before GC-MS analysis, as well as well-trained personnel for sample handling, and rigorous data preprocessing steps of the chromatographic data (the output) before statistical analysis. For these reasons, GC-MS is an expensive and time consuming method that is currently not suitable for cancer mass screening.

1.2.2.3 Pattern recognition-based methods - "The electronic nose"

Sensor array technologies or electronic noses (e-noses) mimic canine (and human) olfaction systems (37) and aim to identify the differences between the compared groups when looking into total compositions of the samples. E-noses can contain several chemical sensors (which can be similar or of different technology) and are usually optimized for a particular application, such as for detection of excess nitric oxide in asthma patients (88).

Chemical sensors react to agents in the sample, and the reactions produce multiple signals which correspond to the whole composition of the sample. To distinguish for example cancer patients from cancer-free individuals, first, a training procedure is applied using samples for which disease status is known. The successful classification on the new unknown samples greatly depends on the training procedure, where the size of the training set and how good these samples represent tested populations are the crucial factors.

In practice, e-noses should meet several requirements, e.g., to provide consistent output that is specific to a given exposure, meaning that the signals from the sensors are always the same for the same input. Furthermore, for specific disease detection, such as cancer, chemical sensors should be sensitive for very low VOC concentrations (20, 111). The reduction in the complexity of the e-nose data can be achieved by using sensors selective for specific target compounds, and high sensitivities were demonstrated for ethanol and several other polar VOCs (116, 203). However, data on cancer-specific volatile compounds in exhaled breath are sparse, as the understanding of cancer-specific biomarkers is still limited to the information obtained by standard analytical methods, such as GC-MS.

Aims of the dissertation

The aim of this dissertation is to provide further insight into exhaled breath analysis for cancer detection and to explore the potential of breath testing for gastrointestinal cancer screening. In particular, the aims are:

1. To review and summarize the literature on the performance of exhaled breath analysis for cancer detection. Specific attention is focused on:
    - Validity of the reported results
    - Differences in breath sampling and analysis protocols
    - Remaining challenges of breath analysis
2. To explore potentially important factors that may influence breath tests results using data obtained by gas chromatography-mass spectrometry, including:
    - Socio-demographic factors
    - Lifestyle factors
    - Medical conditions
    - Dietary patterns
3. To explore the potential of breath testing to be used as a screening tool in CRC and GC screening programs in various populations, by:
    - Demonstrating how the disease prevalence effects performance of screening tests
    - Demonstrating how breath testing can help to improve current screening programs by preselecting screening populations for further invasive testing

2. Systematic review on cancer detection through exhaled breath

To my knowledge, systematic data on exhaled breath analysis for detection of cancer of any type, as well as the performance of exhaled breath analysis for cancer detection by any type of breath analysis methods has not been published before. I have comprehensively summarized available data on exhaled breath analysis for the detection of any cancer type using all breath analysis methods potentially suitable for practical application in cancer screening (therefore sniffer dogs were not considered), and my results were published in (94). Differently from previous reviews, my review includes information on methodological issues, such as breath sampling, storing and analyses protocols and statistical methods used for data analysis. I also systematically summarize individual VOCs that are often reported by studies on cancer detection through exhaled breath, either when comparing concentrations of individual VOCs between cancer patients and controls, or by including them in classification models for cancer detection.

2.1. Methods

2.1.1 Study selection for systematic review

To identify studies which used exhaled breath analysis to detect cancer, a systematic literature search was performed in Pubmed and Web of Science databases on April 30, 2015, using the following combination of keywords: **(cancer OR carcinoma OR adenocarcinoma OR tumor OR malignancy OR malignant disease) AND ((volatile AND (compound OR compounds OR marker OR markers OR biomarker OR biomarkers)) OR VOC OR VOCs OR breathprint OR breath-print OR breath print) AND (breath OR exhaled OR air)**. Only full-text original studies in English language which reported any results on diagnostic performance (i.e., sensitivity, specificity, accuracy or area under the receiver operating characteristic curve (AUC)) for discrimination of cancer cases from controls, or studies which reported significance of individual VOCs between cancer patients and controls were included in this systematic review. Studies exclusively done *in vitro* were not considered for this review. In addition, reference lists of the studies included in this review were checked for other relevant published studies which were potentially missed.

2.1.2 Data extraction for systematic review

2.1.2.1 Study design and characteristics of populations

For each study included in this systematic review, study design (e.g. case-control, including patients with known disease status, or including symptomatic population with breath sampling prior final diagnosis), the type of control group (healthy, benign conditions or mixture of both) were extracted, along with the analyzed cancer type and country in which study participants were recruited. Extracted information on the characteristics of the study populations included numbers of cases and controls, mean or median age, sex and smoking prevalence. Missing information was calculated where possible; for example, when statistics of two groups comparison (cases and controls) was provided but study population description included information on separate smaller subgroups, then numbers were added up or weighted averages (e.g. of age) were calculated.

2.1.2.2 Patient recruitment and breath sampling protocols

Time of breath sampling in cancer patients (before or after treatment) was recorded, as well as usage of antibiotics in both cancer patients and controls. Extracted exclusion criteria for control group consisted of presence or absence of alarm symptoms and presence of chronic underlying diseases, such as diabetes.

The information on breath sample collection and handling protocols used in each of the study, including breath collection time (e.g., morning after fasting) and applied restrictions before the sampling (e.g., no smoking for 2 hours allowed), and the ways to prevent the potential contamination of the samples by using bacterial filters when inhaling, collecting the samples in well-ventilated rooms and other methods, such as additional analyzes of ambient air (surrounding air in the room), was extracted. The time between sample collection and analysis was recorded as well.

2.1.3 Breath analysis methods, collected breath, storage techniques and statistical approaches

Given a broad range of possible breath analysis methods (see chapter 1.2.2), methods used to analyze breath samples were recorded, as well as information on which breath part was used for the analysis (normal breath, deep breath or only last part of the breath corresponding to the air in lungs, i.e., alveolar breath) and where collected samples were stored (types of bags,

vials, cans or specific sampling devices used) were recorded. Statistical approaches for data analysis were also extracted.

2.1.4 Performance of breath analysis for cancer detection

Indicators for diagnostic performance were extracted both for individual VOCs as well as for multi-VOCs classifiers where provided: sensitivity and specificity, accuracy and AUC. Missing accuracy of the test was calculated as the sum of correctly classified cases and controls divided by total number of people in a classification model. Missing sensitivity and specificity was calculated when absolute numbers of true positives (number of cancer cases classified as "cancer" by breath testing) and true negatives (number of cancer-free individuals classified as "healthy" by breath testing) were provided. Sensitivity was calculated as a ratio of true positives to total number of cancer patients in a study, and specificity as a ratio of true negatives to total number of cancer-free individuals.

Particular attention was paid to assure whether reported results were validated to avoid overoptimism. Overoptimistic estimates of diagnostic performance of the test (e.g., sensitivity and specificity) can appear when the classification rule to distinguish patients from controls is built and tested using same study populations. Such classification rules tend to be overfiting – including too many parameters in a model or adding irrelevant predictors that will worsen the performance of such a model when tested on independent datasets (75). Therefore, the most reliable information was extracted, such as bootstrapped or cross-validated values, wherever such results were provided. For studies which used random sample split to create a model and validate it separately, only values corresponding to validation set were considered.

2.1.5 Cancer-related compounds

To determine potential cancer-related or cancer-specific VOCs, names of individual volatile compounds which showed a significantly different concentration in exhaled breath between cancer cases and controls, or which were used by authors to build a classification model to distinguish cancer cases from controls, were extracted from all of the studies. The International Union of Pure and Applied Chemistry (IUPAC) name (92, 129) was checked for all extracted compounds to detect synonyms and to ensure comparable results.

2.2. Results

In the following subchapters, studies selected for this systematic review are described and the results on breath testing performance for cancer detection is provided.

2.2.1 Systematic literature search results

In total 1,277 papers were identified of which 262 were duplicates, 24 non-English papers and one book chapter (Figure 1). The remaining titles and abstracts were checked and studies not relevant to the topic were excluded. For 17 studies, no full paper could be accessed as they mostly were published as conference abstracts. Additionally, 15 papers were excluded after full-text revision as some of the required information was missing (see Appendix 1). Among these, a study on simulated breath samples (135) and a study reporting differences in VOCs concentration between cancer patients and controls, but not providing names of these VOCs (162) were excluded.

Besides of distinguishing cancer cases from controls by exhaled breath analysis, several included studies reported additional information, such as differences in exhaled breath composition of cancer patients before and after tumour resection (36, 51, 146), also VOCs released by cancer cells or tissues (58, 192), and performance of other diagnostic methods, including canine detection (38) and DNA hypermethylation in sputum (81). In these studies only data related to cancer detection through exhaled breath analysis were considered in this systematic review.

In total, 73 studies that met inclusion criteria are listed in Appendix 2. 46 studies were conducted on lung cancer, followed by studies on breast cancer (N=11) and GC (N=5). Malignancies investigated exclusively by one study were cancer of prostate, thyroid, liver, ovaries and gynecological and heamatological cancers. Studies in this systematic review reported the performance of breath testing from various populations, with studies from the U.S. (N=17) and China (N=15) being the most frequent.

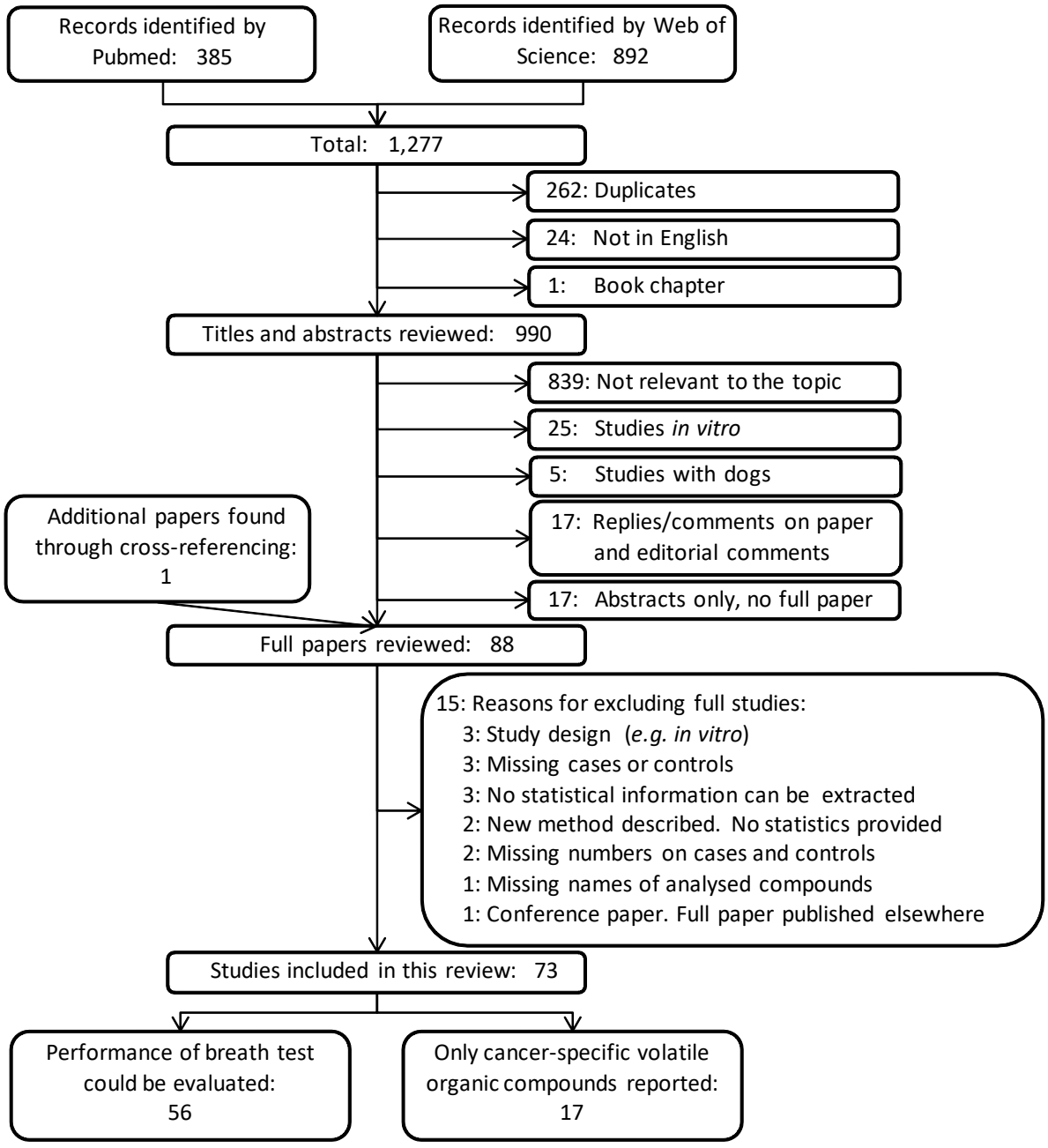

Figure 1. Flow diagram for literature search process to identify studies on exhaled breath analysis for cancer detection.

2.2.2 Study design and characteristics of populations of the included studies

An overview on the participants in the studies, including number of men and women, mean or meaning age and smoking prevalences of cases and controls in each of the studies is provided in Appendix 3. Numbers of participants in analyzed studies varied from 14 (6 cases and 8 controls (201)) to 477 (99 cases and 378 controls (7)). While it is well-known that smoking status can confound the results from the classification model on lung cancer detection, information on smoking status was not provided by 8 out of 46 studies on lung cancer detection.

The majority of studies used healthy controls; however, a mixture of healthy controls and individuals with some benign medical conditions were used as a reference group in 9 studies, and 8 studies exclusively used controls with benign medical conditions. Furthermore, several repeated studies using different statistical approaches to classify cancer cases and controls of the same study populations were conducted, including four studies on lung cancer (137, 138) and (17, 108), and two on breast cancer (142, 143) and malignant mesothelioma (50, 54).

The vast majority of studies were conducted in a case-control approach, where well-defined groups of clinically diagnosed patients and cancer-free controls were compared, while eight studies collected breath samples before the final diagnosis of participants was determined (such a study design resembles a design in the screening setting).

2.2.3 Patients recruitment and breath sampling protocols

Further information on study design and data collection details is presented in Appendix 4. Around 25% of the studies did not perform lung washout or did not analyze ambient air which might lead to exogenous (inhaled) compounds to be included into classification models. Breath sampling was performed in the morning after fasting in 25 studies in order to guarantee that compounds arising immediately after food consumption do not influence the classification model.

Time between breath collection and analysis was very short (analysis done immediately or within few hours) in most studies but extended up to six months in one study (7). Although most of the studies included newly diagnosed untreated cancer patients, few studies recruited patients under different treatment regimens (in these, treatment might have had an influence on exhaled volatiles).

2.2.4 Breath analysis methods, collected breath, storage containers and statistical approaches

The breath sample analysis methods for each of the study, collected breath parts and their storage containers, as well as used statistical analysis methods are provided in Appendix 5. GC-MS analysis was used in 42 studies, e-noses were used by 24 studies, while in some studies other analysis methods were applied. The most commonly used e-noses were gold nanoparticles sensors-based e-nose from the TECHNION group (109) (N=8) and the commercially available Cyranose 320 (165) (N=6).

Differences in collected and analyzed breath parts as well as sample storage containers were also observed. While the majority of the studies collected alveolar breath (last part of air from the lungs), 7 studies focused on the analysis of tidal breath (normal breath) and 12 studies collected maximum amount of exhaled breath (vital capacity). Several studies applied sample analysis techniques allowing the data analysis to be performed at the moment samples are collected, thus no sample storing was employed (72, 74, 114, 115, 154, 198). In case sample storing was needed, various storage containers were reported, including different bags (Tedlar, Mylar, Nalophan and Rapak bags), glass vials (Tenax sorption tubes, sterile vials), cans or techniques specifically design for breath sampling (119, 136). Commercially available Tedlar bags were the most frequently used storage containers (N=33), followed by Tenax sorption tubes (N=9).

Studies also differed regarding statistical methods applied to create classification rules for cancer detection, including custom made rules (30, 61) and computational models, such as neural networks (4, 40, 158, 191).

2.2.5 Performance of breath tests for cancer detection

48 out of 73 studies included in this systematic review reported sensitivity and specificity, diagnostic accuracy or AUC for distinguishing cancer cases from controls based on exhaled breath analysis. Cancer type-specific breath test performance is described in the following subchapters.

2.2.5.1 Lung cancer

Performance of breath tests for lung cancer detection is presented in Table 2. Overall, 31 studies reported diagnostic performance for lung cancer detection out of which 5 studies

validated their results in independent datasets and 15 employed internal validation procedures. Overoptimistic results might have been reported from the studies without validation (N=11), which sometimes reported a perfect discrimination (100% accuracy) (66, 198). Very good diagnostic performance was also reported in some studies with validation showing AUCs of 0.97 (158) and 0.986 (132), or overall accuracy of 96% (71, 195). An independent validation study for (61) using the same classification rule of classifying participants "positive" when 2 or more out of 4 selected VOCs have higher concentrations than the set cut-offs, had demonstrated accuracy of 85% (30). Two studies that used the same study population reported sensitivities of 51 and 52% (with 100% specificity) when including 4 (17) and 8 (108) VOCs in their classification rules, respectively - in both classification models three VOCs (1-propanol, 2-butanone, benzaldehyde) were the same. However, these VOCs were not used together when building a classification model by any other study on lung or other cancer types.

Table 2. Breath test performance for lung cancer detection

First author, year reference	Cases / Controls	Sens	Spec	AUC	Acc	Validation technique or model construction
Diagnostic performance with validation (N=20)						
Boumsamra, 2014 (30)	107 / 40	87.9	77.5	-	85.0	Validation of the same model as in (61)
Broza, 2013 (36)	12 / 5	100	80.0		94.1	LOOCV
Chen, 2005 (40)	5 / 5	80.0	80.0	-	80.0	Validation set
D'Amico, 2010 (49)	28 / 36 28 / 28[c]	85.0 92.8	100 78.6	- -	93.8 85.7	LOOCV
Dragonieri, 2009 (53)	10 / 10 10 / 10[b]	- -	- -	- -	90.0 85.0	Cross-validation
Hakim, 2011 (71)	20 / 26	100	92.3	-	95.7	Average of all sample splits
Hubers, 2014 (81)	18 / 8	94.4	12.5	-	69.2	Validation set
Machado, 2005 (111)	14 / 62	71.4	91.9	-	88.2	Validation set
Mazzone, 2007 (114)	49 / 94	73.3	72.4	-	-	RRS - 70:30%
McWilliams, 2015 (118)	25 / 166	-	-	0.803	-	Average of 10 RSS - 2:1
Peled, 2012 (132)	50 / 19	86.0	96.0	0.986	88.0	LOOCV
Phillips, 1999 (145)	60 / 48	71.7	66.7	-	69.4	LOOCV
Phillips, 2003 (141)	67 / 41 - / 91[a]	85.1 -	80.5 37.4	- -	83.3 -	LOOCV Validation set
Phillips, 2007 (137)	193 / 211	84.6	80.0	0.88	-	RSS - 2:1
Poli, 2010 (147)	40 / 38	90.0	92.1	-	91.0	LOOCV
Santonico, 2012 (160)	20 / 10	85.0	85.0	-	85.0	LOOCV
Rudnicka, 2014 (158)	108 / 121	74.0	73.0	0.97	-	RSS - 50:25:25%
Wang D, 2012 (191)	47 / 42	93.6	83.4	-	88.8	LOOCV
Wang Y, 2012 (192)	85 / 158	96.5	97.5	-	97.1	LOOCV
Wehinger, 2007 (195)	17 / 170	54.0	99.0	-	96.0	Average of 1,000 RSS - 60:40%
Diagnostic performance without validation (N=11)						
Bajtarevic, 2009 (17)	65 / 31 65 / 31 65 / 31	52.0 71.0 80.0	100 100 100	- - -	- - -	Model on selected 4 VOCs Model on selected 15 VOCs Model on selected 21 VOCs
Fu, 2014 (61)	97 / 32	92.8	81.3	-	89.9	Model on selected 4 VOCs

(Table 2 continues on the next page)

(Table 2 Continued)

First author, year reference	Cases / Controls	Sens	Spec	AUC	Acc	Validation technique or model construction
Gordon, 1985 (66)	12 / 9	-	-	-	93.0	Model on selected 3 VOCs
	12 / 9	100	100	-	100	Model on selected 22 VOCs
Handa, 2014 (72)	50 / 39	76.0	100	-	-	Model on selected 10 VOCs
Ligor, 2009 (108)	65 / 31	51.0	100	-	-	Model on selected 8 VOCs
Mazzone, 2012 (115)	83[d] / 137	-	-	0.701	-	Model on selected sensor parameters
	9[e] / 137	-	-	0.8	-	
Phillips, 2008 (138)	193 / 211	-	-	0.87	-	RSS, yet preselection of VOCs performed on total study population
Poli, 2005 (146)	36 / 110	72.2	93.6	-	88.4	Model on selected VOCs
Steeghs, 2007 (179)	11 / 57	-	-	0.81	-	Model on selected VOCs
Westhoff, 2009 (198)	32 / 54	100	100	-	100	LOOCV, yet preselection of VOCs performed on total study population
Yu, 2011 (205)	9 / 9	100	88.9	-	94.4	Model on selected peaks

Acc, accuracy; AUC, area under the receiver operating characteristic curve; LOOCV, leave-one-out cross-validation; RSS, random sample split - training set size: validation set size of total study population. Performance of breath test corresponds to validation set and numbers of cases and controls correspond to total study population. Sens, sensitivity; Spec, specificity; VOC, volatile organic compound.

[a] Abnormal X-rays, no cancer;

[b] Chronic obstructive pulmonary disease;

[c] Lung diseases;

[d] Non-small cell lung cancer;

[e] Small cell lung cancer.

2.2.5.2 Breast cancer

Performance of breath tests for breast cancer detection is presented in Table 3. Overall, six studies reported diagnostic performance for breast cancer detection and all of them employed internal validation procedures to control for overoptimism. The best discriminatory performance was achieved by Phillips *et al.* in 2006 (143), who reported AUC of 0.9. However, the performance of the same model on a validation set using women who were cancer-free but had abnormal mammography findings was lower: only 32% of these women were classified as healthy. Other studies also showed better performance of the classification models when comparing breast cancer patients with healthy women rather than with women with abnormal mammography findings (140, 142).

Table 3. Breath test performance for breast cancer detection.

First author, year reference	Cases / Controls	Sens	Spec	AUC	Acc	Validation technique
Li, 2014 (107)	22 / 24	68.2	91.7	-	80.4	LOOCV
Patterson, 2011 (130)	20 / 20	72.0	64.0	-	77.0	Average of 10,000 RSS - 60:40%
Phillips, 2003 (142)	51 / 42 51 / 50[a]	88.2 60.8	73.8 82.0	- -	81.7 71.3	LOOCV
Phillips, 2006 (143)	51 / 42 - / 50[a]	93.8 -	84.6 32.0	0.9 -	- -	RSS - 70:30% Validation set
Phillips, 2010 (144)	54 / 204	75.3	84.8	0.83	-	Average of 10 RSS -2:1
Phillips, 2014 (140)	35 / 93 35 / 79[a]	-	-	0.73 0.67	-	LOOCV

Acc, accuracy; AUC, area under the receiver operating characteristic curve; LOOCV, Leave-one-out cross-validation; RSS, random sample split - training set size: validation set size of total study population. Performance of breath test corresponds to validation set and numbers of cases and controls correspond to total study population. Sens, sensitivity; Spec, specificity.

[a] Abnormal mammography findings

2.2.5.3 Gastrointestinal cancers

Performance of breath tests for CRC and GC detection is presented in Table 4. Overall, five studies reported diagnostic performance for CRC and GC detection. Accuracy of 76% was reported for CRC detection (N=1) and accuracies between 85 and 92% were reported for GC detection (N=4). The study on CRC detection reported results validated on an independent validation set, whereas all studies on GC detection applied internal validation methods.

Table 4. Breath test performance for colorectal and gastric cancer detection.

First author, year reference	Cancer site	Cases / Controls	Sens	Spec	AUC	Acc	Validation technique
Altomare, 2013 (4)	CRC	15 / 10	80.0	70.0	-	76.0	Validation set
Xu, 2013 (202)	GC	37 / 93	89.0	90.0	-	90.0	RSS - 75:25%
Amal, 2016 (7)	GC	99 / 325 99 / 53[a]	73.3 86.7	97.9 86.7	- -	92.0 86.7	RSS - 70:30%
Kumar, 2015 (96)	GC	81 / 121	86.7	81.2	0.87	-	Average of 10 RSS - 2:1
Shehada, 2015 (167)	GC	30 / 77	71.0	89.0	-	85.0	RSS - 75:25%

Acc, accuracy; AUC, area under the receiver operating characteristic curve; CRC, colorectal cancer; GC, gastric cancer; RSS, random sample split - training set size: validation set size of total study population. Performance of breath test corresponds to validation set and numbers of cases and controls correspond to total study population. Sens, sensitivity; Spec, specificity.

[a] Gastric ulcer

2.2.5.4 Other malignancies

Performance of breath tests for detection of other malignancies is presented in Table 5. Overall, seven studies on four different cancer types, including cancer of head and neck, liver and ovaries, as well as malignant mesothelioma, were identified. Accuracy of 88% was demonstrated for liver cancer detection, however, the classification model included VOCs preselected in the total study population which might have led to overoptimistic estimates. Good diagnostic performance was also reported for head and neck cancer and malignant mesothelioma detection with AUC or accuracy higher than 90% (39, 71). A study on ovarian cancer detection demonstrated accuracies of 58, 72 and 89% when discriminating cancer patients from benign conditions, benign conditions mixed with healthy individuals and exclusively healthy individuals, respectively (8).

The control for optimistic performance was assured in all the studies by internal validation techniques. Additionally to reported sensitivity and specificity of 90 and 91%, respectively, for malignant mesothelioma detection, the authors validated their model on an independent dataset of 18 individuals with benign lung diseases, and demonstrated 83% specificity (39).

Table 5. Breath test performance for other cancer detection.

First author, year Reference	Cancer site	Cases / Controls	Sens	Spec	AUC	Acc	Validation technique
Gruber, 2014 (67)	HNC	22 / 19 22 / 21[a]	77.0 77.0	90.0 90.0	- -	83.0 84.0	LOOCV
Hakim, 2011 (71)	HNC	16 / 26	100.0	92.3	-	95.2	Average of all sample splits
Leunis, 2014 (105)	HNC	36 / 23	-	-	0.85	-	Bootstrapped value
Chapman, 2012 (39)	MM	20 / 42 - / 18[c]	90.0 -	91.0 83.3	- -	90.5 -	RSS[b] Validation set
Dragonieri, 2012 (54)	MM	13 / 13 13 / 13[d]	- -	- -	0.893 0.917	84.6 80.8	LOOCV
Qin, 2010 (151)	LVC	30 / 36 - / 27[e]	83.3 -	91.7 66.7	- -	87.9 -	LOOCV, yet preselection of VOCs performed on total study population
Amal, 2015 (8)	OC	48 / 48 48 / 86[f] 48 / 134[g]	78.6 57.1 71.4	100.0 59.0 71.8	- - -	89.3 58.0 71.7	RSS - 70:30%

Acc, accuracy; AUC, area under the receiver operating characteristic curve; HNC, head and neck cancer; LOOCV, leave-one-out cross-validation; LVC, liver cancer, MM, malignant mesothelioma, RSS, random sample split - training set size: validation set size of total study population. Performance of breath test corresponds to validation set and numbers of cases and controls correspond to total study population; Sens, sensitivity; Spec, specificity, OC, ovarian cancer.

[a] Benign head and neck conditions

[b] Training set included 10 cases and 10 controls and validation set consisted of 10 cases and 32 controls;

[c] Benign lung diseases;

[d] Exposed to asbestos

[e] Hepatocirrhosis

[f] Benign ovarian conditions

[g] Healthy individuals + benign ovarian conditions

2.2.6 Cancer-related individual volatile compounds

To assess whether individual VOCs are specific for particular cancer, the performance of single-VOC classification models was recorded (Table 6). Sensitivity and specificity of >80% both were demonstrated for butan-1-ol and hexadecanal by studies on lung cancer (175, 192), as well as for 3-hydroxybutan-2-one by studies on lung (175) and liver cancer (209). High levels of specificity were reported for discriminating healthy controls from individuals with lung or breast cancer for several fatty aldehydes, e.g., hexanal, octanal and nonanal (62, 107). Reported concentrations of VOCs were generally higher in cancer patients compared to healthy controls, although one study demonstrated lower abundance of 4 VOCs in breast cancer patients than in cancer-free women (112). Only one study validated the performance of selected compounds in independent data samples and demonstrated superb performance of hexadecanal for lung cancer detection (AUC=1.00) (209).

VOCs which were used to build a classification model or whose concentrations were significantly different between cancer cases and controls in at least three independent studies are presented in Appendix 6. Ethenylbenzene (styrene), heptanal and nonanal were the most commonly described compounds (each in 9 independent studies). Interestingly, these studies were performed on different cancer types, including cancer of lung, breast, CRC and GC. By contrast, 1-propanol was described just by the studies on lung cancer, where 3 studies showed significantly different concentrations in exhaled breath and 4 others included this compound when building a classification model for distinguishing lung cancer patients from controls.

Table 6. Diagnostic performance of individual VOCs for cancer detection.

Author, Year, reference	Cancer site	Volatile compound	Cut-off	Sens	Spec	AUC	Level[a]
Fuchs, 2010 (62)	LC	Pantanal	0.275 nmol/L	75.0	95.8	-	up
		Hexanal	1.208 nmol/L	8.3	91.7	-	up
		Octanal	1.068 nmol/L	58.3	91.7	-	up
		Nonanal	8.433 nmol/L	33.3	95.8	-	up
Song, 2010 (175)	LC	Butan-1-ol	3.67 ng/L	95.3	85.4	0.94	up
		3-hydroxybutan-2-one	3.81 ng/L	93.0	92.7	0.96	up
Wang, 2012 (192)	LC	Hexadecanal	-	96.5	89.2	0.949	-
Handa, 2014 (72)	LC	Dodecane	-	70.0	89.7	-	up
Zou, 2014 (209)	LC	5-(2-methylpropyl)nonane	-	-	-	0.845[b]	up
		VOC-1	-	-	-	0.724[b]	up
		2,6,11-trimethyldodecane	-	-	-	0.846[b]	up
		Hexadecanal	-	-	-	1.00[b]	up
		8-hexylpentadecane	-	-	-	0.672[b]	up
Mangler, 2012 (112)	BC	3-methylhexane	-0.55 µg/m³	100.0	40.0	-	down
		Dec-1-ene	-0.125 µg/m³	100.0	40.0	-	down
		VOC-2	-0.05 µg/m³	100.0	60.0	-	down
		Naphthalene	0.05 µg/m³	90.0	70.0	-	down
		Trichloroethene	0.05 µg/m³	80.0	70.0	-	up
Li, 2014 (107)	BC	Hexanal	10.32 ppbv	77.3	79.2	0.79	up
		Heptanal	9.98 ppbv	68.2	91.7	0.823	up
		Octanal	12.9 ppbv	63.6	87.5	0.734	up
		Nonanal	23.14 ppbv	72.7	95.8	0.832	up
Qin, 2010 (151)	LVC	3-hydroxybutan-2-one	2.44 ng/L	83.3	91.7	0.926	up
		Ethenylbenzene	14.92 ng/L	66.7	94.4	0.812	up
		Decane	1.64 ng/L	86.7	58.3	0.798	up

AUC, area under the receiver operating characteristic curve; BC, breast cancer; LC, lung cancer; LVC, liver cancer; Sens, sensitivity; spec, specificity;

VOC-1, 2,6-di-tert-butyl-4-methylphenol.

VOC-2, 4,11,11-trimethyl-8-methylidenebicyclo[7.2.0]undec-4-ene.

[a] Concentration level of the compound in cancer patients compared to healthy controls.

[b] Results from the validation set.

2.3. Current challenges in breath analysis and the way forward

In the studies included in this systematic review, good diagnostic performance of exhaled breath analysis for the detection of various cancers was demonstrated. However, in order to confirm the potential of exhaled breath analysis for cancer screening, the validation of already available results in diverse populations will be necessary, followed by further confirmation of the method's performance in a true screening setting. To my knowledge, currently no data on breath testing performance in a screening setting is available. Several included studies used a mixture of healthy individuals with benign conditions as controls to resemble more realistic characteristics of average-risk populations. However, more pronounced population heterogeneity is expected in average-risk populations undergoing cancer screening. In particular factors other than presence of the cancer that can influence the results of breath tests, are not yet fully understood.

The studies included in this systematic review differed profoundly on aspects such as analyzed breath parts, applied protocols for patients' recruitment and sampling, sample storage time and type of breath analysis as well as statistical methods used to analyze their samples. To what extent these potential differences may affect the results of the studies is not fully understood currently - internationally accepted standardization processes for breath analysis are still lacking, although some recommendations on how to standardize breath sample collection and analysis were already proposed in 2010 (9).

While available data suggest that different sampling, storing and analysis protocols can influence overall performance of breath analysis (e.g., higher VOCs concentrations were demonstrated in alveolar breath compared to normal breath (121, 187)), there remains the need for data on factors potentially affecting the results regarding:

- Breath sampling – does the sampling time (e.g., morning *vs*. any time) have an impact on concentrations of measured VOCs? Are samples collected in several locations comparable (e.g., sampling in hospitals *vs*. other places)? Is the potential contamination in ambient air, such as by alcohols from disinfectors in indoor air in hospitals (25), relevant for exhaled breath analysis?
- Breath sample storing – can VOCs evaporate through storage bag walls and what contamination can be caused by the storage container?

- Breath analysis method – can results from different breath analysis methods be directly compared (e.g., GC-MS vs. e-noses)?
- Recruitment of the patients – do typical covariates, such as age, sex and smoking prevalence also influence the results of exhaled breath analysis? Can other comorbidities be detected when distinguishing cancer patients from controls? Do differences in diet and lifestyle factors result in differences in the composition of exhaled breath?

In the next chapter (chapter 3), I examine the potential influence of covariates on analysis results obtained by GC-MS data of cancer-free individuals. Various socio-demographic, lifestyle factors and dietary patterns among these individuals are investigated by looking into the association between these factors and VOCs in exhaled breath.

3. Associations of VOCs with socio-demographic, lifestyle factors and dietary patterns

The distinct composition of VOCs in exhaled breath (Figure 2) reflect the body's metabolic state at a given time as influenced by infections, diet, smoking, socio-demographic factors (e.g., sex), as well as underlying diseases including cancer if present. In general, the aim of exhaled breath analysis for cancer detection is to identify a "breath print" that is unique for cancer patients compared to cancer-free individuals. To achieve a correct classification of cancer patients and cancer-free individuals, the underlying classification model (constructed before either by pattern recognitions methods (e-noses) or data from analytical methods (array of VOCs) should be specific to the disease of interest while other factors, such as age, gender and smoking status should not confound the results. Knowledge of the potential influence of covariates on breath analysis results is crucial for proper design, analysis and interpretation of breath tests studies and practical application of breath tests.

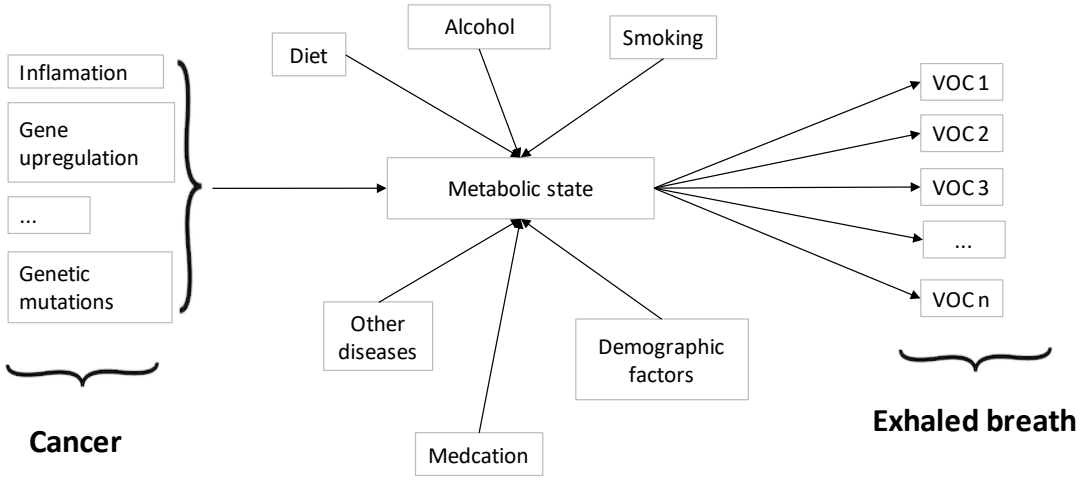

Figure 2. A scheme demonstrating relationship of cancer and covariates with volatile compounds in breath.

Factors potentially influencing breath testing results are not fully investigated. Some evidence exists that breath sampling and storing can influence the results; however, data on other potentially relevant factors, particularity the effect of lifestyle factors and nutrition on exhaled breath analysis results are sparse. I aimed to evaluate the potential co-determinants of GC-MS results using one of the biggest datasets of exhaled breath measurements up-to-date (part of

the GISTAR study). The analysis includes examining the associations of VOCs with socio-demographic, lifestyle, medical factors and dietary patterns among cancer-free individuals. A manuscript regarding the results of this analysis is in preparation (see List of own publications).

3.1. The GISTAR study

GISTAR is a population-based randomized clinical trial designed primarily to evaluate the effect of *H. Pylori* eradication on GC mortality in the asymptomatic, average-risk population in Latvia. Individuals in the trial arm receive triple therapy for *H. Pylori* eradication when *H. Pylori* is present, while participants in the control arm receive no treatment. Additionally, *H. Pylori* positive individuals with increased pepsinogen I/II ratio are offered UGE testing. Differences in GC mortality are expected to be measured in 15 years from the recruitment. At the time of enrollment and before randomization, breath samples were collected from all participants after they provided informed consent.

3.2. Recruitment of the study population

All participants enrolled in this study were 40-64 years old, with good health status: a life expectancy of at least 5 years, no problems with mobility, as well as good expected compliance with the diagnostic work-up or later treatment. Participants with history of any cancer or gastric resection due to benign conditions, presence of alarm symptoms of digestive or other diseases (e.g., bleeding, weight loss, vomiting, chest pain, difficulties swallowing, *etc*.), also patients with pathological findings at examination were excluded, as were patients who underwent *H. Pylori* eradication within 12 months prior to enrollment.

Figure 3. Recruitment centers in Latvia.

Complete medical examination was performed and a questionnaire was filled at the time of breath sample collection, which was from October 1, 2013 until December 31, 2015 in social service buildings in four towns of Latvia (Figure 3).

3.3. Breath sample collection and storage

All breath samples were collected in the morning after participants fasted for 7 hours and restrained from smoking for at least 2 hours. Before sample collection, participants were asked to breath normally through a disposable bacterial filter for 3 minutes. After that, participants took a deep breath and exhaled, where first dead space bag was filled with the air from mouth and airways up to lungs and then the alveolar breath (the last part of the breath) was saved in Mylar bags for later analysis. Collected samples were immediately transferred to ORBO 402 Tenax TA sorption tubes (Sigma Aldrich) for storage in the fridge at +4C temperature until shipment for analyses. Mylar bags were flushed for 10 min with 99.99% Nitrogen before reusing them to ensure the removal of contaminants.

3.4. Breath sample analysis by GC-MS

Breath samples were analyzed by thermal desorption (TD)-GC-MS in TECHNION, Haifa, Israel. The detailed description of GC-MS analysis can be found in (133, 134). Briefly, breath samples were analyzed with a GC-MS-QP2010 instrument (Shimadzu Corporations) coupled with a thermal desorption system. Immediately prior to analysis, the samples were transferred from the ORBO 402 Tenax TA sorption tubes to empty glass TD tubes compatible with the TD system. The TD tube was heat-treated (270°C) and filled TD tubes were injected into the GC-system in splitless mode at 30 cm/s constant linear speed and under 0.70 ml/min column flow. The following oven temperature profile was set to: (a) 10 min at 35°C; (b) 4°C/min ramp up to 150°C; (c) 10°C/min ramp up to 300°C; and (d) 15 min at 300°C.

In general, GC-MS analysis is a combination of two techniques – GC (gas chromatography), where the sample is separated into individual molecules which are then sorted based on their size, and MS (mass spectrometry), where each of these molecules are split into comprising particles using ionization technique. Briefly, due to high temperature at the GC compartment, the molecules included in the breath sample start to move along the column (long isolated, i.e., coated tube) towards the MS chamber; smaller (lighter) VOCs travel faster and reach the MS chamber first. Molecules are ionized when entering the MS compartment and they split into ions that are sorted and counted based on their mass-to-charge ratio (M/Z). The counts of all different M/Z together with the retention time (RT), i.e., the absolute time from the sample

insertion to the GC to the time when the molecule was detected by MS, comprise chromatographic data.

GC-MS is a time consuming method, and analysis time of one individual's breath sample was around 1 h 30 min. GC-MS analysis was run in parallel with sample collection in Latvia, and overall analysis time of all samples used in this analysis was around 1 year and 4 months. During a long analysis period, GC columns degraded, as well as other technical issues appeared (GC-MS system was out of work for nearly 2 months) which resulted in batch effects in the chromatographic data over analysis time. Methods used to correct the batch effects are described in chapter 3.5.2.

3.5. Methods

3.5.1 Questionnaire data

Detailed questionnaire data regarding socio-demographic factors, lifestyle factors, medical conditions and dietary patterns were collected from all participants at the time breath samples were collected. The following factors were extracted from the questionnaire and included in this analysis:

Socio-demographic factors

- Sex (male and female).
- Nationality (Latvian, Russian, other).
- Body mass index (BMI). BMI was calculated as a ratio between weight and squared height, and was grouped into the 4 following categories: underweight (BMI<18.5), normal (18.5≤BMI<25), overweight (25≤BMI<30) and obese (BMI≥30).

Given the relatively narrow age range of the participants (40-64 years), the variable age was not considered in my analysis.

Lifestyle factors

- Smoking status (smoker, ex-smoker, non-smoker). Individuals were classified as smokers if they smoked at least 100 cigarettes during lifetime and the time since last time smoked is less than 6 months, whereas individuals were classified as ex-smokers, if they smoked at least 100 cigarettes during lifetime and the time since last time smoked is more than 6 months. Other participants were classified as non-smokers.

- Alcohol consumption (non-drinker, light, moderate and heavy). Data on frequency of monthly drinking occasions, together with type and amount of alcohol per occasion was available for the following drinks: beer (4-5.5^{o}), fortified beer (5.6-9^{o}), bottled cocktails, cider, wine (8-14^{o}), fortified wine (15-17.5^{o}), spirits up to 30^{o}, spirits up to 40^{o} and spirits up to 70^{o}. To determine drinking status, the maximum amount of pure ethanol per occasion of any drinks listed above was used. A status "non-drinker" was assigned for 0 grams of ethanol per occasion (i.e., participant did not list any of these drinks listed above), "little" for <14 grams of ethanol per occasion, "moderate" for 14-59 grams of ethanol and "heavy" for ≥60 grams of ethanol per occasion.

Medical conditions

For the statistical analysis, only diseases where at least 25 individuals had the condition were included:

- Asthma
- Diabetes type 2
- Gallstone disease
- Gastric or duodenal ulcer
- Thyroid illness. Given higher prevalence of thyroid illness among women, men were not included in this specific comparison. Also, individuals with both thyroid illness and diabetes type 1 or 2 were excluded due to potential interactions between these two illnesses.
- Other allergies. Individuals reporting the presence of allergy (type was not indicated) were included, yet, only if they had not indicated asthma as well.

Medication

- Acid reduction drugs. Individuals were classified positive if they reported the use of proton pump inhibitors, H2 blockers or omeprazole over the last month, and negative otherwise.
- Anti-inflammatory drugs. Individuals were classified positive if they reported the use of aspirin or other NSAID (non-steroidal anti-inflammatory drugs) over the last month, and negative otherwise.

- Antibiotics. Individuals were classified positive if they reported the use of metronidazole, clarithromycin, amoxicillin, penicillin, tetracycline or levofloxacin over the last month, and negative otherwise.

Dietary patterns

Data on frequency per year as well as frequency per month of consumption of separate food items were available. These frequencies were multiplied to calculate yearly consumption and the mean of the multiplied frequencies was used to classify individuals into two categories: eating certain food items not often (when the person eats that product less often than the mean value among all individuals) and eating the same food items often (when the person eats that product more often than the mean value among all individuals). The following food items were analyzed: meet (beef (cooked), pork (cooked), chicken, other meat products (e.g., sausages)); dairy products (kefir (i.e., fermented milk), curds, cheese); greens (vegetables (such as beet root, potatoes and carrots), and separately onions, onion leaves and garlic, as well as local fruits and imported fruits); processed products (pickled, salted, smoked, dried), other products (porridge; dark bread; legumes, fish (cooked); eggs); and drinks (green tea, black tea, instant coffee, ground coffee, sweetened and carbonated drinks).

3.5.2 Preparation of GC-MS data for statistical analysis

In general, the counts of all M/Z at each time point of analysis when plotted over RT will form structures, so-called peaks (Figure 4). Using software specifically designed for GC-MS data, information on each peak can be investigated, including distribution of M/Z for that peak, time points where a peak starts and ends (with RT corresponding to the maximum of M/Z counts over time) and area under the peak (sum of all M/Z (counts) comprising that peak). In my analysis, I considered each peak to be an individual VOC.

Visual inspection of chromatogram is a commonly used method to estimate the overall quality of the data, where missing common VOCs or high background noise can be easily detected. During my analysis, decisions on several data preparation steps were based on visual data inspection as well, including estimation of the quality of baseline reduction (chapter 3.5.2.1) and peak alignment (chapter 3.5.2.2) processes.

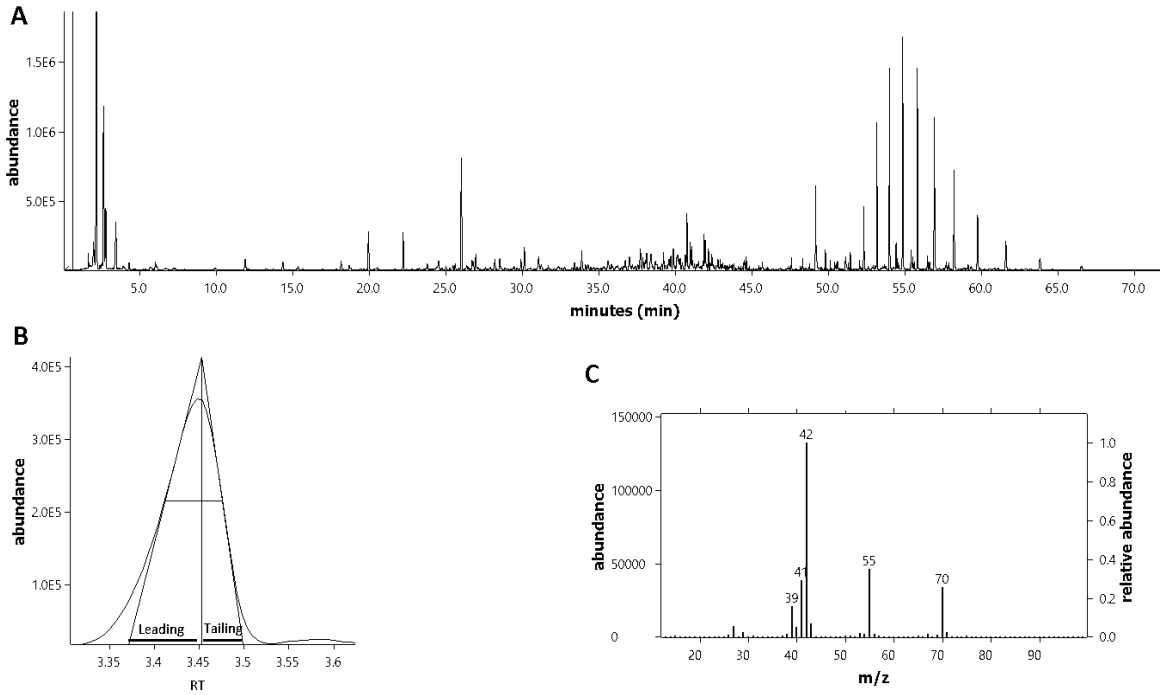

Figure 4. Chromatographic data.

(A) Full chromatogram. (B) A peak at 3.45 min enlarged. (C) Main ions (M/Z) of the peak.

Raw GC-MS data were pre-processed in seven steps according to current recommendations (174) to remove various sources of artifacts before the statistical analysis (Figure 5). Each of these data preparation steps are discussed in the following subsections below.

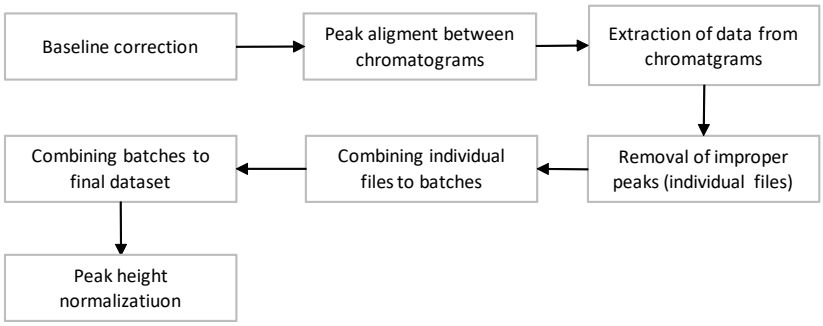

Figure 5. GC-MS data preparation steps.

3.5.2.1 Baseline correction

Together with the ions (i.e., M/Z) from the original sample, additional unrelated M/Z can appear, e.g., due to GC column releasing ions, which may form a baseline of the chromatogram. To remove the baseline from each of chromatograms, B-splines and Asymmetric Least Square method (56) was applied. The results of baseline correction are demonstrated in Figure 6. A strong baseline increase around 36-46 min, as well as overall baseline throughout the whole chromatogram was removed.

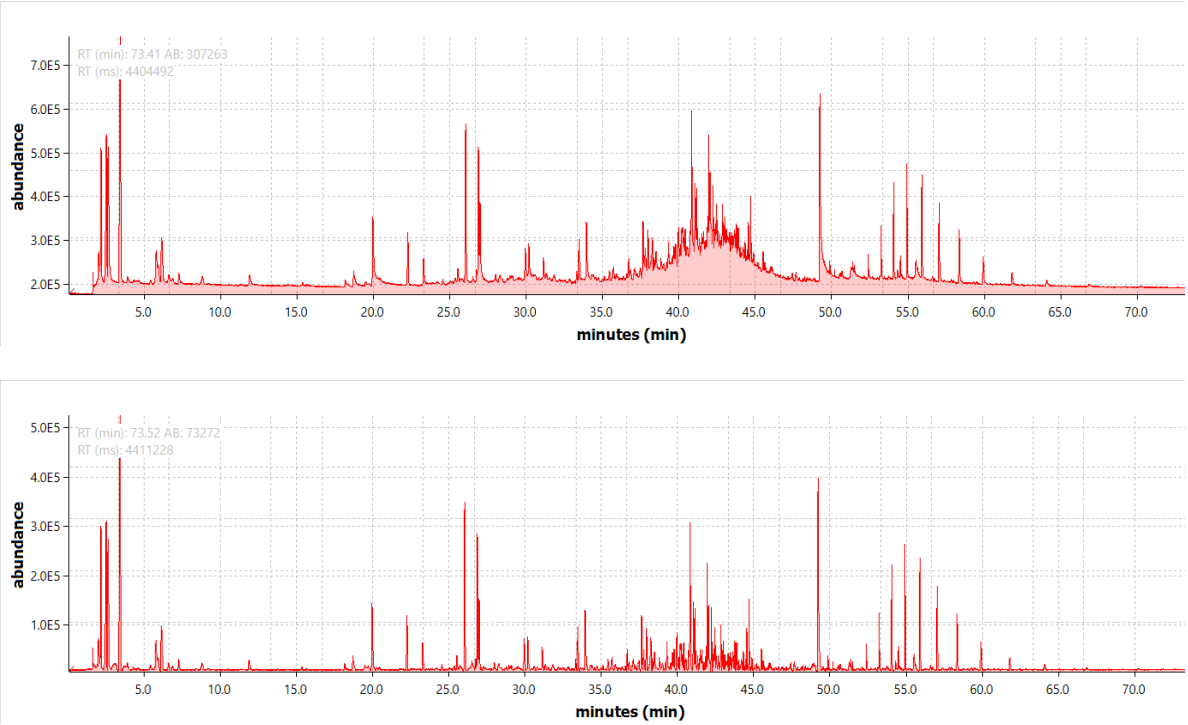

Figure 6. Baseline correction.
Top panel. Uncorrected chromatogram. Bottom panel. Chromatogram with corrected baseline.

3.5.2.2 Peak alignment

RT is not a fixed parameter for the compound and many factors can influence it, including, column length of the GC-MS machine, temperature profile in GC compartment and column coating degradation over the time. Despite use of the same column length and same temperature profile in the analysis of this study, RT shifts between individual files were noticed. In order to group same peaks (same VOCs) among all chromatograms (study

population), RT shifts were eliminated using the Obiwarp method (150) implementation in the R package 'xcms' (173).

The peak alignment is illustrated in Figure 7. The top panel of the figure demonstrates discrepancies between two chromatograms, whereas in the bottom panel the same peaks between two chromatograms appear at similar RT.

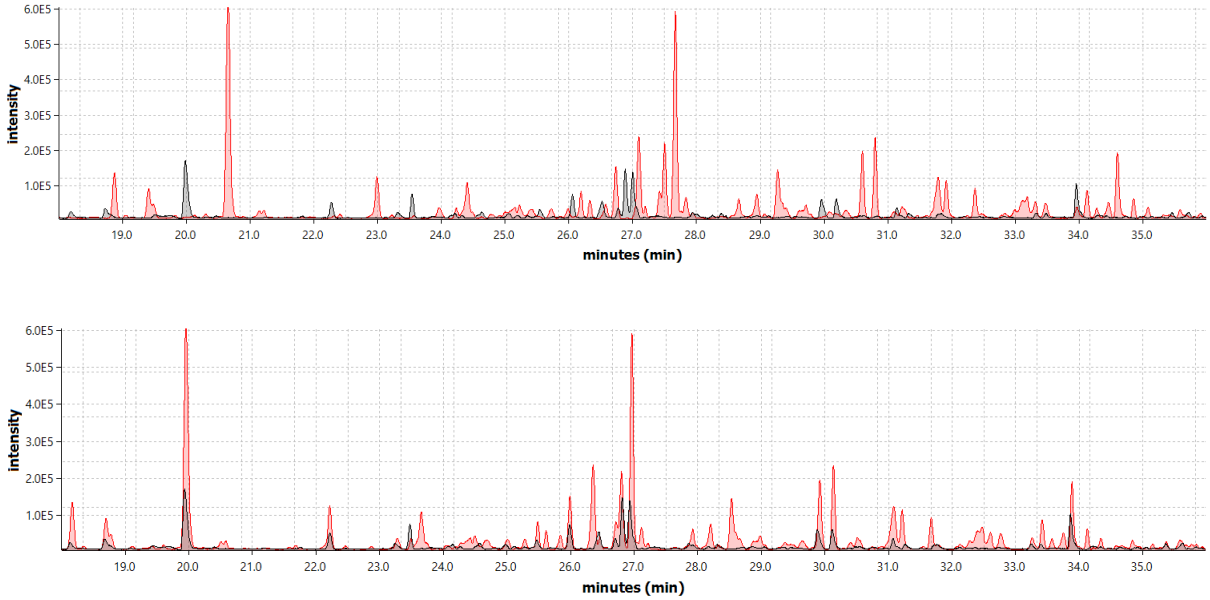

Figure 7. Alignment of chromatograms.

Top panel. The overlay of two chromatograms (red and black) before alignment (18-36 mins). Bottom panel. The overlay of the same two chromatograms after alignment.

3.5.2.3 Extraction of data from chromatograms

The following information of each peak in the single chromatogram was extracted using Openchrom software (197): RT, 5 main M/Z (i.e., most abundant M/Z value for the peak, second most abundant M/Z value, *etc*.), signal to noise ratio (S/N, level of peak signal over the level of background noise), leading value (time between the start of the peak and peaks' center), tailing value (time between the peaks' center and the end of the peak) and area (area under the peak, i.e., sum of all M/Z counts making up the peak).

3.5.2.4 Removal of improper peaks

To select only significant peaks for the analysis, peaks with small area (intensity <10,000) were excluded, as were peaks having 4 out of 5 M/Z values below 34 together with low S/N value (<5) as these peaks most likely are artifacts arising from the GC-system. Additionally, similar M/Z profile of two peaks (e.g., differently ordered same three MZ) in a small RT window (the time between these peaks is less than 0.4 sec) were considered to be one compound, and the one with higher S/N value was included for analysis.

3.5.2.5 Combining individual data files to batches

Individual peak lists of each of the study participants were combined to the final dataset for the analysis in batches. Individual files appeared divided into 10 batches based on systematic differences between individual files depending on GC-MS analysis time. For each of the batches, individual files were combined applying determined rules – starting with very strict rule to combine peaks with exact same RT and exact same 5 M/Z values and continuing with less strict rules, such as peaks are less than 0.4 sec apart and have 2 out of 3 same main M/Z. The combination of individual files to batches was done using the statistical package Stata 14. Finally, the dataset for each batch was checked manually to find peaks that were not covered by the used rules, and these found peaks were also added to the final peak list.

3.5.2.6 Combining batches to final dataset

Prepared final batch specific-datasets were combined together to one dataset based on RT and most common M/Z values among all included peaks. Only peaks having similar abundance levels between batches (percentage of non-missing data among all individual files) were selected for the analysis. I restricted this analysis to the peaks appearing at 0-40 min of the chromatogram following the recommendation of the chemist who conducted GC-MS analysis (19), as for a given GC-MS device and column used, VOCs appearing after 40 min are very heavy molecules that are unlikely to be associated with the disease, e.g., cancer.

3.5.2.7 Peak height (area) normalization

The final dataset consisted of RT (average of individual RT values that were combined for particular peaks), 5 main M/Z values (most common M/Z among all files), abundance (proportion of individuals having particular peak) and area (individual area values of that peak

in each chromatogram). For the statistical analysis, only area is a relevant measurement. Area was normalized using the ratio scale normalization principle (29), where area values were rescaled between 0 and 1. First, squares of area values were summed up and square root calculated. Then, the original area values were divided by the calculated square root value. Normalization was done for each batch separately.

3.6. Statistical methods for associations analysis

Association between measured compounds and factors of interest were assessed using the implementation of the two-part test (183) in Stata 14. Two-part tests are statistical tests combining information on differences in abundance of the peak between two groups ("zero" / "non-zero") together with information on differences in means of area (size of the peak) among those who had that peak detected (98). Missing peak for a study participant can indicate absence of that specific peak for a person, or it can come from technical specification, e.g., peak below detection limit.

Correction for multiple testing was done using Bonferroni adjustment (28) corresponding to the number of volatiles in the final dataset, i.e., fourteen. The alpha level for significance was set to 0.05/14=0.00357.

3.7. Results

In the following subchapters, the final dataset consisting of VOCs selected during data preparation steps is described, followed by results on the association between these VOCs and socio-demographic factors, lifestyle factors, medical conditions and diet.

3.7.1 VCOs included in the analysis

Overall, 14 VOCs were selected for this analysis (Table 7). These VOCs were mostly between 22 and 40 min, except from 3 VOCs appearing at 1.66, 3.45 and 11.80 min. The majority of the selected compounds were present in 88-100% of the samples, meaning that these are common compounds expected to be present in all people in general. Additionally, four VOCs had substantially lower abundance levels, i.e., 30.3, 33.6, 37.7 and 41.0% corresponding to VOC 10, VOC 5, VOC 9 and VOC 12, respectively.

Table 7. VOCs included in the association analysis.

VOC	RT [range]	M/Z 1	M/Z 2	M/Z 3	M/Z 4	M/Z 5	Abundance [range]
VOC 1	1.66 [1.54-1.92]	28	32	14	44	16	97.1
VOC 2	3.45 [3.38-3.51]	42	55	41	39	70	98.8
VOC 3	11.80 [11.43-12.22]	91	92	39	65	63	97.7
VOC 4	22.22 [22.17-22.25]	93	91	39	92	41	96.9
VOC 5	22.44 [22.01-22.77]	57	28	43	41	29	33.6
VOC 6	26.86 [26.73-27.30]	57	41	43	29	71	93.6
VOC 7	29.92 [29.66-30.20]	43	57	41	71	29	90.9
VOC 8	31.07 [30.95-31.12]	73	267	45	268	69	99.7
VOC 9	32.26 [32.05-32.46]	71	41	81	43	55	37.7
VOC 10	33.54 [33.25-33.73]	55	97	41	43	69	30.3
VOC 11	33.88 [33.85-33.93]	43	57	41	71	29	88.7
VOC 12	34.88 [34.62-35.19]	55	57	70	41	28	41.0
VOC 13	37.55 [37.38-37.86]	43	57	41	71	29	91.8
VOC 14	39.90 [30.68-40.16]	57	43	71	41	29	97.5

M/Z, mass to charge ration; RT, retention time; VOC, volatile organic compound.

3.7.2 Study population

Characteristics of the study populations are presented in Table 8. In total, 1,447 individuals from 4 different recruitment centers were included in the analysis. Statistically significant differences (alpha≤0.05) in distributions between the cities were observed for sex, smoking and alcohol consumption, whereas distribution of BMI values between the cities was not different (p-value=0.266). The majority of participants in all four centers were overweight or obese.

Table 8. Study population characteristics: gender, smoking prevalence, alcohol consumption and body mass index.

	Aluksne	**Cesis**	**Ludza**	**Saldus**	**Total**	**p-value[a]**
Total	245	284	412	506	1,447	
Gender						
Men	123 (50.2)	109 (38.4)	252 (61.2)	226 (44.7)	710 (49.1)	**<0.001**
Women	122 (49.8))	175 (61.6)	160 (38.8)	280 (55.3)	737 (50.9)	
Smoking						
Non-smoker	133 (54.3)	157 (55.3)	189 (46.0)	266 (52.8)	745 (51.6)	
Ex-smoker	44 (18.0)	52 (18.3)	92 (22.4)	120 (23.8)	308 (21.3)	**0.034**
Smoker	68 (27.8)	75 (26.4)	130 (31.6)	118 (23.4)	391 (27.1)	
Alcohol consumption						
Non-drinker	36 (14.9)	43 (15.2)	115 (28.2)	158 (31.2)	352 (24.5)	
Light	28 (11.6)	24 (8.5)	22 (5.6)	58 (11.5)	132 (9.2)	**<0.001**
Moderate	90 (37.3)	134 (47.4)	111 (27.2)	166 (32.8)	501 (34.8)	
Heavy	87 (36.1)	82 (29.0)	160 (39.2)	124 (24.5)	453 (31.5)	
Body mass index (grouped)						
Underweight	2 (0.8)	3 (1.1)	2 (0.5)	2 (0.4)	9 (0.6)	
Normal	71 (29.0)	69 (24.3)	106 (27.7)	129 (25.5)	375 (25.9)	0.266
Overweight	83 (33.9)	111 (39.1)	145 (35.2)	216 (42.7)	555 (38.4)	
Obese	89 (36.3)	101 (35.6)	159 (38.6)	159 (31.4)	508 (35.1)	

[a] Chi square test between variables and cities (significant values in **bold** (alpha = 0.05)).

3.7.3 Association between VOCs and socio-demographic and lifestyle factors

Statistically significant associations between socio-demographic and lifestyle factors and VOCs are presented in Table 9. Five VOCs showed significant associations with sex and four with nationality, i.e., between Latvians and Russians. Alcohol consumption showed significant associations for 3 VOCs, although dose-response patterns were not present, as VOC 7 was associated with moderate alcohol consumption while no significant association for this VOC was found in heavy drinkers. No associations were found for smoking.

Table 9. Significant associations between VOCs and socio-demographic and lifestyle factors

Compared groups	VOC 1	VOC 2	VOC 3	VOC 4	VOC 5	VOC 6	VOC 7	VOC 8	VOC 9	VOC 10	VOC 11	VOC 12	VOC 13	VOC 14
Latvians (1,131) vs. Russians (192)			x			x			x				x	
Latvians (1,131) vs. others (124)														
Male (710) vs. female (737)			x			x	x				x		x	
Menopause (yes (339) vs. no(398)														
Non-smoker (745) vs. smoker (391)														
Non-smoker (745) vs. ex-smoker (308)														
Smoked today (352) vs. >1 month ago (1,056)														
Non-drinker vs. light (354 vs. 134)														
Non-drinker vs. moderate (354 vs. 501)							x							
Non-drinker vs. heavy (354 vs. 457)				x										x
Normal BMI vs. underweight (375 vs. 9)														
Normal BMI vs. overweight (375 vs. 555)									x					
Normal BMI vs. obese (375 vs. 508)														

BMI, Body mass index.

Quantitative description of mean, standard deviation, minimum and maximum values of each of the VOCs in men and women are presented in Figure 8 and Table 10. Box plots represents interquartile range of normalized area values (first-third quartiles) with median value highlighted, and whiskers are the upper and lower adjacent values, which are the most extreme area values within [upper end of IQR + 1.5*IQR] and [lower end of IQR – 1.5*IQR], respectively. Outlier values in these graphs were excluded. Box plots were calculated with Stata 14. Box plots for men and women are overlapping for all VOCs.

Wilcoxon test alone that is used to check if two independent samples are selected from the populations having same distribution did not indicate any significant differences between men and women (data not shown). However, two-part test I used (combines a test for zero/non-zero distribution comparison with Wilcoxon test for non-zero data) also calculates statistics for distribution of zero data between the groups that in this specific case corresponds to significant overall findings for several VOCs and sex.

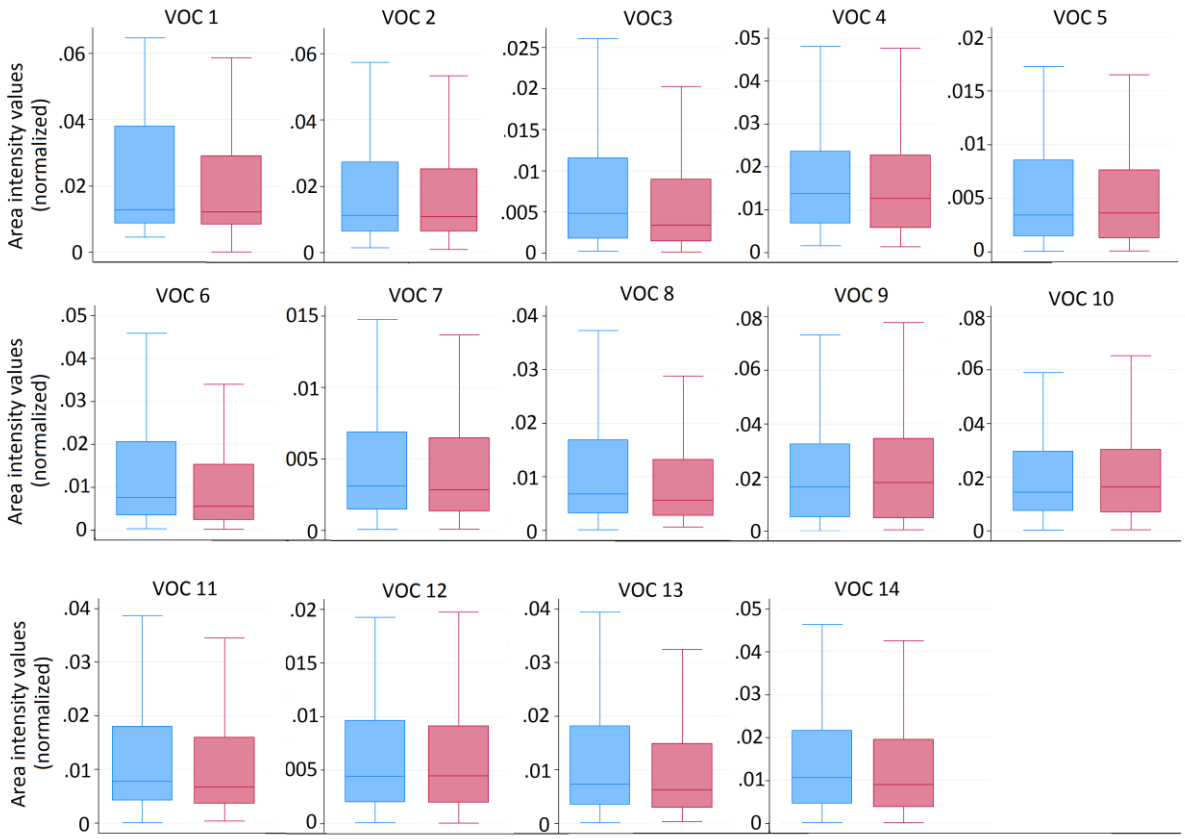

Figure 8. Boxplots of interquartile range of VOC values with median value and upper whisker at [upper end of IQR+1.5*IQR] and lower whiskers at [lower end of IQR–1.5*IQR] (without outliers).

Boxes in blue are for men and boxes in red are for women.

IQR, interquartile range.

Table 10. Descriptive summary of each VOC for men and women.

		Men					Women				
	"0"[a]	**Non-"0"[b]**	**Mean**	**s.d.**	**Min**	**Max**	**Non-"0"[b]**	**Mean**	**s.d.**	**Min**	**Max**
VOC 1	42	695	22.3	17.0	4.6	64.7	710	19.4	15.5	0.1	64.4
VOC 2	17	699	19.3	18.9	1.6	148.0	731	18.3	18.0	1.2	176.7
VOC 3	**33**	**694**	**14.1**	**26.2**	**0.2**	**298.3**	**720**	**11.3**	**19.9**	**0.1**	**152.5**
VOC 4	45	687	19.3	21.3	1.6	226.8	715	17.3	16.9	1.3	138.4
VOC 5	961	252	17.1	58.4	0.1	543.8	234	8.0	15.5	0.1	157.5
VOC 6	**93**	**662**	**16.3**	**28.5**	**0.3**	**511.9**	**692**	**12.0**	**16.2**	**0.2**	**147.7**
VOC 7	**132**	**664**	**11.0**	**24.3**	**0.1**	**206.0**	**651**	**11.6**	**25.5**	**0.1**	**224.6**
VOC 8	5	709	14.5	21.5	0.1	283.0	733	13.4	22.8	0.6	309.6
VOC 9	902	249	26.5	34.7	0.2	246.5	296	27.0	32.3	0.6	255.4
VOC 10	1,009	196	28.6	43.8	0.3	349.2	242	26.5	35.0	0.4	266.7
VOC 11	**163**	**651**	**16.8**	**23.4**	**0.1**	**189.7**	**633**	**15.5**	**21.7**	**0.4**	**123.3**
VOC 12	854	278	16.3	38.4	0.1	292.8	315	15.5	37.4	0.1	248.4
VOC 13	**119**	**670**	**16.4**	**24.4**	**0.2**	**220.8**	**658**	**14.2**	**20.8**	**0.3**	**179.3**
VOC 14	36	698	17.5	22.5	0.1	172.1	713	14.9	19.4	0.1	148.8

s.d., standard deviation; Max, maximum value; Min, minimum value; VOC, volatile organic compound.

[a] number of "zero" values

[b] number of "non-zero" values

For this descriptive table, mean value, standard deviation and minimum and maximum values were multiplied by 1,000.

Lines in bold mark VOCs that were associated with gender (Table 9).

Stratified analysis for each of the cities was performed to investigate the associations between VOCs and gender, smoking status, alcohol consumption and BMI. Results are presented in Table 11. In contrast to results in Table 9, no sex-specific significant associations were found in any of the cities. Significant associations between non-smokers and ex-smokers were found in the city Ludza for 7 VOCs. Furthermore, alcohol consumption was significantly associated with 9 VOCs, where VOC 1 and VOC 2 were found to be significantly associated with moderate alcohol intake in both Ludza and Saldus cities.

Table 11. City-specific significant associations between VOCs and socio-demographic and lifestyle factors.

Compared groups	VOC 1	VOC 2	VOC 3	VOC 4	VOC 5	VOC 6	VOC 7	VOC 8	VOC 9	VOC 10	VOC 11	VOC 12	VOC 13	VOC 14
Male vs. female														
Non-smoker vs. smoker														
Non-smoker vs. ex-smoker	L	L					L	L			L		L	L
Smoked today vs. >1 month ago														
Non-drinker vs. light														
Non-drinker vs. moderate	L,S	L,S	S	S		S	S				S		S	S
Non-drinker vs. heavy		S		S							S		S	
Normal BMI vs. overweight														
Normal BMI vs. obese														

L, Ludza; S, Saldus, VOC, volatile organic compound

3.7.4 Association between VOCs and medical conditions and medications

Prevalences of self-reported medical conditions are presented in Table 12. The most common self-reported illness was gastric or duodenal ulcers (in 12.4% of study participants), followed by thyroid disease reported by 8.1% of study participants. Diseases more common in women than in men were thyroid illness, gallstone disease and allergies other than asthma with 12-fold, 5-fold and 2-fold higher reported prevalence, respectively. Higher prevalence of gastric or duodenal ulcers was reported among men (16.1% *versus* 9.0%). Less common diseases that

Table 12. Number (prevalence in %) of self-reported illnesses.

Disease	Men	Women	Total
Gastric or duodenal ulcer	114 (16.1)	66 (9.0)	180 (12.4)
Thyroid illness	9 (1.3)	108 (14.7)	117 (8.1)
Other allergies	40 (5.6)	76 (10.3)	116 (8.0)
Gallstone disease	18 (2.5)	82 (11.1)	100 (6.9)
Asthma	41 (5.8)	46 (6.2)	87 (6.0)
Diabetes 2	20 (2.8)	25 (3.4)	45 (3.1)
Tuberculosis	11 (1.6)	9 (1.2)	20 (1.4)
Diabetes 1	10 (1.4)	9 (1.2)	19 (1.3)
Diverticulitis or diverticulosis	1 (0.1)	5 (0.7)	6 (0.4)
Ulcerative colitis	2 (0.3)	3 (0.4)	5 (0.4)
Barrett's oesophagus	0 (0)	1 (0.1)	1 (0.1)
Crohn's disease	1 (0.1)	0 (0)	1 (0.1)
Celiac disease	0 (0)	1 (0.1)	1 (0.1)
HIV	1 (0.1)	0 (0)	1 (0.1)

HIV, human immunodeficiency virus.

were not included in the statistical analysis (less than 25 prevalence cases) were tuberculosis (N=20), diabetes type 1 (N=19), diverticulitis or diverticulosis (N=6), ulcerative colitis (N=5) and single cases of Barret's oesophagus, Crohn's disease, celiac disease and infection with human immunodeficiency virus.

Associations between VOCs and self-reported medical conditions together with medication used one month prior recruitment are presented in Table 13. No significant associations were found for 4 out of 6 diseases. VOC 6 was associated with ulcerative colitis (although only 5 individuals had this diseases) and VOC 10 with allergies other than asthma. 5 out of 14 VOCs were shown to be significantly associated with the presence of gastric or duodenal ulcers.

In total, 64 people (4.4%) used antibiotics in the last month before providing a breath sample, however, no significant associations with analyzed VOCs were found. Acid reduction drugs and anti-inflammatory drugs (i.e., aspirin and other NSAIDs) were associated with 1 and 5 out of 14 VOCs, respectively.

Table 13. Significant associations between VOCs and medical conditions and medication use.

Compared groups	**VOC 1**	**VOC 2**	**VOC 3**	**VOC 4**	**VOC 5**	**VOC 6**	**VOC 7**	**VOC 8**	**VOC 9**	**VOC 10**	**VOC 11**	**VOC 12**	**VOC 13**	**VOC 14**
Diseases (N. Yes *versus* N. No)														
Gastric or duodenal ulcer (180 vs. 1,267)			x	x		x							x	x
Thyroid illness[a] (103 vs. 634)														
Other allergies (116 vs. 1,331)										x				
Gallstone disease (100 vs. 1,347)														
Asthma (87 vs. 1,360)														
Diabetes 2 (45 vs. 1,402)														
Medication use (N. Yes *versus* N. No)														
Acid reduction drugs (199 vs. 1,248)				x										
Antibiotics (64 vs. 1,383)														
Anti-inflammatory drugs (451 vs. 996)			x	x			x	x						x

N, number.

[a] only women. Participants with both thyroid illness and diabetes type 1 or 2 were excluded.

3.7.5 Association between VOCs and food items

Dietary patterns were investigated by comparing people eating certain food items not often with people eating them often. Associations between VOCs and food items are presented in Table 14. Significant associations between VOCs and 15 out of 27 different analyzed foods were demonstrated. Among these, VOC 7, VOC 9 and VOC 14 were associated with 8 food items each. Consumption of onion leaves was significantly associated with 8 out of 14 analyzed VOCs. Five of those VOCs were also found to be associated with garlic consumption, and two with onion (onion bulb) as well as garlic consumption. Instant coffee and ground coffee were associated with 4 and 5 VOCs, respectively, and 4 of these VOCs were the same for both.

Table 14. Significant associations between VOCs and consumption of individual food items.

	Food item (N. of eat less often *versus* N. of eat often)	VOC 1	VOC 2	VOC 3	VOC 4	VOC 5	VOC 6	VOC 7	VOC 8	VOC 9	VOC 10	VOC 11	VOC 12	VOC 13	VOC 14
Meat	Beef cooked (1,063 vs. 384)				x		x								x
	Pork cooked (364 vs. 1,083)	x					x								x
	Chicken (604 vs. 843)														
	Meat products (776 vs. 671)														
Dairy products	Kefir (fermented milk) (725 vs. 722)							x	x		x	x			
	Curds (617 vs. 830)														
	Cheese (1,004 vs. 443)														
Greens	Vegetables (802 vs. 645)			x				x		x					
	Onions (867 vs. 580)							x		x					
	Onion leaves (1,095 vs. 352)			x	x		x	x		x			x	x	x
	Garlic (985 vs. 462)			x	x			x		x					x
	Fruits (local) (858 vs. 589)				x										x
	Fruits (imported) (935 vs. 512)							x				x			
Processed products	Pickled products (836 vs. 611)														
	Salted products (899 vs. 548)				x							x			
	Smoked products (858 vs. 589)														
	Dried products (1,096 vs. 351)														
Other products	Porridge (802 vs. 645)				x										
	Bread (dark) (496 vs. 951)														
	Legumes (902 vs. 545)				x			x	x	x					x
	Fish cooked (1,065 vs. 382)														
	Eggs (509 vs. 938)														
Drinks	Tea (green) (902 vs. 545)														
	Tea (black) (842 vs. 605)									x					
	Coffee (instant) (942 vs. 505)		x				x			x					x
	Coffee (ground) (589 vs. 858)		x				x	x		x					x
	Carbonated drinks (1,139 vs. 308)														

N, number; VOC, volatile organic compound

4. Use of breath testing for cancer screening

Preselection of screening populations for further invasive testing can improve overall efficiency of the screening programs through reducing number of invasive follow-up procedures and harms related to them, as well as through lowering the overall screening cost. Clearly, both medically and economically, it is desired to minimize the number of false positive results of a prescreening test which directly translates to the number of unnecessary follow-up procedures. In general, screening tests are usually described based on their ability to identify disease cases (sensitivity) and correctly identify healthy controls (specificity), which are both fixed parameters of the test. In contrast, the numbers of true positive, true negative, false positive and false negative results, as well as predictive values of a screening test are additionally dependent on the prevalence of potentially detectable (preclinical) cancer in the target population of screening (207). Accordingly, predictive values may be quite different for the same screening test, translating into different screening performance in two different populations with differing disease prevalence. However, data on the prevalence of such preclinical cancers and resulting potential differences in screening performance in target populations are seldom reported and remain sparse.

In this chapter, I aim to explore the potential of breath testing for CRC and GC detection. The work flow of this chapter is demonstrated in Figure 9. First, I estimated the prevalence of preclinical CRC and GC in the asymptomatic population, which is the target group for cancer screening in various geographical regions and countries worldwide. Then I used these data together with sensitivity and specificity of breath tests for CRC and GC detection to calculate positive and negative predictive values (PPVs, NPVs) for breath tests in those populations. Both actually reported breath tests as well as hypothetical screening tests are used in this analysis. Additionally, I modeled the expected effect of preselection for more invasive screening using breath tests in a population of 100,000 individuals with various GC prevalence levels.

In this chapter, the terms “breath tests” or “breath testing” are used as a general term describing any test that measures the changes in the concentrations of VOCs in exhaled breath. The results on each of cancer type were prepared and submitted for separate publications – a manuscript with results on CRC has been accepted for publication (95) and manuscript with

results of GC was submitted for publication at the time of submission of this thesis (see List of own publications).

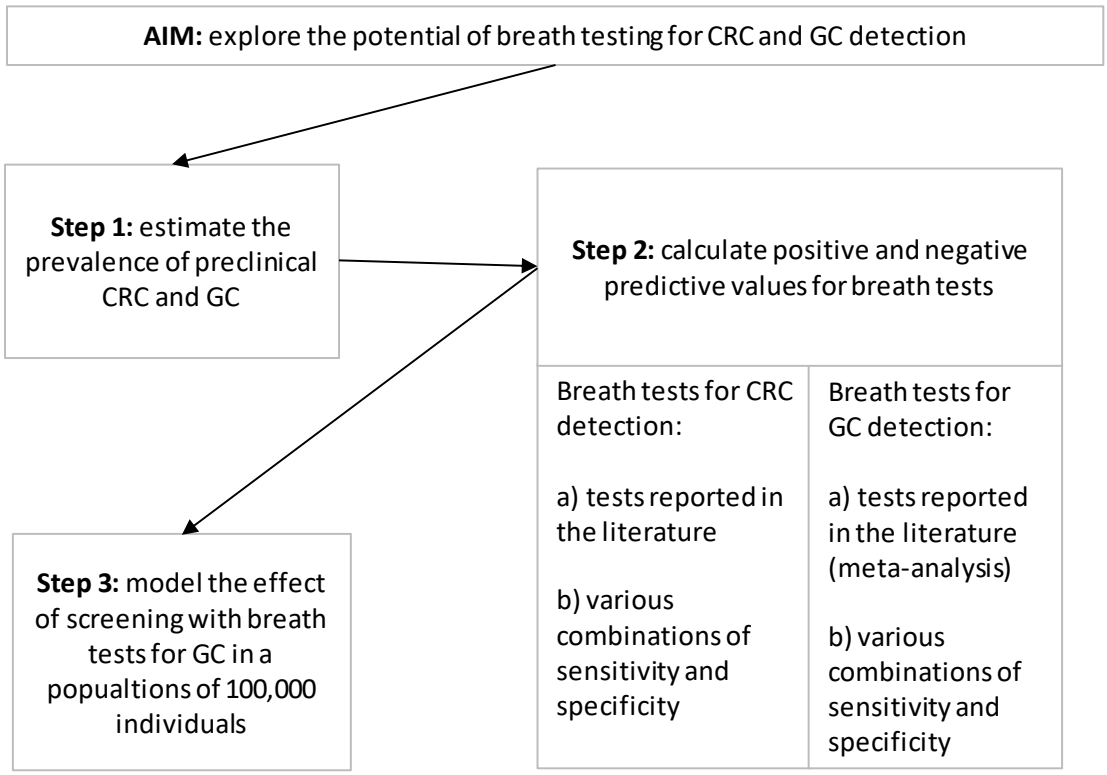

Figure 9. Flow diagram for the current chapter.

4.1. Methods

The performance of breath testing for CRC and GC detection will be described in the context of positive and negative predictive values (PPVs, NPVs) expected in various target populations for screening. The PPVs are of particular interest in this context as they reflect the measured prevalence of CRC and GC in those populations with a positive result from a screening test, and the reciprocal value, the number needed to screen (NNS), reflect the number of participants that would have to undergo invasive screening to detect one cancer case.

4.1.1 Estimation of age- and sex-specific cancer prevalence (unselected target population)

The development of both CRC and GC is a multi-step process (32, 47). The speed at which one development stage turns into another is called annual transition rate between these stages. In the preclinical stage, cancer is still asymptomatic but can be already detected with screening tests (see Figure 10). Targets for screening are preclinical cancers, and the prevalence of preclinical cancer (the proportion of people with of asymptomatic cancer cases in a target population) can be estimated if annual transition rates and cancer incidence are known.

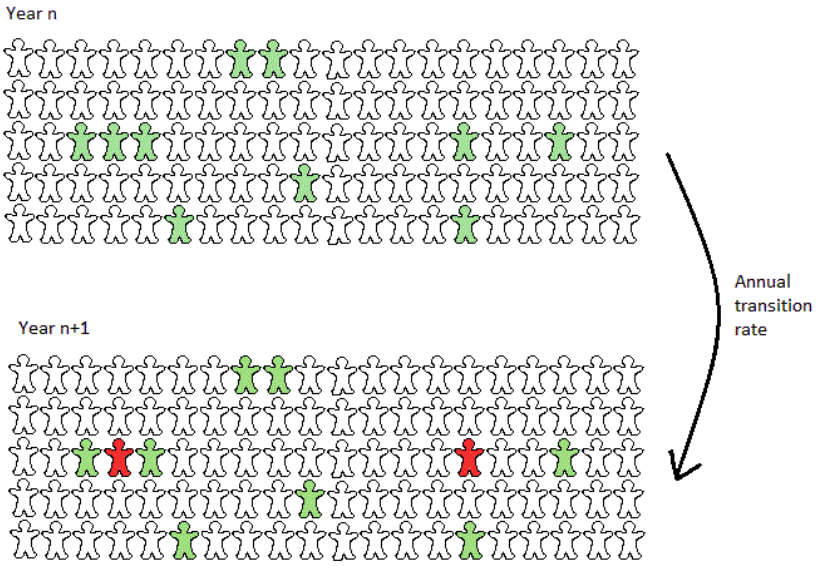

Figure 10. Development of cancer from preclinical to clinically manifested disease.

Top panel. A population with number of already screening-detectable, yet asymptomatic cancers (labeled in green). Bottom panel. Some cancers will manifest symptoms (labeled in red) and will be detected through the symptoms as incident cases.

Using this concept, age- and sex-specific prevalence of preclinical CRC and GC, denoted $P_{age,\ sex}$, was calculated as follows:

$$P_{age,\ sex} = I_{age,\ sex} \ / \ T_{age,\ sex},$$

where $I_{age,sex}$ denotes age- and sex-specific cancer incidence and $T_{age,\ sex}$ denotes age- and sex-specific transition rates, i.e., the annual probability of transition from already existing but not clinically manifest (preclinical) to clinically manifest cancer

Data sources on cancer type-specific transition rates and CRC and GC incidence rates, as well as populations for which CRC and GC prevalence was estimated, are listed below.

4.1.1.1 Transition rates

Colorectal cancer

Transition rates from preclinical to clinical CRC were based on previous analyses results from the German screening colonoscopy registry data (32). Age-specific estimates of annual transition rates ranged from 18.1% (95% CI, 16.7-19.5) to 22.5% (95% CI, 20.9-24.2) which corresponds to mean sojourn times from preclinical to clinical cancers ranging from 5.5 years to 4.5 years, respectively (see Appendix 7). Transition rates for the 50-54 year old population were not available, therefore the same transition rates as for the age group 55-59 were applied. Transition rates in the same order of magnitude, but without stratification by age and/or sex have been estimated for a few other countries (100, 149, 188). Given the lack of detailed transition rates by age and sex from other countries and assuming regional variation of such transition rates to be small, the age- and sex-specific transition rates derived from the German screening colonoscopy registry were applied for all calculations.

Gastric cancer

Transition rates from preclinical to clinical GC were based on results from a study from Korea (14), where transition rates for 50-59 and 60-69 year old men were provided. Reported mean sojourn time between preclinical and clinical GC was 3.18 years (which corresponds to an annual transition rate of 31.4%) for 50-59 year old men and 3.74 years (corresponding to an annual transition rate of 26.7%) for 60-69 year old men (see Appendix 7). The latter estimate was applied for the 70-74 year old populations as well. Given missing data on transition rates for women, we applied the same transition rates for women as for men. Given the lack of data on transition rates from other populations, the age- and sex-specific transition rates derived from the study in Korea (14) were applied for all calculations.

4.1.1.2 Incidence data

CRC and GC incidence data for major geographic regions (Europe: North, West, South and Central-East; America: North, Central and South; Asia: West, South-Central, South-East and East; Africa; and Australia) and individual countries for each of the cancer sites (see cancer-specific subsections below) were taken from the GLOBOCAN 2012 database (57). The list of countries in each geographical region is listed in Appendix 8.

Colorectal cancer

Prevalence of preclinical CRC was estimated for major geographical regions and, to explore the differences within European regions, for individual European countries. CRC prevalence estimates were also calculated for six exemplary countries from each continent: Australia, Brazil (South America), Germany (Europe), India (Asia), Morocco (Africa) and the U.S. (North America).

Gastric cancer

Prevalence of preclinical GC was estimated for major geographical regions and for exemplary countries with high GC incidence (Korea, Japan, China, Albania and Belarus) and relatively low GC incidence (India, the United Kingdom and the U.S.).

4.1.2 Estimation of positive and negative predictive values of breath tests

PPV is the probability that a positive result from a screening test is truly a correct diagnosis, whereas NPV is the probability of negative result from the screening test to be a correct diagnosis. In the screening setting, PPVs is the prevalence of cancer in the populations preselected with screening tests for further diagnostic workup.

Age- and sex-specific positive and negative predictive values, denoted $PPV_{age,\ sex}$ and $NPV_{age,\ sex}$, respectively, were calculated as follows:

$$PPV_{age,\ sex} = \frac{Se * P_{age,\ sex}}{Se * P_{age,\ sex} + (1 - Sp) * (1 - P_{age,\ sex})} \quad \text{and}$$

$$NPV_{age,\ sex} = \frac{Sp * (1 - P_{age,\ sex})}{(1 - Sp) * P_{age,\ sex} + Sp * (1 - P_{age,\ sex})} \quad ,$$

where $P_{age,\ sex}$ denotes the derived prevalence estimates of preclinical CRC and GC (see 4.3.2 and 4.4.1), and *Se* and *Sp* denote the sensitivity and specificity of the screening test for CRC and GC detection. For the calculations of PPVs and NPVs, breath tests and other non-invasive screening tests for comparison were included. These tests are listed in the cancer-specific subsections below.

4.1.2.1 Sensitivity and specificity of breath tests

Colorectal cancer

Age- and sex-specific and overall PPVs and NPVs were calculated for currently available actual screening tests: FIT (sensitivity: 79%, specificity: 94% (102)), the Cologuard test (sensitivity: 92.3%, specificity: 86.6% (82)) and Epi proColon test (sensitivity: 48.2%, specificity: 91.5% (43)) as well as for two breath tests reported in the literature – "breath test 1" described by Amal *et al*. (sensitivity: 85%, specificity: 94% (6)) and "breath test 2" described by Altomare *et al*. (sensitivity: 80%, specificity: 70% (4)).

Other hypothetical PPVs and NPVs potentially achievable with screening tests with various levels of sensitivities and specificities were additionally calculated. Specifically, potential screening tests with combination values of sensitivity and specificity of 70, 80, and 90% each, as well as a test with very high performance level, i.e. a sensitivity and specificity of 95%, were used to derive PPVs and NPVs for CRC detection.

Country-specific PPVs and NPVs were calculated for Australia, Germany, the U.S., Brazil, Morocco and India.

Gastric cancer

To estimate the sensitivity and specificity of breath testing for GC detection, a meta-analysis was performed using studies that evaluated GC detection through exhaled breath analysis. Systematic literature search to identify these studies was performed on the 1st of May, 2017 in Pubmed and Web of Science databases using the following combination of keywords: **(cancer OR carcinoma OR adenocarcinoma OR tumor OR malignancy OR malignant disease) AND ((volatile AND (compound OR compounds OR marker OR markers OR biomarker OR biomarkers)) OR VOC OR VOCs OR breathprint OR breath-print OR breath print) AND (breath OR exhaled OR air) AND (gastric OR stomach).** A similar search strategy to identify studies on breath analysis for cancer detection, yet without restriction to GC, is described in chapter 2. However, in this search, only original studies reporting sensitivity and specificity of breath testing for GC detection or the numbers of true positive and true negative results when discriminating GC patients from healthy controls were eligible for inclusion. Pooled estimates of sensitivity and specificity of breath testing were obtained by

performing a bivariate random-effects meta-analysis using methods proposed by Reitsma *et al.* (153) in the R-package *mada* (52).

Other hypothetical PPVs and NPVs potentially achievable with screening tests with various levels of sensitivities and specificities were calculated. Specifically, potential screening tests with combination values of sensitivity and specificity of 70, 80, and 90% each were used to derive PPVs and NPVs for GC detection.

Country-specific PPVs and NPVs were calculated for Korea, Japan, China, Belarus, Albania, India, the United Kingdom and the U.S.

4.1.3 Overall estimates of prevalence and predictive values

Overall sex-specific prevalence of preclinical CRC and GC in the 50-74 year old population (potential target population for cancer screening) was calculated by weighting age-specific prevalence estimates by the proportions of people in the different age groups in each of analyzed populations. Similarly, overall sex-specific PPVs and NPVs were calculated by weighting age-specific PPVs and NPVs, respectively, by the proportions of people in each of the age groups in each of analyzed populations. The underlying population size, i.e., the proportions of men and women in each of the age groups can be found in Appendix 9. Estimation of number needed to screen to detect one relevant cancer case

4.1.4 Overall estimates of number needed to screen

Age- and sex-specific and overall NNS to detect one CRC and GC case in the populations without prescreening was assessed as one divided by age- and sex-specific and overall prevalence estimates. However, in the population prescreened by breath testing, CRC and GC prevalence is higher compared to that one in the average-risk population. Therefore, age- and sex-specific and overall NNS to detect one CRC and GC case in the populations prescreened with non-invasive tests was calculated as one divided by cancer prevalence in prescreened population, i.e. age- and sex-specific and overall PPVs.

4.1.5 Sensitivity analysis – predictive values of screening tests for different cancer prevalence levels

To demonstrate the impact of the disease prevalence in a target population on predictive values, PPVs and NPVs were calculated for individual screening tests for each prevalence

level between 0.2% and 3% by 0.2 units (such prevalences of preclinical CRC and GC were estimated in different age- and sex- groups).

4.2. Micro-simulation – screening with breath tests for GC in a population of 100,000

For modelling, the potential effect of screening among 100,000 individuals with breath tests was estimated for GC using pooled sensitivity and specificity of breath tests (see 4.4.2). The following parameters were calculated: number of true positive, true negative, false positive and false negative results and NNS to detect one cancer case among 100,000 individuals without preselection by breath tests, as well as PPV, NPV and NNS to detect one cancer case after preselection by breath testing, together with number of total positive tests results. Populations with GC prevalences between 0.01 and 1.6% (the range of GC prevalences observed in different sex and age groups between 50-59 and 70-74 years in different countries, see 4.4.1) were of interest. For this micro-simulation, PPVs were calculated as the ratio of true positives to combined true and false positives, and NPVs were calculated as the ratio of true negatives to combined true and false negatives. NNS in the population without prescreening was assessed as one divided by prevalence and NNS in the population after prescreening by breath testing was calculated as one divided by the GC prevalence in prescreened population, i.e. PPV. The total positives were accessed as the sum of true positive and false positive results from breath testing.

4.3. Results – colorectal cancer

In the following subchapters, the estimated prevalences of preclinical CRC in various geographical regions are presented, followed by PPVs and NPVs of various screening tests expected in exemplary countries from various parts of the world. Finally, detailed PPVs and NPVs are given for populations with varying levels of preclinical CRC.

4.3.1 Prevalence of CRC

Age-specific prevalences of preclinical CRC in major geographical regions of the world are presented in Table 15. Very low rates are found for the 50-54 year old populations in all parts of the world (0.08-0.36% for men and 0.05-0.23% for women), while increasing variation is seen in older age groups, e.g. prevalences of around 0.13% in 70-74 year old men in South-Central Asia and Africa compared to 1.4-1.8% in Europe and Australia. The prevalence of

preclinical CRC in men is higher than in women, with an overall male-to-female prevalence ratio of 1.9 in Southern Europe (0.84% *vs.* 0.44%, respectively) compared to a male-to-female prevalence ratio of 1.3 in South America (0.31% *vs.* 0.23%, respectively). The highest prevalence of preclinical CRC is estimated for Australia in all age groups and for both sexes.

To assess the variation of cancer prevalence within geographical regions, age-specific and overall prevalences of preclinical CRC for all European countries were calculated and are presented in Appendix 10. High variation between prevalence estimates is seen even within the same geographical regions. For example, estimated overall prevalence in 50-74 year old men in Southern Europe varied from 0.19% in Albania to 1.01% in Slovenia. The highest overall prevalences in men were found for Hungary (1.27%), Slovakia (1.19%) and Czech Republic (1.13%), and in women for Denmark (0.68%), the Netherlands and Norway (0.64%, both). The lowest overall prevalence estimates in Europe were seen in Albania, where 0.19% of men and 0.14% of women would have screening detectable preclinical CRC.

Table 15. Prevalence of preclinical colorectal cancer and number needed to screen to detect one cancer case in various geographical regions, by age and sex.

Region or country	Prevalence [%] (number needed to screen)						Ratios in prevalence		
	50-54	55-59	60-64	65-69	70-74	Overall[a]	Age[b]	Sex[c]	Region[d]
Men									
Northern Europe	0.26 (387)	0.49 (204)	0.77 (131)	1.07 (94)	1.51 (67)	0.74 (194)	5.8	1.6	1.2
Western Europe	0.32 (317)	0.58 (174)	0.87 (116)	1.17 (86)	1.60 (63)	0.82 (168)	5.1	1.8	1.1
Southern Europe	0.35 (289)	0.62 (161)	0.92 (109)	1.16 (86)	1.53 (66)	0.84 (157)	4.4	1.9	1.1
Central and Eastern Europe	0.27 (377)	0.53 (190)	0.83 (120)	1.12 (90)	1.42 (71)	0.69 (204)	5.4	1.7	1.3
Northern America	0.31 (320)	0.47 (213)	0.62 (161)	0.76 (132)	1.00 (101)	0.56 (207)	3.2	1.5	1.6
Central America	0.10 (1041)	0.14 (722)	0.18 (565)	0.21 (471)	0.29 (351)	0.16 (715)	3.0	1.4	5.7
South America	0.15 (684)	0.23 (432)	0.34 (291)	0.46 (216)	0.67 (149)	0.31 (423)	4.6	1.3	2.9
Western Asia	0.17 (602)	0.27 (368)	0.39 (258)	0.48 (211)	0.62 (162)	0.31 (388)	3.7	1.6	2.9
South-Central Asia	0.08 (1,284)	0.12 (862)	0.15 (677)	0.17 (599)	0.22 (466)	0.13 (891)	2.8	1.6	7.2
South-Eastern Asia	0.19 (527)	0.31 (322)	0.45 (220)	0.59 (170)	0.82 (123)	0.40 (313)	4.3	1.7	2.2
Eastern Asia	0.14 (694)	0.24 (422)	0.34 (295)	0.41 (245)	0.52 (192)	0.27 (446)	4.3	1.7	2.2
Africa	0.08 (1,207)	0.12 (867)	0.14 (696)	0.17 (587)	0.20 (496)	0.13 (866)	2.4	1.4	7.1
Australia	0.36 (277)	0.66 (152)	0.99 (101)	1.30 (77)	1.77 (57)	0.90 (149)	4.9	1.6	n/a
Women									
Northern Europe	0.18 (548)	0.30 (336)	0.43 (232)	0.66 (152)	0.94 (107)	0.46 (295)	5.1	n/a	1.2
Western Europe	0.20 (497)	0.31 (328)	0.42 (238)	0.61 (164)	0.85 (119)	0.44 (289)	4.2	n/a	1.2
Southern Europe	0.22 (463)	0.33 (307)	0.44 (230)	0.60 (168)	0.79 (128)	0.44 (275)	3.6	n/a	1.2
Central and Eastern Europe	0.19 (536)	0.31 (321)	0.43 (231)	0.60 (167)	0.76 (131)	0.42 (306)	4.1	n/a	1.3

(Table 15 continues on the next page)

`

(*Table 15 continued*)

Region or country	Prevalence [%] (number needed to screen)						Ratios in prevalence		
	50-54	55-59	60-64	65-69	70-74	Overall[a]	Age[b]	Sex[c]	Region[d]
Northern America	0.20 (497)	0.28 (360)	0.36 (280)	0.51 (196)	0.71 (142)	0.36 (327)	3.5	n/a	1.5
Central America	0.07 (1,469)	0.09 (1,071)	0.11 (893)	0.16 (630)	0.23 (434)	0.11 (1,027)	3.4	n/a	4.9
South America	0.11 (892)	0.16 (629)	0.23 (443)	0.35 (283)	0.54 (184)	0.23 (568)	4.9	n/a	2.4
Western Asia	0.12 (863)	0.16 (609)	0.22 (463)	0.29 (349)	0.37 (270)	0.20 (589)	3.2	n/a	2.8
South-Central Asia	0.05 (1,853)	0.07 (1,375)	0.09 (1,166)	0.11 (953)	0.14 (738)	0.08 (1,357)	2.5	n/a	6.9
South-Eastern Asia	0.11 (903)	0.17 (584)	0.24 (420)	0.34 (299)	0.49 (203)	0.23 (545)	4.5	n/a	2.4
Eastern Asia	0.10 (1,000)	0.14 (703)	0.18 (545)	0.24 (411)	0.32 (314)	0.17 (687)	4.5	n/a	2.4
Africa	0.06 (1,626)	0.08 (1,218)	0.10 (1,000)	0.12 (827)	0.14 (703)	0.09 (1,186)	2.3	n/a	6.0
Australia	0.23 (427)	0.38 (265)	0.54 (185)	0.82 (122)	1.16 (86)	0.55 (243)	5.0	n/a	n/a

[a] Weighted sum of age-specific prevalence estimates (population size weights can be found in Appendix 9).

[b] Ratio in prevalence estimates between the age groups 70-74 and 50-54 years old;

[c] Ratio in overall prevalence estimates between men and women;

[d] Ratio in overall prevalence estimates between Australia and other regions.

4.3.2 Predictive values for CRC screening tests

Table 16 provides the calculated overall PPVs and NPVs for a screening test with various levels of sensitivities and specificities in the populations of Australia, Germany, the U.S., Brazil, Morocco and India. In general, PPVs remain below 15% for any screening test, even in countries with relatively high CRC prevalence, such as Australia and Germany. In populations with low disease prevalence (below 0.2%, Morocco and India) all screening tests with the studied characteristics would perform rather poorly - only a PPV of 2.4-3.5% can be achieved with a test with 95% sensitivity and specificity. Age-specific PPVs and NPVs for each of the countries are shown in Appendix 11-Appendix 16.

In Germany, a screening test with a sensitivity and specificity of 90% would result in PPVs of 7.0% in men and 3.6% in women, lower than the positive result from FIT test, which confers 10.5% probability for a true diagnosis for men and 5.2% for women, respectively. PPVs for the Cologuard test (with 92.3% sensitivity and 86.6% specificity (82)) are very similar to those with any screening test with sensitivity of 70% and specificity of 90%. A positive result from the “breath test 2” (80% sensitivity and 70% specificity (4)) in Germany would have PPV of 2.2% in men and 1.1% in women, while a “breath test 1” (85% sensitivity and 94% specificity (6)) would result into PPV of 10.5% in men and 5.6% in women.

The strong dependence of the PPV especially on the test’s specificity can be seen in Table 16. Due to low disease prevalence in asymptomatic population, improving specificity of the screening test from 70% to 90% (at the same sensitivity level) resulted in to 3-fold higher PPV in Germany, e.g., among men aged 50-74, PPV increased from 2.0% to 5.6% (sensitivity = 70%), respectively. At the same time, an increase in the test’s sensitivity level from 70% to 90% (at 70% specificity) only would add 25 percent to PPV (PPV would increase from 2.0 to 2.5%).

Table 16. Overall positive and negative predictive values of various screening tests for colorectal cancer detection in selected countries.

Screening test (sensitivity / specificity [%])	Positive predictive value [%][a] / Negative predictive value [%][a]					
	Australia	**Germany**	**the U.S.**	**Brazil**	**Morocco**	**India**
Men						
Prevalence [%][a]	0.90	0.85	0.53	0.30	0.19	0.13
70 / 70	2.1 / 99.6	2.0 / 99.6	1.2 / 99.8	0.7 / 99.9	0.5 / 99.9	0.3 / 99.9
70 / 80	3.1 / 99.7	2.9 / 99.7	1.8 / 99.8	1.1 / 99.9	0.7 / 99.9	0.5 / 100
70 / 90	5.9 / 99.7	5.6 / 99.7	3.6 / 99.8	2.1 / 99.9	1.3 / 99.9	0.9 / 100
80 / 70	2.4 / 99.7	2.2 / 99.8	1.4 / 99.8	0.8 / 99.9	0.5 / 99.9	0.3 / 100
80 / 80	3.5 / 99.8	3.3 / 99.8	2.1 / 99.9	1.2 / 99.9	0.8 / 100	0.5 / 100
80 / 90	6.7 / 99.8	6.3 / 99.8	4.1 / 99.9	2.4 / 99.9	1.5 / 100	1.0 / 100
90 / 70[b]	2.6 / 99.9	2.5 / 99.9	1.6 / 99.9	0.9 / 100	0.6 / 100	0.4 / 100
90 / 80	3.9 / 99.9	3.7 / 99.9	2.3 / 99.9	1.3 / 100	0.9 / 100	0.6 / 100
90 / 90	7.5 / 99.9	7.0 / 99.9	4.6 / 99.9	2.7 / 100	1.7 / 100	1.1 / 100
95 / 95	14.3 / 100	13.5 / 100	9.1 / 100	5.4 / 100	3.5 / 100	2.4 / 100
Breath test 1[c]	11.2 / 99.9	10.5 / 99.9	7.0 / 99.9	4.1 / 100	2.7 / 100	1.8 / 100
FIT[d]	10.5 / 99.8	9.9 / 99.8	6.5 / 99.9	3.8 / 99.9	2.5 / 100	1.7 / 100
Cologuard[e]	5.8 / 99.9	5.5 / 99.9	3.5 / 100	2.0 / 100	1.3 / 100	0.9 / 100
Epi proColon[f]	4.9 / 99.5	4.6 / 99.5	2.9 / 99.7	1.7 / 99.8	1.1 / 99.9	0.7 / 99.9
Women						
Prevalence [%][a]	0.55	0.42	0.35	0.23	0.13	0.08
70 / 70	1.3 / 99.8	1.0 / 99.8	0.8 / 99.8	0.5 / 99.9	0.3 / 99.9	0.2 / 100
70 / 80	1.9 / 99.8	1.4 / 99.8	1.2 / 99.9	0.8 / 99.9	0.4 / 100	0.3 / 100
70 / 90	3.7 / 99.8	2.8 / 99.9	2.4 / 99.9	1.6 / 99.9	0.9 / 100	0.5 / 100
80 / 70	1.5 / 99.8	1.1 / 99.9	0.9 / 99.9	0.6 / 99.9	0.3 / 100	0.2 / 100
80 / 80	2.2 / 99.9	1.6 / 99.9	1.4 / 99.9	0.9 / 99.9	0.5 / 100	0.3 / 100
80 / 90	4.2 / 99.9	3.2 / 99.9	2.7 / 99.9	1.8 / 99.9	1.0 / 100	0.6 / 100
90 / 70[b]	1.6 / 99.9	1.2 / 99.9	1.0 / 99.9	0.7 / 100	0.4 / 100	0.2 / 100
90 / 80	2.4 / 99.9	1.9 / 99.9	1.6 / 100	1.0 / 100	0.6 / 100	0.4 / 100
90 / 90	4.7 / 99.9	3.6 / 100	3.1 / 100	2.0 / 100	1.1 / 100	0.7 / 100
95 / 95	9.3 / 100	7.3 / 100	6.2 / 100	4.2 / 100	2.4 / 100	1.5 / 100

(Table 16 continues on the next page)

(Table 16 continued)

Screening test (sensitivity / specificity [%])	Positive predictive value [%][a] / Negative predictive value [%][a]					
	Australia	Germany	the U.S.	Brazil	Morocco	India
Breath test 1[c]	7.2 / 99.9	5.6 / 99.9	4.7 / 99.9	3.2 / 100	1.8 / 100	1.1 / 100
FIT[d]	6.7 / 99.9	5.2 / 99.9	4.4 / 99.9	2.9 / 99.9	1.7 / 100	1.0 / 100
Cologuard[e]	3.7 / 100	2.8 / 100	2.4 / 100	1.6 / 100	0.9 / 100	0.5 / 100
Epi proColon[f]	3.0 / 99.7	2.3 / 99.8	1.9 / 99.8	1.3 / 99.9	0.7 / 99.9	0.4 / 100

[a] Weighted sum of age-specific estimates using region-specific relative population size weights (weights can be found in Appendix 9).

[b] Breath test 2 by Altomare *et al*., sensitivity 90%, specificity 70% (4).

[c] Breath test 1 by Amal *et al*, sensitivity 85%, specificity 94% (6).

[d] Fecal immunochemical test (FIT), sensitivity 79% and specificity 94% (102).

[e] Cologuard, sensitivity 92.3% and specificity 86.6% (82).

[f] Epi proColon, sensitivity 48.2% and specificity 91.5% (43).

4.3.3 Predictive values for CRC detection – sensitivity analysis

Positive and negative predictive values for CRC prevalences between 0.2 and 3% were calculated. Based on the results of age- and sex-specific prevalence estimates in various regions and countries (see Table 15 and Appendix 10), such prevalences would be expected between different age groups in countries with low and high CRC incidence.

PPVs are presented in Table 17. In general, PPVs exceeded 10% in populations with the prevalence as low as 1.2, 1.4 and 1.6% for the tests with 90% specificity and 90%, 80% and 70% sensitivities, respectively. Among the actual screening tests, PPVs higher than 10% are also expected with FIT, Cologuard and Epi proColon in populations with at least 0.8, 1.6 and 2% CRC prevalence, respectively. PPVs of FIT and "breath test 1" are consistently similar – the difference remains less than 2 percent units for all prevalence levels, with "breath test 1" performing slightly better than FIT.

Table 18 presents calculated NPVs for prevalences between 0.2 and 3%. In general, NPVs stayed between 98 and 100% for all tests and between 99.5 and 100% for most of the assessed scenarios. The NPV drops to 98% only for the Epi ProColon test in populations with at least 2.8% prevalence of preclinical CRC.

Table 17. Positive predictive values for various screening tests for populations with different prevalence of preclinical CRC.

Screening test[a]	**Positive predictive value [%]**														
Prevalence [%]	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.6	2.8	3.0
70 / 70	0.5	0.9	1.4	1.8	2.3	2.8	3.2	3.7	4.1	4.5	5.0	5.4	5.9	6.3	6.7
70 / 80	0.7	1.4	2.1	2.7	3.4	4.1	4.7	5.4	6.0	6.7	7.3	7.9	8.5	9.2	9.8
70 / 90	1.4	2.7	4.1	5.3	6.6	7.8	9.0	10.2	11.4	12.5	13.6	14.7	15.7	16.8	17.8
80 / 70	0.5	1.1	1.6	2.1	2.6	3.1	3.6	4.2	4.7	5.2	5.7	6.2	6.6	7.1	7.6
80 / 80	0.8	1.6	2.4	3.1	3.9	4.6	5.4	6.1	6.8	7.5	8.3	9.0	9.6	10.3	11.0
80 / 90	1.6	3.1	4.6	6.1	7.5	8.9	10.2	11.5	12.8	14.0	15.3	16.4	17.6	18.7	19.8
90 / 70[b]	0.6	1.2	1.8	2.4	2.9	3.5	4.1	4.7	5.2	5.8	6.3	6.9	7.4	8.0	8.5
90 / 80	0.9	1.8	2.6	3.5	4.3	5.2	6.0	6.8	7.6	8.4	9.2	10.0	10.7	11.5	12.2
90 / 90	1.8	3.5	5.2	6.8	8.3	9.9	11.3	12.8	14.2	15.5	16.8	18.1	19.4	20.6	21.8
95 / 95	3.7	7.1	10.3	13.3	16.1	18.8	21.2	23.6	25.8	27.9	29.9	31.8	33.7	35.4	37.0
Breath test 1[c]	2.8	5.4	7.9	10.3	12.5	14.7	16.7	18.7	20.6	22.4	24.2	25.8	27.4	29.0	30.5
FIT[d]	2.6	5.0	7.4	9.6	11.7	13.8	15.8	17.6	19.4	21.2	22.9	24.5	26.0	27.5	28.9
Cologuard[e]	1.4	2.7	4.0	5.3	6.5	7.7	8.9	10.1	11.2	12.3	13.4	14.5	15.5	16.6	17.6
Epi proColon[f]	1.1	2.2	3.3	4.4	5.4	6.4	7.5	8.4	9.4	10.4	11.3	12.2	13.1	14.0	14.9

[a] sensitivity [%] / specificity [%] of hypothetical test or actual test.

[b] Breath test 2 by Altomare *et al*., sensitivity 90%, specificity 70% (4).

[c] Breath test 1 by Amal *et al*, sensitivity 85%, specificity 94% (6).

[d] Fecal immunochemical test (FIT), sensitivity 79% and specificity 94% (102).

[e] Cologuard, sensitivity 92.3% and specificity 86.6% (82).

[f] Epi proColon, sensitivity 48.2% and specificity 91.5% (43).

Positive predictive value in grey marks estimates exceeding 5%.

Table 18. Negative predictive values for various screening tests for populations with different prevalence of preclinical CRC.

Screening test[a]	Negative predictive value [%]														
Prevalence [%]	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.6	2.8	3.0
70 / 70	99.9	99.8	99.7	99.7	99.6	99.5	99.4	99.3	99.2	99.1	99.0	99.0	98.9	98.8	98.7
70 / 80	99.9	99.8	99.8	99.7	99.6	99.5	99.5	99.4	99.3	99.2	99.2	99.1	99.0	98.9	98.9
70 / 90	99.9	99.9	99.8	99.7	99.7	99.6	99.5	99.5	99.4	99.3	99.3	99.2	99.1	99.0	99.0
80 / 70	99.9	99.9	99.8	99.8	99.7	99.7	99.6	99.5	99.5	99.4	99.4	99.3	99.2	99.2	99.1
80 / 80	99.9	99.9	99.8	99.8	99.7	99.7	99.6	99.6	99.5	99.5	99.4	99.4	99.3	99.3	99.2
80 / 90	100	99.9	99.9	99.8	99.8	99.7	99.7	99.6	99.6	99.5	99.5	99.5	99.4	99.4	99.3
90 / 70[b]	100	99.9	99.9	99.9	99.9	99.8	99.8	99.8	99.7	99.7	99.7	99.6	99.6	99.6	99.6
90 / 80	100	99.9	99.9	99.9	99.9	99.8	99.8	99.8	99.8	99.7	99.7	99.7	99.7	99.6	99.6
90 / 90	100	100	99.9	99.9	99.9	99.9	99.8	99.8	99.8	99.8	99.8	99.7	99.7	99.7	99.7
95 / 95	100	100	100	100	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.8	99.8
Breath test 1[c]	100	99.9	99.9	99.9	99.8	99.8	99.8	99.7	99.7	99.7	99.6	99.6	99.6	99.5	99.5
FIT[d]	100	99.9	99.9	99.8	99.8	99.7	99.7	99.6	99.6	99.5	99.5	99.5	99.4	99.4	99.3
Cologuard[e]	100	100	99.9	99.9	99.9	99.9	99.9	99.9	99.8	99.8	99.8	99.8	99.8	99.7	99.7
Epi proColon[f]	99.9	99.8	99.7	99.5	99.4	99.3	99.2	99.1	99.0	98.9	98.7	98.6	98.5	98.4	98.3

[a] sensitivity [%] / specificity [%] of hypothetical test or actual test.

[b] Breath test 2 by Altomare *et al*., sensitivity 90%, specificity 70% (4).

[c] Breath test 1 by Amal *et al*, sensitivity 85%, specificity 94% (6).

[d] Fecal immunochemical test (FIT), sensitivity 79% and specificity 94% (102).

[e] Cologuard, sensitivity 92.3% and specificity 86.6% (82).

[f] Epi proColon, sensitivity 48.2% and specificity 91.5% (43).

Negative predictive value in grey marks estimates exceeding 99.5%.

4.4. Results - Gastric Cancer

In the following subchapters, the estimated prevalences of preclinical GC in various geographical regions and countries are presented, followed by the results on pooled sensitivity and specificity estimates of breath testing for GC detection derived by meta-analysis. Then, calculated expected PPVs in various populations achievable with breath testing are presented. Finally, results describing the expected effect of screening an asymptomatic population of 100,000 individuals with various levels of GC prevalence are given.

4.4.1 Prevalence of preclinical GC

Age-specific and overall prevalence of preclinical GC in various regions of the world together with NNS to detect one GC are presented in Table 19. The lowest detectable GC prevalences are expected in Africa, Australia and North America, where only 1 out of >1,095 men (0.05-0.09%) and 1 out of >2,495 women (0.03-0.04%) aged 50-74 years would have preclinical GC. By contrast, 1 out of 220 men (0.46%) and 1 out of 608 women (0.16%) in Eastern Asia are expected to have preclinical GC, followed by Central-Eastern Europe, with GC prevalence of 0.28% among men and 0.12% among women. Among individual countries, estimated prevalence of detectable GC was the highest in Korea in all age groups and for both men and women. Higher than 1% prevalence was estimated for men in Korea and Japan, but only for those older than 60 and 70 years of age, respectively. Compared to men, substantially lower GC prevalence was estimated among women, for whom expected overall prevalence varied from <0.05% in the U.S. and the United Kingdom to around 0.30% in Korea.

Table 19. Prevalence of preclinical gastric cancer and number needed to screen to detect one cancer case in various populations.

Region or country	Prevalence [%] (number needed to screen)			
	50-59	60-69	70-74	Overall[a]
Men				
Northern Europe	0.04 (2,334)	0.13 (779)	0.24 (420)	0.10 (960)
Western Europe	0.05 (1,854)	0.15 (678)	0.28 (362)	0.12 (822)
Southern Europe	0.08 (1,289)	0.21 (480)	0.35 (287)	0.17 (602)
Central and Eastern Europe	0.15 (647)	0.39 (257)	0.60 (168)	0.28 (355)
Northern America	0.04 (2,781)	0.10 (1,018)	0.16 (634)	0.07 (1,385)
Central America	0.06 (1,560)	0.18 (559)	0.31 (320)	0.13 (775)
South America	0.09 (1,137)	0.25 (399)	0.44 (227)	0.18 (557)
Western Asia	0.08 (1,186)	0.22 (458)	0.33 (299)	0.15 (674)
South-Central Asia	0.07 (1,378)	0.17 (585)	0.23 (442)	0.12 (843)
South-Eastern Asia	0.07 (1,494)	0.14 (730)	0.20 (505)	0.10 (993)
Eastern Asia	0.23 (427)	0.62 (161)	1.00 (100)	0.46 (220)
Africa	0.04 (2,846)	0.07 (1,363)	0.11 (944)	0.05 (1,824)
Australia	0.04 (2,541)	0.12 (842)	0.21 (473)	0.09 (1,095)
Korea	0.50 (202)	1.19 (84)	1.60 (63)	0.84 (119)
Japan	0.27 (368)	0.91 (111)	1.54 (65)	0.76 (132)
China	0.22 (446)	0.56 (178)	0.88 (114)	0.41 (245)
Albania	0.18 (547)	0.42 (241)	0.70 (144)	0.32 (310)
Belarus	0.22 (446)	0.58 (174)	0.81 (124)	0.39 (255)
India	0.07 (1,462)	0.16 (609)	0.21 (484)	0.11 (894)
The United Kingdom	0.03 (3,164)	0.11 (905)	0.22 (455)	0.09 (1,115)
The U.S.	0.04 (2,847)	0.10 (1,053)	0.15 (663)	0.07 (1,432)
Women				
Northern Europe	0.02 (4,857)	0.06 (1,791)	0.11 (945)	0.05 (2,100)
Western Europe	0.03 (3,341)	0.06 (1,635)	0.11 (879)	0.06 (1,807)
Southern Europe	0.04 (2,771)	0.09 (1,158)	0.15 (668)	0.07 (1,357)
Central and Eastern Europe	0.06 (1,594)	0.16 (637)	0.25 (396)	0.12 (804)
Northern America	0.02 (5,613)	0.04 (2,510)	0.07 (1,469)	0.03 (3,112)
Central America	0.06 (1,751)	0.13 (782)	0.21 (475)	0.10 (1,026)

(Table 19 continues on the next page)

(Table 19 Continued)

Region or country	Prevalence [%] (number needed to screen)			
	50-59	60-69	70-74	Overall[a]
South America	0.04 (2,489)	0.11 (936)	0.21 (480)	0.08 (1,213)
Western Asia	0.05 (1,936)	0.13 (787)	0.20 (512)	0.09 (1,091)
South-Central Asia	0.04 (2,643)	0.07 (1,514)	0.09 (1,175)	0.05 (1,919)
South-Eastern Asia	0.03 (3,014)	0.07 (1,513)	0.09 (1,062)	0.05 (1,990)
Eastern Asia	0.09 (1,171)	0.21 (471)	0.36 (275)	0.16 (608)
Africa	0.02 (4,398)	0.06 (1,748)	0.07 (1,342)	0.04 (2,495)
Australia	0.02 (5,177)	0.05 (2,020)	0.08 (1,180)	0.04 (2,545)
Korea	0.17 (580)	0.39 (256)	0.55 (181)	0.30 (339)
Japan	0.11 (878)	0.25 (400)	0.45 (224)	0.23 (426)
China	0.08 (1,253)	0.20 (496)	0.34 (293)	0.15 (661)
Albania	0.09 (1,062)	0.27 (368)	0.42 (238)	0.20 (508)
Belarus	0.09 (1,157)	0.22 (460)	0.32 (310)	0.17 (605)
India	0.03 (2,902)	0.06 (1,722)	0.07 (1,385)	0.05 (2,162)
The United Kingdom	0.02 (6,552)	0.05 (2,172)	0.09 (1,092)	0.04 (2,560)
The U.S.	0.02 (5,629)	0.04 (2,554)	0.07 (1,505)	0.03 (3,158)

[a] weighted sum of age-specific estimates (population size weights can be found in Appendix 9).

4.4.2 Performance of breath test for GC detection – meta-analysis

The pooled estimates of sensitivity and specificity of breath tests were derived by meta-analysis from the studies identified in the systematic review. Details of the study selection process for the systematic review are shown in Figure 11. In total, 51 papers were identified, of which 10 remained for full text review after removal of duplicates and application of eligibility criteria. Of these, 4 reported sensitivity and specificity of breath testing for GC detection. (7, 96, 167, 202). Further details on the included studies can be found in Appendix 17. Data from a total of 98 GC patients and 255 healthy controls were available to derive pooled sensitivity and specificity estimates of breath testing for GC detection.

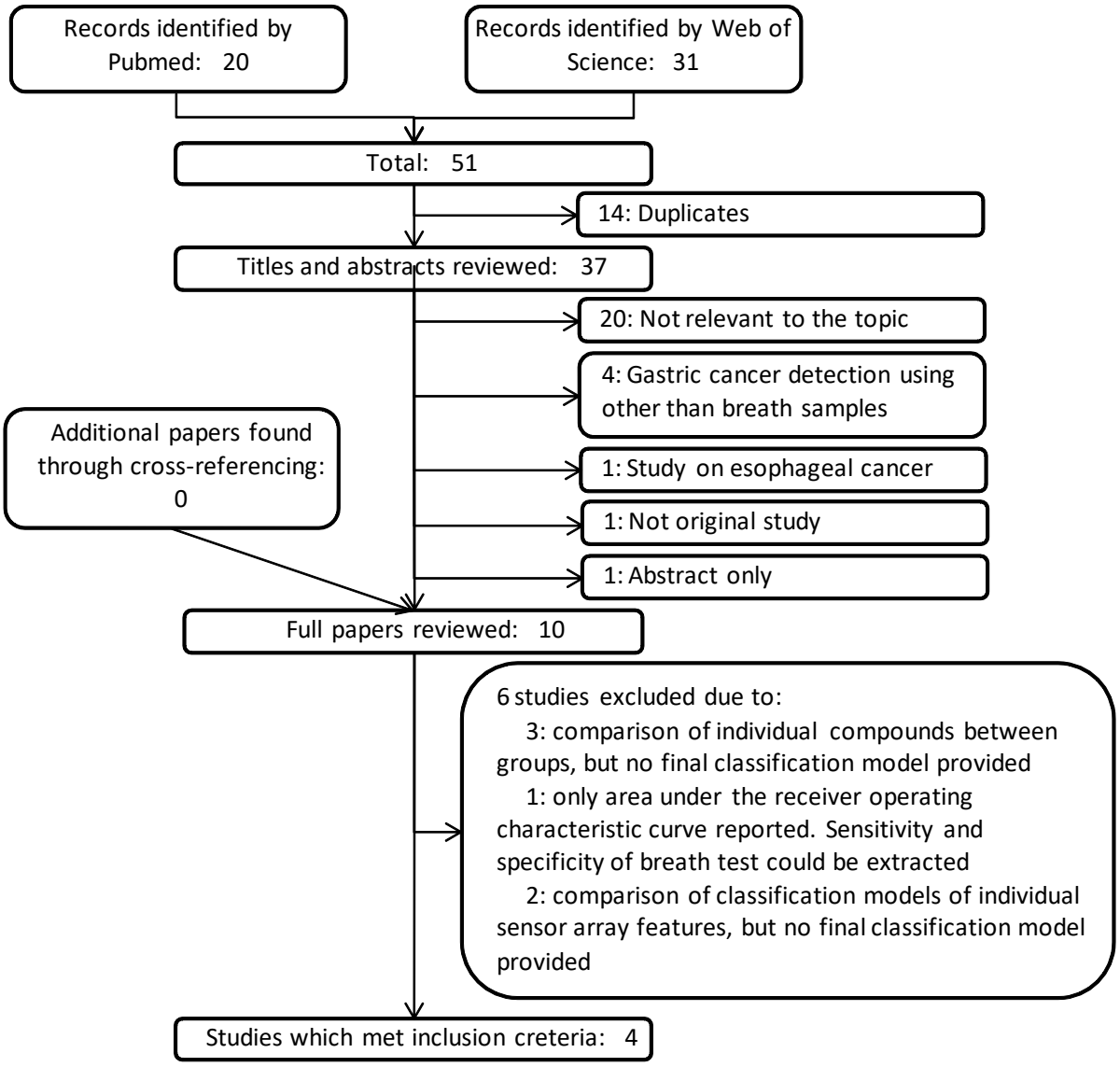

Figure 11. Flow diagram for literature search process to identify studies on exhaled breath analysis for GC detection.

Pooled summary estimates of sensitivity and specificity of breath test performance are shown in Figure 12. Estimated sensitivity and specificity was 82.7% (CI, 71.0 - 90.4%) and 91.4% (CI, 80.4 - 96.5%), respectively.

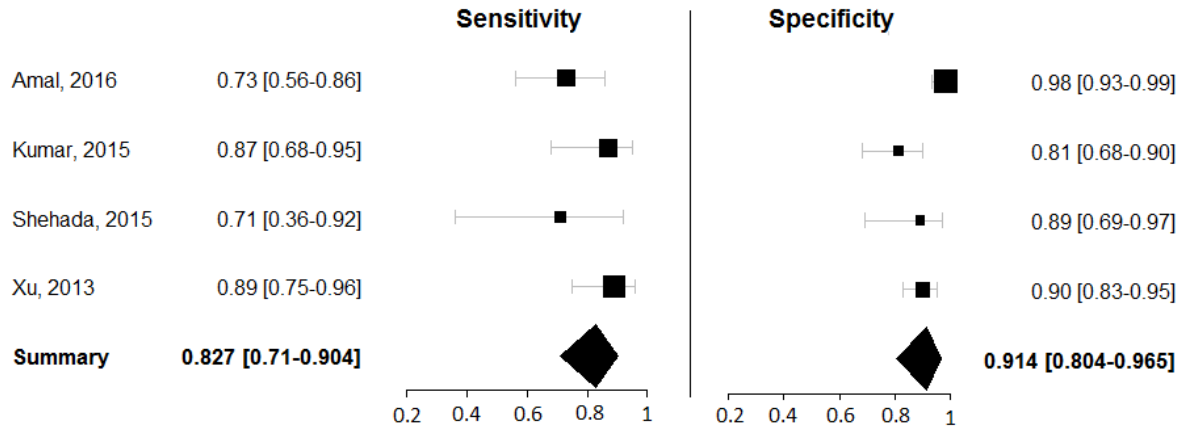

Figure 12. Summary pooled estimates of sensitivity and specificity of breath tests for gastric cancer detection.

4.4.3 Prevalence of GC in populations preselected for screening with breath testing

To estimate the prevalence of GC among those with a positive breath test result, expected PPVs were calculated combining estimates of sensitivity and specificity derived by meta-analysis (82.7% and 91.4%, respectively) in different populations with various levels of GC prevalence. These results and NNS to detect one GC case in prescreened populations are presented in Table 20. PPVs in the 50-74 year old population varied between 0.5-4.2% in men and 0.3-1.6% in women, with the highest PPVs in men seen in Eastern Asia, Central-East Europe and South America, with overall PPVs of 4.2%, 2.6% and 1.7%, respectively. PPVs between 5 and 14% may be achieved with breath tests in 60 years and older men in Korea, Japan, Belarus and China and 70 years and older men in Albania. As for the women, overall PPVs of 1.6 and 1.2% are expected in Eastern Asia and Central-East Europe. Highest PPVs were estimated for women in Korea and Japan, where estimates increased from >1% up to 5% between age groups 50-59 and 70-74.

A comparison of the prevalences of preclinical GC in the unscreened populations (provided in Table 19) and the PPVs, which reflect the prevalences in those prescreened with the positive breath test (provided in Table 20), shows that the latter estimates are close to 10-fold higher than the former across all age groups, both sexes, and all populations.

The NPVs consistently stayed between 99-100% in all analyzed populations, regardless of age, gender and geographical location (Appendix 18).

Table 20. Positive predictive value of breath test for gastric cancer detection and number needed to screen to detect one cancer in various geographical regions, by age and sex.

Region or country	Prevalence [%] (number needed to screen)			
	50-59	60-69	70-74	Overall[a]
Men				
Northern Europe	0.41 (243)	1.23 (82)	2.26 (45)	0.99 (101)
Western Europe	0.52 (193)	1.41 (72)	2.60 (39)	1.16 (87)
Southern Europe	0.74 (135)	1.98 (51)	3.27 (31)	1.57 (64)
Central and Eastern Europe	1.47 (68)	3.65 (28)	5.47 (19)	2.64 (38)
Northern America	0.35 (289)	0.94 (107)	1.50 (67)	0.69 (145)
Central America	0.62 (163)	1.70 (59)	2.94 (34)	1.23 (82)
South America	0.84 (119)	2.37 (43)	4.11 (25)	1.70 (59)
Western Asia	0.81 (124)	2.07 (49)	3.14 (32)	1.41 (71)
South-Central Asia	0.70 (144)	1.63 (62)	2.15 (47)	1.13 (89)
South-Eastern Asia	0.64 (156)	1.31 (77)	1.88 (54)	0.96 (104)
Eastern Asia	2.22 (46)	5.69 (18)	8.92 (12)	4.18 (24)
Africa	0.34 (296)	0.70 (143)	1.01 (99)	0.53 (191)
Australia	0.38 (265)	1.13 (89)	2.00 (50)	0.87 (115)
Korea	4.60 (22)	10.42 (10)	13.54 (8)	7.46 (14)
Japan	2.57 (39)	8.12 (13)	13.11 (8)	6.76 (15)
China	2.13 (48)	5.19 (20)	7.92 (13)	3.77 (27)
Albania	1.74 (58)	3.87 (26)	6.36 (16)	3.01 (34)
Belarus	2.13 (48)	5.30 (19)	7.33 (14)	3.64 (28)
India	0.66 (153)	1.56 (64)	1.96 (51)	1.07 (94)
The United Kingdom	0.30 (329)	1.06 (95)	2.09 (48)	0.86 (117)
The U.S.	0.34 (296)	0.91 (110)	1.44 (70)	0.67 (150)
Women				
Northern Europe	0.20 (505)	0.54 (187)	1.01 (99)	0.46 (219)
Western Europe	0.29 (348)	0.59 (171)	1.09 (92)	0.53 (189)
Southern Europe	0.35 (288)	0.83 (121)	1.43 (71)	0.71 (142)
Central and Eastern Europe	0.60 (166)	1.49 (67)	2.39 (42)	1.18 (85)
Northern America	0.17 (583)	0.38 (261)	0.65 (154)	0.31 (324)
Central America	0.55 (183)	1.22 (82)	2.00 (51)	0.93 (108)
South America	0.39 (259)	1.02 (98)	1.98 (51)	0.79 (127)
Western Asia	0.50 (202)	1.21 (83)	1.85 (54)	0.88 (115)

(Table 20 continues on the next page)

(Table 20 continued)

Region or country	Prevalence [%] (number needed to screen)			
	50-59	**60-69**	**70-74**	**Overall[a]**
South-Central Asia	0.36 (275)	0.63 (158)	0.82 (123)	0.50 (200)
South-Eastern Asia	0.32 (314)	0.63 (158)	0.90 (111)	0.48 (208)
Eastern Asia	0.82 (123)	2.02 (50)	3.41 (30)	1.56 (65)
Africa	0.22 (457)	0.55 (182)	0.71 (140)	0.39 (260)
Australia	0.19 (538)	0.48 (211)	0.81 (124)	0.38 (265)
Korea	1.64 (61)	3.66 (28)	5.09 (20)	2.76 (37)
Japan	1.09 (92)	2.36 (43)	4.15 (25)	2.21 (46)
China	0.77 (131)	1.91 (53)	3.20 (32)	1.44 (70)
Albania	0.90 (111)	2.56 (40)	3.92 (26)	1.86 (54)
Belarus	0.83 (121)	2.06 (49)	3.03 (34)	1.57 (64)
India	0.33 (302)	0.56 (180)	0.69 (145)	0.44 (225)
The United Kingdom	0.15 (680)	0.44 (226)	0.88 (115)	0.38 (267)
The U.S.	0.17 (585)	0.38 (266)	0.64 (157)	0.30 (329)

[a] weighted sum of age-specific estimates (population size weights can be found in Appendix 9).

4.4.4 Predictive values for GC detection – sensitivity analysis

To consider different performances of breath tests, predictive values for tests with various levels of sensitivity and specificity were calculated for selected countries and can be found in Appendix 19. With a test having 70% sensitivity and 70% specificity, overall PPVs stayed below 2% in all countries. In countries with low GC prevalence (India, the U.S. and the United Kingdom), PPVs reached up to 1-2% only for 70-74 year old men and only with tests having at least 90% specificity. For 70-74 year old women, even a screening test with 90% sensitivity and 90% specificity would yield PPVs <0.4% in these countries. Overall PPVs higher than 5% would be expected in men in Korea and Japan for tests with 90% specificity, while PPVs in women for the same screening tests would remain below 3%.

4.5. Screening a population of 100,000 individuals

Expected results when using breath testing (sensitivity of 82.7%, specificity of 91.4%) to screen 100,000 individuals with various GC prevalence levels are presented in Table 21. In general, PPVs, i.e., expected prevalences of preclinical GC after prescreening by breath testing with a positive result, are consistently close to 10-fold higher than the corresponding prevalences in the total population without prescreening, with PPVs ranging from 0.1% to 13.5% compared to population-wide prevalences between 0.01% and 1.6%. The NNS to detect one GC are less than 100 in the population with initial GC prevalence of 0.1%. The expected number of total positives is hardly affected by GC prevalence, with around 9-10% of the total screened population having a positive result from breath tests. Among those with a negative test result (around 90% of the total population), >99.7% would truly be GC-free individuals.

Table 21. Screening with breath tests 100,000 individuals with various gastric cancer prevalence

Prevalence [%]	GC	HC	NNS[a]	TP	TN	FN	FP	Total positives		PPV[c] [%]	NPV[d] [%]	NNS[e]
								N[b]	%			
0.01	10	99,990	10,000	8	91,391	2	8,599	8,607	8.6	0.1	100	1,000
0.03	30	99,970	3,334	25	91,373	5	8,597	8,622	8.6	0.3	100	334
0.06	60	99,940	1,667	50	91,345	10	8,595	8,645	8.6	0.6	100	167
0.09	90	99,910	1,112	74	91,318	16	8,592	8,666	8.7	0.9	100	112
0.1	100	99,900	1,000	83	91,309	17	8,591	8,674	8.7	1.0	100	100
0.3	300	99,700	334	248	91,126	52	8,574	8,822	8.8	2.8	99.9	36
0.5	500	99,500	200	414	90,943	87	8,557	8,971	9.0	4.6	99.9	22
0.7	700	99,300	143	579	90,760	121	8,540	9,119	9.1	6.3	99.9	16
0.9	900	99,100	112	744	90,577	156	8,523	9,267	9.3	8.0	99.8	13
1.0	1,000	99,000	100	827	90,486	173	8,514	9,341	9.3	8.9	99.8	12
1.2	1,200	98,800	84	992	90,303	208	8,497	9,489	9.5	10.5	99.8	10
1.4	1,400	98,600	72	1,158	90,120	242	8,480	9,638	9.6	12.0	99.7	9
1.6	1,600	98,400	63	1,323	89,938	277	8,462	9,785	9.8	13.5	99.7	8

FN, false negative; FP, false positive; GC, gastric cancer; HC, healthy controls; N, number; NNS, number needed to screen; NPV, negative predictive value; PPV, positive predictive value; TN, true negative; TP, true positive.

[a] NNS = 1 / prevalence * 100. Number needed to screen to detect one gastric cancer case before preselection by breath testing.

[b] Total positives = TP + FP

[c] PPV = TP / (TP + FP) *100

[d] NPV = TN / (TN + FN) *100

[e] NNS = 1 / PPV * 100. Number needed to screen to detect one gastric cancer case after preselection by breath testing.

5. Discussion

In this chapter, the results of the systematic literature review on cancer detection with exhaled breath analysis, potentially important covariates to be considered in breath analysis and the potential of breath testing for cancer screening are discussed in the context of recent literature. Strengths and limitations of the research are discussed, followed by an outline of current challenges and the recommendations for further research.

5.1. Systematic review on breath analysis for cancer detection

Results of my systematic review present a comprehensive up-to-date overview of studies on diagnostic performance of VOCs in cancer detection, paying particular attention to breath sample collection, storage and analysis methods, whose standardization is a critical step toward the practical breath analysis application. My review identified 73 studies which used breath analysis for classification of cancer cases and controls or analyzed individual VOCs in exhaled breath of cancer patients and healthy individuals. The majority of the studies focused on lung cancer; however, recent reports addressed other common malignancies including breast, gastric and other types of cancer. In general, very good diagnostic performance of breath tests was reported. At the same time, it is notable that 25% studies lacked appropriate correction for overoptimism (i.e., meaning that reported results might have been achieved by the model overfitting the study population of that study), and studies exhibited large variation with respect to breath analysis techniques as well as data analysis methods. Based on the available published evidence, VOCs seem to hold a great potential in cancer diagnostics; nevertheless, the role of these markers for cancer screening needs to be examined and confirmed in large scale studies conducted in real life screening setting in diverse populations.

Despite the fact that it has been known for decades that hundreds of VOCs are present in human breath (131), the studies reporting diagnostic performance of breath testing for cancer detection accumulated only in the last decade. In the 1980s, the first studies reported higher levels of some volatiles in the breath of lung cancer patients (66, 126, 148), whereas nowadays breath analysis is employed to detect various types of cancer, including cancers of the ovaries (8), thyroid (68), prostate (133) and the colon (4, 6). This is mostly due to better understanding of the origin of individual VOCs and their connection to carcinogenesis, as well as due to establishment of new analysis techniques (mainly e-noses) allowing faster and cheaper

recruitment of patients and easier data handling compared to standard GC-MS analysis. Nevertheless, MS methods are still often used, e.g., among the studies included in the systematic review that were published in 2015, three used e-noses and two used MS techniques. According to the analysis method, studies on exhaled breath analysis can be grouped into two categories: studies using e-noses that aim to establish evidence on practical application of breath analysis for cancer detection, and studies with MS techniques that aim also to potentially find cancer-specific biomarkers in exhaled breath. Both of these directions are important, and the combination of these techniques is also possible, where specific VOCs identified by MS techniques can be used to select sensors in e-noses most sensitive to target compounds (133).

Regardless of the analysis technique, breath sample collection and storage are challenging tasks. Several different storage containers were used by the studies included in this systematic review (e.g., Tedlar bags, Bio-VOC samplers, Tenax sorption tubes) that are not equally suited for storage (in case analyses are not performed immediately after the collection). The stability of compounds in different bags has been investigated (23, 123, 124), which showed that some polar compounds, including water, diffuse rather quickly through Tedlar bag walls, while other compounds are quite stable. Aldehydes were shown to be rather stable in Bio-VOC sampler in the first 10 hours after collection (147). Sample storage times in studies included in this review were generally very short – the five studies which exceeded few months for storage, used thermal sorption tubes which were shown to be suitable for long-term storing (9). Apart from loss of compounds due to diffusion through the bags' walls, some compounds might be released by the bags themselves and accumulate in the collected air sample (120). The absence of compounds possibly lost due to long storage and the presence of artificial compounds arising from the storage container makes direct comparison of breath analysis results rather complicated, which calls for the need of standardized sample storing protocols to be established.

Reusing the same bag might represent another challenge as flushing and heating do not remove some of the compounds from Tedlar bag (117). Finally, concentration of VOCs also strongly depends on the breath collection method. Alveolar breath has higher levels of exhaled components than the whole breath without separation (121, 187) and also the lowest concentrations of contaminants (164). While the majority of the studies included in this

systematic review used alveolar breath, deep breath or normal breath were also analyzed. This might have resulted in different VOCs to be present in breath samples, as low concentrations VOCs might be detected only in alveolar breath.

Overall, standardization of the breath collection process appears crucial for further advances in breath-based biomarker research. Additionally to ambient air analysis or lung washout before breath sampling (93, 155) (which were done by 44 and 18 studies in this systematic review, respectively (8 studies did both)), other standardization processes including recommendations for sample storage in thermal desorption tubes or ways to avoid some covariates while recruiting hospital personnel rather than healthy individuals outside the hospital were recently suggested as well (9). However, the latter suggestion might increase the risk of other biases, such as selection bias.

A key issue in the analysis of high dimensional data such as those obtained from breath analysis is rigorous control for overoptimism by internal and/or external validation. Internal validation by performing, for example, random sample split or leave-one-out cross-validation can help to initially verify a classification model, but does not guarantee good performance once a different study population is targeted (84, 180). Therefore, external validation is a particularly important step in which the performance of a classification model is demonstrated on a population different from the one the model was calibrated with. Only 6 studies out of 73 reported performing external validation. Currently, the replication of the results in breath analysis remains challenging as different methods and analysis techniques are being used by different research groups. In fact, different results were achieved even in the same study while applying different computational approaches for data analysis (50, 74).

Identifying factors potentially influencing a classification model for breath analysis are another challenge, as little is known about covariates that should be taken into account. Both significant differences as well as non-significant associations between VOCs measured with MS techniques and age, gender and smoking status were reported (62, 97, 101, 141, 198), suggesting that these factors might be important to consider when building a classification model for cancer detection with data from MS techniques. On the other hand, results of "breath-prints" pattern analysis with e-nose showed to be insensitive to various covariates including the ones mentioned above (71, 133). Still, it will require further research to clarify factors that may influence study results using electronic nose. While matching or adjusting for

covariates is crucial for evaluating the discriminatory potential of VOCs *per se*, combined use of VOCs and covariates may provide the most powerful discriminatory algorithms for screening practice.

To date, despite of research ongoing for several decades, cancer-specific biomarker has not been discovered in exhaled breath, and the set of reported VOCs varies considerably among the studies. The same VOCs were reported for several cancer sites, such as formaldehyde (methanal) in studies on breast (55), prostate and bladder cancers (177), suggesting that breath analysis might be sensitive not to a specific cancer site, but rather cancer in general. Outstanding diagnostic performance has been reported for certain single markers in individual studies (e.g.., hexadecanal for lung cancer (209)); however, enhanced accuracy for classification of cancer cases and controls is likely to be achieved by the combination of several compounds. Potential metabolic pathways for volatiles to arise in the body and their transformations in a blood-breath pathway were addressed by Haick *et al*. (70), yet, further research is needed to fully understand the origin of VOCs and their connection to cancer.

A particular strength of this systematic review is that it includes all cancer sites and analysis methods in contrast to previous studies (31, 125, 152, 182) which were selective. Since the publication of the review, breath testing studies for cancer detection were published on CRC (6), GC or oesophageal cancer (167, 208) and lung cancer (64, 139, 159, 161). I extensively discuss key shortcomings of methodological issues, such as correction for overoptimism, performance of the validation studies and influence of potentially relevant covariates. Nevertheless, this review has certain limitations that need to be acknowledged. Despite a comprehensive research in two independent databases, I cannot rule out the possibility of having missed relevant studies. Standardized summarization by meta-analysis and presentation of results was hampered by heterogeneity in the reporting in the original studies.

In conclusion, breath analysis is a young field of research with great potential in cancer screening. In order to foster implementation in practice, larger studies should be implemented in true screening settings, paying particular attention to standardization in breath collection, consideration of covariates, adjustment for overoptimism, and validation in independent population samples. With further advancements in the area, breath tests may have the potential to become a useful supplement and improve existing screening tools for a variety of cancers.

5.2. Association analysis of VOCs in breath with various factors from the questionnaire data

Using exhaled breath analysis data from one of the biggest cohorts on cancer-free individuals conducted so far, I investigated whether socio-demographic factors, lifestyle factors, medical conditions and dietary patterns are associated with individual volatile compounds in exhaled breath. Univariate analysis revealed that all of 14 analyzed VOCs are associated with at least one of the analyzed factors. Statistically significant associations were found for gender and nationality, as well as 10 out 26 analyzed food items, including coffee, onion leaves and garlic. No significant associations were found between VOCs and vegetables or fruit consumption, and neither for smoking. Same volatiles were associated with several factors, such as VOC 6 with gender, nationality, consumption of coffee and onion leaves and presence of gastroduodenal ulcers. Further research is needed to confirm the results from my analysis in independent study populations. Further research is also needed on less common VOCs that were not analyzed in the current work.

So far, reported results of breath analysis for cancer detection were based on rather small sample size (number of participants in the studies varied from 14 to 477, see chapter 2), which rendered the investigation of the influence of covariates on breath test results virtually impossible. Studies on healthy, i.e., cancer-free, individuals are particularly suitable to investigate the role of covariates for breath analysis, as it is easier to achieve larger sample sizes. However, due to the batch effect that appeared in GC-MS analysis time in this study, I limited statistical analysis to the compounds that were present in all batches, i.e., common compounds. In total, 14 VOCs were included and 10 of them were present in more than 89% of the samples. Analysis of the common compounds prohibits finding significant associations for rare events, such as rare VOCs specific to the particular factor, e.g., underlying disease.

Recently, Blanchet *et al*. used GC-MS analysis to examined the profiles of VOCs in exhaled breath of 1,417 healthy individuals, and correlated these VOC profiles to 14 characteristics of all participants, including age, sex, BMI, smoking prevalence and various blood measurements (26). Statistically significant associations were demonstrated between VOC profiles and smoking, as well as age, BMI and gender. No associations between VOC profiles and standard blood measurements, such as blood cell counts, were found. While this large study aimed to

investigate potential covariates on breath analysis, results need to be confirmed in independent populations when using same analysis approaches.

A previous review on the impact of dietary factors on the composition of exhaled breath from 2013 concluded that diet has an impact on the composition of exhaled breath (2). General recommendations are to use standardized breath collection and analysis protocols in order to minimize any distortion of breath analysis results, including the ones that might occur due to recent food intake (e.g., breath sampling in the morning before breakfast is recommended). A study using repeated stool samples from individuals switching from normal to gluten-free diet demonstrated changes in the composition of the gut microbiome and alterations in the activity of microbial pathways (27). Differences corresponding to gluten-free diet could also be measured in significant changes in concentrations of 12 VOCs identified in exhaled breath (18). Gluten-free diet is primarily used to treat celiac disease; however, among 1,447 individuals included in my study, only one had celiac disease. Differences in diet in diverse populations are well-known, e.g. consumption of dairy products is less common in South-Eastern Asia, where more than 80% of populations are lactose intolerant compared to only 1% in the Netherlands, for example (122, 171). However, data on breath VOCs related to those diet regimes remain sparse. Overall, 2 individuals reported to follow gluten-free diet (not including the person with celiac disease, who indicated having normal diet), and very low frequencies of other common diets were reported – vegetarian (N=2), vegan (N=1) and lactose-free (N=0) – that made further statistical analysis regarding exhaled breath VOCs and major dietary differences as a whole impossible.

Composition of exhaled breath between diverse populations was investigated by Amal *et al.* (5), where individuals in two remote geographical locations (Latvia and China) were recruited with the primary goal of GC detection through exhaled breath analysis. Variation in VOC concentrations between Caucasian and Asian populations was observed, and authors reasoned that genetics and nutrition might be linked to these findings. The gut microbiome is closely connected to nutrition, and the presence of VOCs released by gut microbiota in exhaled breath is highly expected. A study on VOCs in exhaled breath and colonoscopy use investigated whether bowel cleansing has an effect on breath volatolome, and demonstrated that individuals before bowel cleansing can be separated from individuals after bowel cleansing (two repeated measurements of the same individuals) with 74% diagnostic accuracy (103).

The authors concluded that bowel cleansing has an effect on breath volatiles and therefore needs to be considered in breath testing.

My study is one of the largest studies to-date on exhaled breath samples from cancer-free individuals. However, despite of similarities between my analysis and the study by Blanchet *et al*. (26), including similar study population size (1,447 and 1,417 healthy individuals, respectively), as well as several overlapping analyzed factors (e.g., smoking, BMI), the direct comparison of the results is restrained due to differences in the used analysis approaches. In my analysis, individual VOCs were analyzed, whereas in the study by Blanchet *et al*. all VOCs together as profiles of breath samples were used. Furthermore, in my analysis I focused on common compounds and 71% of VOCs (10 out of 14) were present in 89-100% of study participants, while VOCs in at least of 20% of participants were included in the analysis by Blanchet *et al*.

My study has several limitations that need to be addressed. First, multiple GC-MS data preparation steps were performed to ensure the consistent list of VOCs between individual files, and several decisions of these data preparations steps were based on visual data inspection by one person, i.e., myself. However, all questions that arouse during the data preparation phase were thoroughly discussed and determined with the experts in the TECHNION group, who performed GC-MS analysis, as well with an expert from Maastricht University, who also participated in the study by Blanchet *et al* (26).

I performed univariate analysis to estimate associations between individual VOCs identified in exhaled breath and factors of interest determined from the questionnaire data. Univariate analysis was used to fit a two-part test that is particularly designed for data with many zeroes, such as mass spectrometry data (98). In MS, zeroes or missing data appear when a specific peak or VOC is not present in that sample (true “zero”) or due to technical reasons, such as peak concentration below detection limits. General solutions for dealing with zeroes in data are imputation, which works reasonably well if missing values are up to 20% (79, 194), or statistical tests which can handle missing data (two-part tests). Two-part models combine the information from a binomial test for the difference in the proportion of zero/non-zero values in a data and from a parametric or non-parametric test for the difference in means of the continuous components. Without any distributional assumptions for the continuous part, two-part Wilcoxon test, which was used in my analysis, shows good overall properties (65, 184).

Study participants provided self-reported information in the study questionnaires, which means that inaccurate recall or reporting may have been present, for example with items that carry negative attributes (e.g., heavy alcohol consumption might be underreported (176)). Nevertheless, recruitment of the participants in this study was conducted in the morning after fasting, and self-reports for the alcohol use are generally accurate if conducted under controlled conditions, e.g., collecting data from sober individuals (13). Also, questionnaire data collection was done together with breath sampling and calculated frequencies of yearly food item's consumption might not directly reflect the composition of analyzed compounds in exhaled breath in my analysis. For example, frequency of the consumption of seasonal food items (e.g., local fruits) might be underestimated.

Caution is needed interpreting the results on medication use. Available data allowed grouping study participants into two categories, i.e., whether individuals used certain medication over the last month or not. However, exact information was lacking on the date when the medication was taken and the period between medication usage and breath sampling, as well as the interval length of the medication use and the reason why the person took that treatment was unknown.

The usage of acid reduction drugs (including proton pump inhibitors) one month before sampling was indicated by 199 individuals. Proton pump inhibitors relieve symptoms of acid reflux or gastrooesophageal reflux disease. However, these diseases were not recorded in the questionnaire – only the presence of gastroduadenal ulcers was available, and it was indicated by 180 individuals. Data on these medications were collected to match the purpose of the GISTAR study (evaluate the changes in GC mortality due to *H. Pylori* eradication). However, high number of individuals (13.6%) reporting usage of acid reduction drugs suggests that the presence of some medical conditions was not recorded.

In summary, associations between volatile compounds in exhaled breath and socio-demographic factors, lifestyle factors, medical conditions and dietary patterns were investigated using data on 1,447 healthy individuals. Numerous analyzed associations were found to be significant, including associations of several VOCs with gender, nationality and consumption of certain food items, such as coffee and onion leaves. Taking together these results suggest that nutrition seems to be an important factor modifying breath composition. Further research is needed to determine whether individuals VOCs or composition of all

VOCs together can be truly attributed to any socio-demographic factors or lifestyle factors or dietary patterns.

5.3. Preselection of screening populations with breath tests

To my knowledge, the prevalence of preclinical CRC and GC by age and sex in various populations has not been estimated before. The same applies to PPVs, reflecting the prevalence of cancer of interest in the populations preselected with non-invasive testing. Substantial variation was demonstrated in the prevalence of preclinical CRC and GC between geographical regions and countries, and consistent variation was found between men and women and different age groups. As a consequence, PPVs for screening tests were found to vary substantially between countries, while NPVs were generally high (mostly >99%) in all populations. My results show that the prevalence of GC in the populations prescreened by breath testing is expected to be approximately 10-fold higher than prevalence before prescreening, which provides clear demonstration that preselection of the average risk population for further, more invasive screening, might be a promising approach to enhance effectiveness and cost-effectiveness of current screening programs.

Currently, no preselection criteria are used in GC screening programs, and CRC screening programs, while often offering prior testing with FITs, suffer from low participation rates. For example, a participation rate as low as 7% was reported from Belgium (91). Extensive research is going on regarding the development and evaluation of novel non-invasive screening tests, using a variety of different approaches, such as stool, blood, urine and breath sample-based testing, while addressing different molecular targets, such as genetic, epigenetic or proteomic markers (3, 11, 76, 204). Breath testing may be particularly suitable for mass screening due to its non-invasiveness, potentially low cost and easy sampling processes. However, there are uncertainties on potentially influencing factors for breath analysis. Prevalence of these factors is expected to vary between different populations. Nevertheless, studies using sensor array technologies demonstrated that differences in age, sex and smoking prevalence between study groups have no influence on the results (7, 71, 133, 202). Further validation of already published results in diverse populations together with more thoroughly exploring of potential influencing factors for breath analysis remain essential steps in establishing breath tests-based screening.

Caution is required when diagnostic performance of a screening test is derived from the published literature. The sensitivity and specificity of new screening tests are often estimated in case-control studies, conducted in clinical settings and comparing clinically manifest cases with healthy controls. Such studies may often overestimate sensitivity to be expected in a screening setting (48). While a case-control setting facilitates evaluation of test performance for rare conditions, such as specific cancers, clinically detected cancers often differ from screening-detected cancers in many respects in general. In particular, there is often a shift of the stage distribution towards more advanced cases that may result in overestimation of sensitivity compared to a true screening setting. For example, reported sensitivity of Epi proColon for detection of CRC ranged from 67 (196) to 96% (83) in case-control studies compared to only 48% (the estimate I used in my analyses) when the test was validated in an average-risk population (42).

Particular caution is warranted regarding the interpretation of PPVs reported from case-control studies. Such PPVs, if not adjusted for the actual prevalence in a given population, are often very high, sometimes even reaching 99-100% (185, 193). These seemingly high PPV levels are due to the high "prevalence" of the disease in the study population which results from the ratio of cases and controls determined by study design. As demonstrated in my results, even a CRC screening test with sensitivity and specificity of 95% would result in PPVs below 10% even in most populations with relatively high CRC prevalence. Caution is also required when PPVs are reported from studies oversampling older adults within the target population from the studies conducted in a screening setting, such as in a study on the Cologuard test (82), which may lead to overestimation of the PPV due to the higher prevalence of CRC at older ages. My results may therefore help to provide more realistic estimates of PPVs to be expected in population-wide screening than those reported in many diagnostic studies.

With the results provided here (e.g., Table 16), expected realistic PPVs and NPVs in different populations for any new test with known sensitivity and specificity can be easily estimated. For example, a blood test based on a 29-gene expression panel for CRC detection, with reported sensitivity of 79.5% and specificity of 90% (44), would have PPVs of around 4.1% in men and 2.7% in women and NPVs of 99.9% in both genders in the average-risk population of the U.S.

Several limitations have to be kept in mind when interpreting these results. Firstly, only limited data on transition rates from preclinical to clinical CRC and GC were available in the literature, and I applied the same annual transition rates for CRC as well as for GC in all populations. While transition rates might differ somewhat in distinct populations, reported transition rates for CRC development from few other countries, including the U.S. and France (99, 100, 149, 178, 181, 188), are in a similar range as those reported from Germany (32) (annual transition rates used in my analysis), suggesting that CRC development speed may not differ much between populations. The most precise available data on transition rates between preclinical to clinical GC were used (14), although these data were available only for men, and cancer development speed was assumed to be similar for women. While a transition rate between preclinical and clinical GC stages has been reported for women aged 50-59 (15), this estimate was rather imprecise due to the low number of GC cases, therefore these data were not considered in my study.

Furthermore, I estimated a point-prevalence of preclinical CRC and GC, i.e. corresponding to the year 2012. Changes in cancer prevalence and incidence over the time will affect the relevance of derived predictive values. CRC incidence remains stable or even decreases in highly developed countries, whereas incidence rates are rising rapidly in low- and middle-income countries due to populations' aging and adoption of Western lifestyle (12), suggesting that screening for CRC will remain an important part in controlling the cancer burden. On the contrary, incidence of GC worldwide is declining mainly due to decreasing levels of *H. Pylori* infection and further declines in GC incidence are expected in the future (24). As a result, the expected predictive values for GC screening tests will also decline over time with decreasing GC incidence. Utilization of highly sensitive and specific tools to preselect subpopulations for GC screening will become compulsory to make screening effective and the value of breath testing may strongly increase.

Implementation of new CRC screening programs in countries shortly before the year 2012 could theoretically have artificially increased numbers of "incident" cancer cases, resulting in apparently larger numbers of preclinical cancers and higher PPVs. This particularly applies for derived CRC prevalence estimates in countries where screening was implemented in 2012, e.g. Ireland, Norway and Malta (163). However, for this to have a substantial impact,

presumably a powerful screening tool (e.g. sigmoidoscopy or colonoscopy) and high uptake of screening would be required, which is not known to have occurred.

Screening tests are known to have different sensitivities for cancer subtypes. For example, FITs are more sensitive for the detection of left-side colon cancer compared to right-side colon cancer (73). While cancer subtype-specific prevalence estimates as well as cancer subtype-specific PPVs would be beneficial for health care planners, such estimates could not be derived given the lack of available data. However, variation between the proportions of colon and rectum cancers between men and women and different age groups have been described (69, 170), as well as variation between cardia and non-cardia GC prevalences (45), which may be an additional source of variation in predictive values for tests whose diagnostic performance varies for cancer subtypes. Performance of breath tests for small cell and non-small cell lung cancer were reported (46, 115); however, to my knowledge, no studies on exhaled breath analysis addressing CRC or GC-specific subtypes are available. Future research will be required to assess the extent to which population level variation in cancer subtypes may affect the potential performance of screening programs.

Other cancer type-specific limitations of my work include that I exclusively focused on the prevalence of preclinical cancer, while detection and removal of precancerous lesions, such as AA in CRC screening, is an important screening component (34, 128, 166). Removal of AA prevents their further development to CRC, and for this reason CRC screening generally targets not only cancers but all advanced neoplasms (CRC and AA). Data on performance of breath testing for AA detection are sparse, with only one study reporting 91% diagnostic accuracy for classifying CRC and AA, and 94% for classifying AA and healthy individuals (6), and PPVs for AA detection could not be estimated. Nevertheless, the expected PPVs for a combined endpoint of advanced neoplasms, i.e., either CRC or AA, would be much higher than PPVs for CRC only, given a much higher prevalence of advanced adenomas in screening populations (33, 156).

Furthermore, specific issues related to estimation of breath testing performance for GC detection should be discussed. I derived estimates of breath testing performance from 4 studies which used different analytical approaches to develop their GC detection models, including pattern recognition of different sensor array data, as well as identification of individual compounds in exhaled breath with analytical methods. While standardized procedures for

breath sampling, analysis and interpretation of the results are yet to be agreed upon, I, like others, used a general term “breath testing”. With the understanding that the target of all currently available breath analysis methods is differences in breath composition, this approach, and derivation of summary estimates of diagnostic performance appears justified.

Finally, more comprehensive modeling is required to estimate the overall effectiveness of a certain screening strategy in a particular population. Such analyses which can best be performed by microsimulation models need to take multiple additional factors into account, including time trends in cancer incidence, the composition of and changes in the underlying population’s age structure, and effects of the screening program modalities on the detection (and removal) of cancer precursors, as well as resulting cost implications. While such modeling is beyond the scope of the work presented here, the detailed estimates of prevalence of preclinical CRC and GC together with PPVs by sex and age should be helpful to inform such models.

In summary, I estimated prevalences of preclinical CRC and GC cases in the target population for screening in various regions of the world and found substantial geographic-, age- and gender-specific variation. I demonstrate that these variations profoundly affect the expected PPVs, and thereby the actual performance of screening tests. My analyses also provide valuable, up-to-date information on the potential of using breath testing as a non-invasive prescreening tool for preselecting average-risk people for CRC and GC detection with more invasive and costly screening procedures, such as colonoscopy and UEG, respectively. My results suggest that prescreening by breath testing might be a promising tool to achieve or enhance effectiveness of screening programs. However, further validation of breath testing results derived from diverse populations in screening setting is needed to fully determine the value and potential of breath testing for cancer screening**.** In addition, comprehensive modeling studies are required to evaluate effectiveness and cost-effectiveness of various conceivable approaches to population-based CRC and GC screening by breath testing in detail.

5.4. Conclusion

As of today, individual cancer-specific biomarkers in exhaled breath are not discovered yet. Cancer detection through exhaled breath analysis may only be possible based on systematic

and consistent differences in the “breath print” of cancer patients and cancer-free individuals, as opposed to usual screening tests that look for specific markers, such as blood in stool (FITs), for example.

Mass spectrometry has allowed the identification of several compounds already that are potentially related to cancer metabolism, but reported compounds are not unique for a specific cancer site or not even for cancer in general, suggesting that a combination of inflammatory, oxidative stress and other markers all together might contribute to a sensitive discriminatory model to distinguish between cancer patients and controls.

Standardization of breath sampling, storing and preprocessing of data will be essential to increase the quality of studies and derive reproducible and verifiable results. Further research is also required on the influence of covariates on breath analysis results, specifically, whether common factors that are present in general population (for instance, smoking) also affect the discriminatory value of a cancer detection model.

Overall, MS data are enormously complex, and sample collection, storing and analysis are time consuming and expensive, that makes exhaled breath analysis with MS methods not suited for the mass screening. In cancer screening, a large throughput sampling protocol of participants is required and “breath-print” pattern recognition by electronic noses is the method most likely to be used for practical applications. Well-defined and accepted breath sample handling and analysis protocols are also required for e-noses. Nevertheless, despite of diversity between the published studies in all these aspects, the results on reported diagnostic performance of breath testing for cancer detection are promising. E-noses, while being fast and easy to handle tools, are particularly suitable for mass screening. Screening asymptomatic average-risk population with breath testing and preselecting subpopulations for further invasive testing could make screening effective even for rare cancers.

In conclusion, with further improvements in the field, breath testing could be used to preselect high-risk populations for further invasive testing and substantially reduce the number of following invasive and costly procedures. Further research is needed to confirm breath testing diagnostic performance in diverse populations using standardized procedures for breath sampling, storing and analysis methods.

6. Summary

Cancer is a leading cause of death worldwide. Early detection is essential to improve successful treatment and reduce cancer mortality, and cancer screening in the asymptomatic general population might be a particularly promising approach to achieve this goal. However, direct screening of the asymptomatic population may be suboptimal due to potential harms arising from the invasiveness of the screening methods, high resource utilization and cost. There is a need for reliable non-invasive screening tools that can preselect high-risk populations for further invasive screening and that could achieve high levels of adherence at virtually no risk in population-based screening.

The aim of this dissertation is to provide further insight into exhaled breath analysis for cancer detection and to explore the potential of non-invasive breath testing for cancer screening.

First, a systematic literature review was performed to summarize the current state of exhaled breath analysis for cancer detection. Overall, 73 studies were identified that focused on detecting cancers through exhaled breath, including common malignancies, such as lung, breast and colorectal cancer (CRC), as well as more rare malignant diseases (e.g., malignant mesothelioma). Very good diagnostic performance of breath tests was demonstrated for all cancer types, but overoptimistic results could have been reported as one out of four studies did not employ any validation procedures. Furthermore, substantial differences were revealed in breath collection, storage and analysis methods between the studies. Further studies on exhaled breath analysis for cancer detection, particularly, the validation of already reported results, together with exploring factors that may have an influence on breath analysis results, are needed.

Second, potentially important covariates that can influence breath test results were investigated using data obtained by gas chromatography-mass spectrometry (the gold standard for analytical methods) analysis. Breath samples from more than 1,400 healthy individuals were used to explore whether socio-demographic (e.g., sex, nationality), lifestyle factors (e.g., smoking prevalence, diet) and various medical conditions (e.g., diabetes), can influence breath analysis results. Fourteen volatile organic compounds which were present in the breath samples in the majority of the patients were analyzed, and statistically significant differences were demonstrated for sex and nationality, as well as between people consuming certain food

products (e.g., coffee, fermented milk and onion leaves) at different frequencies. These results suggest, that differences in socio-demographic and lifestyle factors can be measured in exhaled breath and they might influence the results of breath testing for cancer detection as prevalence of such factors varies between populations.

Finally, using available data on breath tests diagnostic performance for CRC and gastric cancer (GC) detection, the potential of breath testing for cancer screening was examined. For that, the prevalence of preclinical CRC and GC (the target for the screening) was estimated and these data were used to derive positive predictive values (PPVs) of breath tests in various populations worldwide. My results suggest that restricting the screening to high-risk populations preselected by breath testing can substantially improve screening programs. For example, the number needed to screen to detect one GC in populations positive for breath testing is approximately 10-fold lower than that in populations without prescreening. Prescreening by breath testing might improve overall screening effectiveness through reducing the number of unnecessary invasive procedures and lowering overall cost, as well as improving participation rate that currently is not satisfactory. However, more data on breath testing for CRC and GC detection derived from diverse populations are needed to determine the full potential of breath testing in cancer screening.

In summary, breath analysis as a non-invasive, harmless and potentially cheap method holds great potential for cancer screening. Existing evidence on diagnostic performance for cancer detection suggests that breath testing can be a powerful tool to enhance cancer screening programs. While breath testing might not be suited to determine the final cancer diagnosis, the use of breath testing for the preselection of high-risk groups for further screening by more invasive methods could substantially improve effectiveness and cost-effectiveness of screening programs.

BIBLIOGRAPHY

1. Agkoc M, Dursun H, Albayrak F, Yilmaz O, Kiziltunc A, Yilmaz A and Gundogdu C (2010). Usefulness of serum pepsinogen levels as a screening test for atrophic gastritis and gastric cancer. Eurasian J Med. 42: 15-18. DOI: 10.5152/eajm.2010.05

2. Ajibola OA, Smith D, Spanel P and Ferns GA (2013). Effects of dietary nutrients on volatile breath metabolites. J Nutr Sci. 2: e34. DOI: 10.1017/jns.2013.26

3. Altomare DF (2016). Breath analysis for colorectal cancer screening. Colorectal Dis. 18: 1127-1128. DOI: 10.1111/codi.13480

4. Altomare DF, Di Lena M, Porcelli F, Trizio L, Travaglio E, Tutino M, Dragonieri S, Memeo V and de Gennaro G (2013). Exhaled volatile organic compounds identify patients with colorectal cancer. Br J Surg. 100: 144-150. DOI: 10.1002/bjs.8942

5. Amal H, Leja M, Broza YY, Tisch U, Funka K, Liepniece-Karele I, Skapars R, Xu ZQ, Liu H and Haick H (2013). Geographical variation in the exhaled volatile organic compounds. J Breath Res. 7: 047102. DOI: 10.1088/1752-7155/7/4/047102

6. Amal H, Leja M, Funka K, Lasina I, Skapars R, Sivins A, Ancans G, Kikuste I, Vanags A, Tolmanis I, Kirsners A, Kupcinskas L and Haick H (2016). Breath testing as potential colorectal cancer screening tool. Int J Cancer. 138: 229-236. DOI: 10.1002/ijc.29701

7. Amal H, Leja M, Funka K, Skapars R, Sivins A, Ancans G, Liepniece-Karele I, Kikuste I, Lasina I and Haick H (2016). Detection of precancerous gastric lesions and gastric cancer through exhaled breath. Gut. 65: 400-407. DOI: 10.1136/gutjnl-2014-308536

8. Amal H, Shi DY, Ionescu R, Zhang W, Hua QL, Pan YY, Tao L, Liu H and Haick H (2015). Assessment of ovarian cancer conditions from exhaled breath. Int J Cancer. 136: E614–E622. DOI: 10.1002/ijc.29166

9. Amann A, Miekisch W, Pleil JD, Risby TH and Schubert J (2010), Methodological issues of sample collection and analysis of exhaled breath. Eur Respir Soc Monogr. Maney Publishing, Leeds, Uk, p. 1-45.

10. Amann A, Mochalski P, Ruzsanyi V, Broza YY and Haick H (2014). Assessment of the exhalation kinetics of volatile cancer biomarkers based on their physicochemical properties. J Breath Res. 8: 016003. DOI: 10.1088/1752-7155/8/1/016003

11. Arasaradnam RP, McFarlane MJ, Ryan-Fisher C, Westenbrink E, Hodges P, Thomas MG, Chambers S, O'Connell N, Bailey C, Harmston C, Nwokolo CU, Bardhan KD and Covington JA (2014). Detection of colorectal cancer (CRC) by urinary volatile organic compound analysis. PLoS One. 9: e108750. DOI: 10.1371/journal.pone.0108750

12. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A and Bray F (2017). Global patterns and trends in colorectal cancer incidence and mortality. Gut. 66: 683-691. DOI: 10.1136/gutjnl-2015-310912

13. Babor TF, Steinberg K, Anton R and Del Boca F (2000). Talk is cheap: measuring drinking outcomes in clinical trials. J Stud Alcohol. 61: 55-63. DOI: 10.15288/jsa.2000.61.55

14. Bae JM, Shin SY and Kim EH (2014). Mean sojourn time of preclinical gastric cancer in Korean men: a retrospective observational study. J Prev Med Public Health. 47: 201-205. DOI: 10.3961/jpmph.2014.47.4.201

15. Bae JM, Shin SY and Kim EH (2015). Optimal Interval for Repeated Gastric Cancer Screening in Normal-Risk Healthy Korean Adults: A Retrospective Cohort Study. Cancer Res Treat. 47: 564-568. DOI: 10.4143/crt.2014.098

16. Bajtarevic A, Ager C, Pienz M, Klieber M, Schwarz K, Ligor M, Ligor T, Filipiak W, Denz H, Fiegl M, Hilbe W, Weiss W, Lukas P, Jamnig H, Hackl M, Haidenberger A, Buszewski B, Miekisch W, Schubert J and Amann A (2009). Noninvasive detection of lung cancer by analysis of exhaled breath. BMC cancer. 9: 348.

17. Bajtarevic A, Ager C, Pienz M, Klieber M, Schwarz K, Ligor M, Ligor T, Filipiak W, Denz H, Fiegl M, Hilbe W, Weiss W, Lukas P, Jamnig H, Hackl M, Haidenberger A, Buszewski B, Miekisch W, Schubert J and Amann A (2009). Noninvasive detection of lung cancer by analysis of exhaled breath. BMC cancer. 9: 348. DOI: 10.1186/1471-2407-9-348. DOI: 10.1186/1471-2407-9-348. DOI: 10.1186/1471-2407-9-348

18. Baranska A, Tigchelaar E, Smolinska A, Dallinga JW, Moonen EJ, Dekens JA, Wijmenga C, Zhernakova A and van Schooten FJ (2013). Profile of volatile organic compounds in exhaled breath changes as a result of gluten-free diet. J Breath Res. 7: 037104. DOI: 10.1088/1752-7155/7/3/037104

19. Barash-Adams O (2016). [Personal communication].

20. Barker M, Hengst M, Schmid J, Buers HJ, Mittermaier B, Klemp D and Koppmann R (2006). Volatile organic compounds in the exhaled breath of young patients with cystic fibrosis. Eur Respir J. 27: 929-936. DOI: 10.1183/09031936.06.00085105

21. Barzi A, Lenz HJ, Quinn DI and Sadeghi S (2017). Comparative effectiveness of screening strategies for colorectal cancer. Cancer. 123: 1516-1527. DOI: 10.1002/cncr.30518

22. Basch CH, C.E. B, Zybert P and R.L. W (2016). Fear as a Barrier to Asymptomatic Colonoscopy Screening in an Urban Minority Population with Health Insurance. J Community Health. 41: 818-824. DOI: 10.1007/s10900-016-0159-9

23. Beauchamp J, Herbig J, Gutmann R and Hansel A (2008). On the use of Tedlar (R) bags for breath-gas sampling and analysis. J Breath Res. 2: 046001. DOI: 10.1088/1752-7155/2/4/046001

24. Bertuccio P, Chatenoud L, Levi F, Praud D, Ferlay J, Negri E, Malvezzi M and La Vecchia C (2009). Recent patterns in gastric cancer: a global overview. Int J Cancer. 125: 666-673. DOI: 10.1002/ijc.24290

25. Bessonneau V, Mosqueron L, Berrube A, Mukensturm G, Buffet-Bataillon S, Gangneux JP and Thomas O (2013). VOC contamination in hospital, from stationary sampling of a large panel of compounds, in view of healthcare workers and patients exposure assessment. PLoS One. 8: e55535. DOI: 10.1371/journal.pone.0055535

26. Blanchet L, Smolinska A, Baranska A, Tigchelaar E, Swertz M, Zhernakova A, Dallinga JW, Wijmenga C and van Schooten FJ (2017). Factors that influence the volatile organic compound content in human breath. J Breath Res. 11: 016013. DOI: 10.1088/1752-7163/aa5cc5

27. Bonder MJ, Tigchelaar EF, Cai X, Trynka G, Cenit MC, Hrdlickova B, Zhong H, Vatanen T, Gevers D, Wijmenga C, Wang Y and Zhernakova A (2016). The influence of a short-term gluten-free diet on the human gut microbiome. Genome Med. 8: 45. DOI: 10.1186/s13073-016-0295-y

28. Bonferroni CE (1935), Il calcolo delle assicurazioni su gruppi di teste, p. 13-60. In: Ortu Carboni S (ed.): *Studi in Onore del Professore Salvatore Ortu Carboni*. Giappichelli, Rome.

29. Borgatti S (2010). Normalizing variables. Available from: http://www.analytictech.com/ba762/handouts/normalization.htm. [last accessed: 2 June 2017].

30. Bousamra Mn, Schumer E, Li M, Knipp RJ, Nantz MH, van Berkel V and Fu XA (2014). Quantitative analysis of exhaled carbonyl compounds distinguishes benign from malignant pulmonary disease. J Thorac Cardiovasc Surg. 148: 1074-1080; discussion 1080-1081. DOI: 10.1016/j.jtcvs.2014.06.006

31. Bouza M, Gonzalez-Soto J, Pereiro R, de Vicente JC and Sanz-Medel A (2017). Exhaled breath and oral cavity VOCs as potential biomarkers in oral cancer patients. J Breath Res. 11: 016015. DOI: 10.1088/1752-7163/aa5e76

32. Brenner H, Altenhofen L, Katalinic A, Lansdorp-Vogelaar I and Hoffmeister M (2011). Sojourn time of preclinical colorectal cancer by sex and age: estimates from the German national screening colonoscopy database. Am J Epidemiol. 174: 1140-1146. DOI: 10.1093/aje/kwr188

33. Brenner H, Altenhofen L, Kretschmann J, Rosch T, Pox C, Stock C and Hoffmeister M (2015). Trends in Adenoma Detection Rates During the First 10 Years of the German Screening Colonoscopy Program. Gastroenterology. 149: 356-366 e1. DOI: 10.1053/j.gastro.2015.04.012

34. Brenner H, Stock C and Hoffmeister M (2014). Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: systematic review and meta-analysis of randomised controlled trials and observational studies. BMJ. 348: g2467. DOI: 10.1136/bmj.g2467

35. Brooks SW, Moore DR, Marzouk EB, Glenn FR and Hallock RM (2015). Canine olfaction and electronic nose detection of volatile organic compounds in the detection of cancer: a review. Cancer Invest. 33: 411-419. DOI: 10.3109/07357907.2015.1047510

36. Broza YY, Kremer R, Tisch U, Gevorkyan A, Shiban A, Best LA and Haick H (2013). A nanomaterial-based breath test for short-term follow-up after lung tumor resection. Nanomedicine. 9: 15-21. DOI: 10.1016/j.nano.2012.07.009

37. Buck L and Axel R (1991). A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell. 65: 175-187. DOI: 10.1016/0092-8674(91)90418-X

38. Buszewski B, Ligor T, Jezierski T, Wenda-Piesik A, Walczak M and Rudnicka J (2012). Identification of volatile lung cancer markers by gas chromatography-mass spectrometry: comparison with discrimination by canines. Anal Bioanal Chem. 404: 141-146. DOI: 10.1007/s00216-012-6102-8

39. Chapman EA, Thomas PS, Stone E, Lewis C and Yates DH (2012). A breath test for malignant mesothelioma using an electronic nose. Eur Respir J. 40: 448-454. DOI: 10.1183/09031936.00040911

40. Chen X, Cao MF, Li Y, Hu WJ, Wang P, Ying KJ and Pan HM (2005). A study of an electronic nose for detection of lung cancer based on a virtual SAW gas sensors array and imaging recognition method. Meas Sci Technol. 16: 1535-1546. DOI: 10.1088/0957-0233/16/8/001

41. Choi KS and Suh M (2014). Screening for gastric cancer: the usefulness of endoscopy. Clin Endosc. 47: 490-496. DOI: 10.5946/ce.2014.47.6.490

42. Church TR, Wandell M, Lofton-Day C, Mongin SJ, Burger M, Payne SR, Castanos-Velez E, Blumenstein BA, Rosch T, Osborn N, Snover D, Day RW and Ransohoff DF (2014). Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. Gut. 63: 317-25. DOI: 10.1136/gutjnl-2012-304149

43. Church TR, Wandell M, Lofton-Day C, Mongin SJ, Burger M, Payne SR, Castanos-Velez E, Blumenstein BA, Rosch T, Osborn N, Snover D, Day RW, Ransohoff DF, Presept Clinical Study Steering Committee I and Study T (2014). Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. Gut. 63: 317-325. DOI: 10.1136/gutjnl-2012-304149

44. Ciarloni L, Ehrensberger SH, Imaizumi N, Monnier-Benoit S, Nichita C, Myung SJ, Kim JS, Song SY, Kim TI, van der Weg B, Meier R, Borovicka J, Beglinger C, Vallet C, Maerten P, Ruegg C and Dorta G (2016). Development and Clinical Validation of a Blood Test Based on 29-Gene Expression for Early Detection of Colorectal Cancer. Clin Cancer Res. 22: 4604-4611. DOI: 10.1158/1078-0432.CCR-15-2057

45. Colquhoun A, Arnold M, Ferlay J, Goodman KJ, Forman D and Soerjomataram I (2015). Global patterns of cardia and non-cardia gastric cancer incidence in 2012. Gut. 64: 1881-1888. DOI: 10.1136/gutjnl-2014-308915

46. Corradi M, Poli D, Banda I, Bonini S, Mozzoni P, Pinelli S, Alinovi R, Andreoli R, Ampollini L, Casalini A, Carbognani P, Goldoni M and Mutti A (2015). Exhaled breath analysis in suspected cases of non-small-cell lung cancer: a cross-sectional study. J Breath Res. 9: 027101. DOI: 10.1088/1752-7155/9/2/027101

47. Correa P (1992). Human gastric carcinogenesis: a multistep and multifactorial process--First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. Cancer Res. 52: 6735-6740. DOI: Published December 1992

48. Cronin KA, Weed DL, Connor RJ and Prorok PC (1998). Case-control studies of cancer screening: theory and practice. J Natl Cancer Inst. 90: 498-504. DOI: 10.1093/jnci/90.7.498

49. D'Amico A, Pennazza G, Santonico M, Martinelli E, Roscioni C, Galluccio G, Paolesse R and Di Natale C (2010). An investigation on electronic nose diagnosis of lung cancer. Lung Cancer. 68: 170-176. DOI: 10.1016/j.lungcan.2009.11.003

50. de Gennaro G, Dragonieri S, Longobardi F, Musti M, Stallone G, Trizio L and Tutino M (2010). Chemical characterization of exhaled breath to differentiate between patients with malignant plueral mesothelioma from subjects with similar professional asbestos exposure. Anal Bioanal Chem. 398: 3043-3050. DOI: 10.1007/s00216-010-4238-y

51. Di Natale C, Macagnano A, Martinelli E, Paolesse R, D'Arcangelo G, Roscioni C, Finazzi-Agro A and D'Amico A (2003). Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors. Biosens Bioelectron. 18: 1209-1218. DOI: 10.1016/S0956-5663(03)00086-1

52. Doebler P. Package ‘mada’. Available from: http://www.R-project.org. [last accessed: 1 May 2017].

53. Dragonieri S, Annema JT, Schot R, van der Schee MP, Spanevello A, Carratu P, Resta O, Rabe KF and Sterk PJ (2009). An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD. Lung Cancer. 64: 166-170. DOI: 10.1016/j.lungcan.2008.08.008

54. Dragonieri S, van der Schee MP, Massaro T, Schiavulli N, Brinkman P, Pinca A, Carratu P, Spanevello A, Resta O, Musti M and Sterk PJ (2012). An electronic nose distinguishes exhaled breath of patients with Malignant Pleural Mesothelioma from controls. Lung Cancer. 75: 326-331. DOI: 10.1016/j.lungcan.2011.08.009

55. Ebeler SE, Clifford AJ and Shibamoto T (1997). Quantitative analysis by gas chromatography of volatile carbonyl compounds in expired air from mice and human. J Chromatogr B Biomed Sci Appl. 702: 211-215. DOI: 10.1016/S0378-4347(97)00369-1

56. Eilers PHC and Boelens HFM (2005). Baseline Correction with Asymmetric Least Squares Smoothing, p. 1-24. In: Leiden University Medical Centre Report.

57. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D and Bray F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality

Worldwide: IARC CancerBase No. 11. Available from: http://globocan.iarc.fr/. [last accessed: 1 August 2016].

58. Filipiak W, Filipiak A, Sponring A, Schmid T, Zelger B, Ager C, Klodzinska E, Denz H, Pizzini A, Lucciarini P, Jamnig H, Troppmair J and Amann A (2014). Comparative analyses of volatile organic compounds (VOCs) from patients, tumors and transformed cell lines for the validation of lung cancer-derived breath markers. J Breath Res. 8: 027111. DOI: 10.1088/1752-7155/8/2/027111

59. Fock KM and Ang TL (2010). Epidemiology of Helicobacter pylori infection and gastric cancer in Asia. J Gastroenterol Hepatol. 25: 479-486. DOI: 10.1111/j.1440-1746.2009.06188.x

60. Fock KM, Katelaris P, Sugano K, Ang TL, Hunt R, Talley NJ, Lam SK, Xiao SD, Tan HJ, Wu CY, Jung HC, Hoang BH, Kachintorn U, Goh KL, Chiba T and Rani AA (2009). Second Asia-Pacific Consensus Guidelines for Helicobacter pylori infection. J Gastroenterol Hepatol. 24: 1587-1600. DOI: 10.1111/j.1440-1746.2009.05982.x

61. Fu XA, Li M, Knipp RJ, Nantz MH and Bousamra M (2014). Noninvasive detection of lung cancer using exhaled breath. Cancer Med. 3: 174-181. DOI: 10.1002/cam4.162

62. Fuchs P, Loeseken C, Schubert JK and Miekisch W (2010). Breath gas aldehydes as biomarkers of lung cancer. Int J Cancer. 126: 2663-2670. DOI: 10.1002/ijc.24970

63. Garcia RA, Morales V, Martin S, Vilches E and Toledano A (2014). Volatile Organic Compounds Analysis in Breath Air in Healthy Volunteers and Patients Suffering Epidermoid Laryngeal Carcinomas. Chromatographia. 77: 501-509. DOI: 10.1007/s10337-013-2611-7

64. Gasparri R, Santonico M, Valentini C, Sedda G, Borri A, Petrella F, Maisonneuve P, Pennazza G, D'Amico A, Di Natale C, Paolesse R and Spaggiari L (2016). Volatile signature for the early diagnosis of lung cancer. J Breath Res. 10: 016007. DOI: 10.1088/1752-7155/10/1/016007

65. Gleiss A, Dakna M, Mischak H and Heinze G (2015). Two-group comparisons of zero-inflated intensity values: the choice of test statistic matters. Bioinformatics. 31: 2310-2317. DOI: 10.1093/bioinformatics/btv154

66. Gordon SM, Szidon JP, Krotoszynski BK, Gibbons RD and O'Neill HJ (1985). Volatile organic compounds in exhaled air from patients with lung cancer. Clin Chem. 31: 1278-1282. Retrieved from http://clinchem.aaccjnls.org/

67. Gruber M, Tisch U, Jeries R, Amal H, Hakim M, Ronen O, Marshak T, Zimmerman D, Israel O, Amiga E, Doweck I and Haick H (2014). Analysis of exhaled breath for diagnosing head and neck squamous cell carcinoma: a feasibility study. Br J Cancer. 111: 790-798. DOI: 10.1038/bjc.2014.361

68. Guo L, Wang C, Chi C, Wang X, Liu S, Zhao W, Ke C, Xu G and Li E (2015). Exhaled breath volatile biomarker analysis for thyroid cancer. Transl. Res. 188-895. DOI: 10.1016/j.trsl.2015.01.005

69. Haggar FA and Boushey RP (2009). Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. Clin Colon Rectal Surg. 22: 191-197. DOI: 10.1055/s-0029-1242458

70. Haick H, Broza YY, Mochalski P, Ruzsanyi V and Amann A (2014). Assessment, origin, and implementation of breath volatile cancer markers. Chem Soc Rev. 43: 1423-1449. DOI: 10.1039/c3cs60329f

71. Hakim M, Billan S, Tisch U, Peng G, Dvrokind I, Marom O, Abdah-Bortnyak R, Kuten A and Haick H (2011). Diagnosis of head-and-neck cancer from exhaled breath. Br J Cancer. 104: 1649-1655. DOI: 10.1038/bjc.2011.128

72. Handa H, Usuba A, Maddula S, Baumbach JI, Mineshita M and Miyazawa T (2014). Exhaled breath analysis for lung cancer detection using ion mobility spectrometry. PLoS One. 9: e114555. DOI: 10.1371/journal.pone.0114555

73. Haug U, Knudsen AB, Brenner H and Kuntz KM (2011). Is fecal occult blood testing more sensitive for left- versus right-sided colorectal neoplasia? A systematic literature review. Expert Rev Mol Diagn. 11: 605-616. DOI: 10.1586/erm.11.41

74. Hauschild AC, Baumbach JI and Baumbach J (2012). Integrated statistical learning of metabolic ion mobility spectrometry profiles for pulmonary disease identification. Genet Mol Res. 11: 2733-2744. DOI: 10.4238/2012.July.10.17

75. Hawkins DM (2004). The problem of overfitting. J Chem Inf Comput Sci. 44: 1-12. DOI: 10.1021/ci0342472

76. Heichman KA (2014). Blood-based testing for colorectal cancer screening. Mol Diagn Ther. 18: 127-135. DOI: 10.1007/s40291-013-0074-z

77. Hietanen E, Bartsch H, Bereziat JC, Camus AM, McClinton S, Eremin O, Davidson L and Boyle P (1994). Diet and oxidative stress in breast, colon and prostate cancer patients: a case-control study. Eur J Nutr. 48: 575-586. Retrieved from http://www.nature.com/ejcn

78. Hisamichi S and Sugawara N (1984). Mass screening for gastric cancer by X-ray examination. Jpn J Clin Oncol. 14: 211-223. DOI: 10.1093/oxfordjournals.jjco.a038974

79. Hrydziuszko O and Viant MR (2012). Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline. Metabolomics. 8: S161-S174. DOI: 10.1007/s11306-011-0366-4

80. Huang YK, Yu JC, Kang WM, Ma ZQ, Ye X, Tian SB and Yan C (2015). Significance of Serum Pepsinogens as a Biomarker for Gastric Cancer and Atrophic Gastritis Screening: A Systematic Review and Meta-Analysis. PLoS One. 10: e0142080. DOI: 10.1371/journal.pone.0142080

81. Hubers AJ, Brinkman P, Boksem RJ, Rhodius RJ, Witte BI, Zwinderman AH, Heideman DAM, Duin S, Koning R, Steenbergen RDM, Snijders PJF, Smit EF, Sterk PJ and Thunnissen E (2014). Combined sputum hypermethylation and eNose analysis for lung cancer diagnosis. J Clin Pathol. 67: 707-711. DOI: 10.1136/jclinpath-2014-202414

82. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, Ahlquist DA and Berger BM (2014). Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med. 370: 1287-1297. DOI: 10.1056/NEJMoa1311194

83. Jin P, Kang Q, Wang X, Yang L, Yu Y, Li N, He YQ, Han X, Hang J, Zhang J, Song L, Han Y and Sheng JQ (2015). Performance of a second-generation methylated SEPT9 test in detecting colorectal neoplasm. J Gastroenterol Hepatol. 30: 830-833. DOI: 10.1111/jgh.12855

84. Justice AC, Covinsky KE and Berlin JA (1999). Assessing the generalizability of prognostic information. Ann Intern Med. 130: 515-524. DOI: 10.7326/0003-4819-130-6-199903160-00016

85. Karim-Kos HE, de Vries E, Soerjomataram I, Lemmens V, Siesling S and Coebergh JWW (2008). Recent trends of cancer in Europe: A combined approach of incidence, survival and mortality for 17 cancer sites since the 1990S. Eur J Cancer. 44: 1345-1389. DOI: 10.1016/j.ejca.2007.12.015

86. Kawai T, Kawakami K, Kataoka M, Itoi T, Takei K, Moriyasu F, Takagi Y, Aoki T, Serizawa H, Rimbard E, Noguchi N and Sasatsu M (2008). A study of the relationship between Helicobacter pylori microbial susceptibility, 13C-urea breath test values. Hepatogastroenterology. 55: 786-790. Retrieved from http://www.thieme.com

87. Kerlin P and Wong L (1988). Breath hydrogen testing in bacterial overgrowth of the small intestine. Gastroenterology. 95: 982-988. Retrieved from http://www.gastrojournal.org

88. Kharitonov SA, Yates D, Robbins RA, Logan-Sinclair R, Shinebourne EA and Barnes PJ (1994). Increased nitric oxide in exhaled air of asthmatic patients. Lancet. 343: 133-135. DOI: 10.1016/S0140-6736(94)90931-8

89. Kischkel S, Miekisch W, Sawacki A, Straker EM, Trefz P, Amann A and Schubert JK (2010). Breath biomarkers for lung cancer detection and assessment of smoking related effects - confounding variables, influence of normalization and statistical algorithms. Clin Chim Acta. 411: 1637-1644. DOI: 10.1016/j.cca.2010.06.005

90. Kivi M and Tindberg Y (2006). Helicobacter pylori occurrence and transmission: a family affair? Scand J Infect Dis. 38: 407-417. DOI: 10.1080/00365540600585131

91. Klabunde C, Blom J, Bulliard JL, Garcia M, Hagoel L, Mai V, Patnick J, Rozjabek H, Senore C and Tornberg S (2015). Participation rates for organized colorectal cancer screening programmes: an international comparison. J Med Screen. 22: 119-126. DOI: 10.1177/0969141315584694

92. Klesney SP and Rigaudy J (1979), Nomenclature of Organic Chemistry, Sections A, B, C, D, E, F, and H. Pergamon Press, Oxford , New York, Beijing.

93. Kneepkens CMF, Lepage G and Roy CC (1994). The Potential of the Hydrocarbon Breath Test as a Measure of Lipid-Peroxidation. Free Radical Bio Med. 17: 127-160. DOI: 10.1016/0891-5849(94)90110-4

94. Krilaviciute A, Heiss JA, Leja M, Kupcinskas J, Haick H and Brenner H (2015). Detection of cancer through exhaled breath: a systematic review. Oncotarget. 6: 38643-38657. DOI: 10.18632/oncotarget.5938

95. Krilaviciute A, Stock C and Brenner H (2017). International variation in the prevalence of preclinical colorectal cancer: implications for predictive values of noninvasive screening tests and potential target populations for screening. Int J Cancer. DOI: 10.1002/ijc.30867. [Accepted for publication]

96. Kumar S, Huang J, Abbassi-Ghadi N, Mackenzie HA, Veselkov KA, Hoare JM, Lovat LB, Spanel P, Smith D and Hanna GB (2015). Mass Spectrometric Analysis of Exhaled Breath for the Identification of Volatile Organic Compound Biomarkers in Esophageal and Gastric Adenocarcinoma. Ann Surg. 262: 981-990. DOI: 10.1097/sla.0000000000001101

97. Kushch I, Arendacka B, Stolc S, Mochalski P, Filipiak W, Schwarz K, Schwentner L, Schmid A, Dzien A, Lechleitner M, Witkovsky V, Miekisch W, Schubert J, Unterkofler K and Amann A (2008). Breath isoprene - aspects of normal physiology related to age, gender and cholesterol profile as determined in a proton transfer reaction mass spectrometry study. Clin Chem Lab Med. 46: 1011-1018. DOI: 10.1515/CCLM.2008.181

98. Lachenbruch PA (1976). Analysis of data with clumping at zero. Biom Z. 18: 351–356. Retrieved from http://www.dkfz.de/en/zbib/

99. Lansdorp-Vogelaar I, van Ballegooijen M, Boer R, Zauber A and Habbema JD (2009). A novel hypothesis on the sensitivity of the fecal occult blood test: Results of a joint analysis of 3 randomized controlled trials. Cancer. 115: 2410-2419. DOI: 10.1002/cncr.24256

100. Launoy G, Smith TC, Duffy SW and Bouvier V (1997). Colorectal cancer mass-screening: estimation of faecal occult blood test sensitivity, taking into account cancer mean sojourn time. Int J Cancer. 73: 220-224. DOI: 10.1002/(SICI)1097-0215(19971009)73:2<220::AID-IJC10>3.0.CO;2-J

101. Lechner M, Moser B, Niederseer D, Karlseder A, Holzknecht B, Fuchs M, Colvin S, Tilg H and Rieder J (2006). Gender and age specific differences in exhaled isoprene levels. Respir Physiol Neurobiol. 154: 478-483. DOI: 10.1016/j.resp.2006.01.007

102. Lee JK, Liles EG, Bent S, Levin TR and Corley DA (2014). Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. Ann Intern Med. 160: 171. DOI: 10.7326/M13-1484

103. Leja M, Amal H, Lasina I, Skapars R, Sivins A, Ancans G, Tolmanis I, Vanags A, Kupcinskas J, Ramonaite R, Khatib S, Bdarneh S, Natour R, Ashkar A and Haick H (2016). Analysis of the effects of microbiome-related confounding factors on the reproducibility of the volatolomic test. J Breath Res. 10: 037101. DOI: 10.1088/1752-7155/10/3/037101

104. Leung WK, Wu MS, Kakugawa Y, Kim JJ, Yeoh KG, Goh KL, Wu KC, Wu DC, Sollano J, Kachintorn U, Gotoda T, Lin JT, You WC, Ng EK, Sung JJ and Asia Pacific Working Group on Gastric C (2008). Screening for gastric cancer in Asia: current evidence and practice. Lancet Oncol. 9: 279-287. DOI: 10.1016/S1470-2045(08)70072-X

105. Leunis N, Boumans ML, Kremer B, Din S, Stobberingh E, Kessels AG and Kross KW (2014). Application of an electronic nose in the diagnosis of head and neck cancer. Laryngoscope. 124: 1377-1381. DOI: 10.1002/lary.24463

106. Levy I and Gralnek IM (2016). Complications of diagnostic colonoscopy, upper endoscopy, and enteroscopy. Best Pract Res Clin Gastroenterol. 30: 705-718. DOI: 10.1016/j.bpg.2016.09.005

107. Li J, Peng Y, Liu Y, Li W, Jin Y, Tang Z and Duan Y (2014). Investigation of potential breath biomarkers for the early diagnosis of breast cancer using gas chromatography-mass spectrometry. Clin Chim Acta. 436: 59-67. DOI: 10.1016/j.cca.2014.04.030

108. Ligor M, Ligor T, Bajtarevic A, Ager C, Pienz M, Klieber M, Denz H, Fiegl M, Hilbe W, Weiss W, Lukas P, Jamnig H, Hackl M, Buszewski B, Miekisch W, Schubert J and Amann A (2009). Determination of volatile organic compounds in exhaled breath of patients with lung cancer using solid phase microextraction and gas chromatography mass spectrometry. Clin Chem Lab Med. 47: 550-560. DOI: 10.1515/CCLM.2009.133

109. LNBD. Nanomaterial-based devices. Available from: http://lnbd.technion.ac.il/research/application-of-the-na-nose-diffrential-diagnosis-of-cancer/. [last accessed: 21 Jannuary 2015].

110. Ma H, Li X, Chen J, Wang H, Cheng T, Chen K and Xu S (2014). Analysis of human breath samples of lung cancer patients and healthy controls with solid-phase microextraction (SPME) and flow-modulated comprehensive two-dimensional gas chromatography (GC x GC). Anal Methods. 6: 6841-6849. DOI: 10.1039/C4AY01220H

111. Machado RF, Laskowski D, Deffenderfer O, Burch T, Zheng S, Mazzone PJ, Mekhail T, Jennings C, Stoller JK, Pyle J, Duncan J, Dweik RA and Erzurum SC (2005). Detection of lung cancer by sensor array analyses of exhaled breath. Am J Respir Crit Care Med. 171: 1286-1291. DOI: 10.1164/rccm.200409-1184OC

112. Mangler M, Freitag C, Lanowska M, Staeck O, Schneider A and Speiser D (2012). Volatile organic compounds (VOCs) in exhaled breath of patients with breast cancer in a clinical setting. Ginekol Pol. 83: 730-736. Retrieved from http://www.dkfz.de/en/zbib/

113. Mazzone PJ (2010), Exhaled volatile organic compounds as biomarkers for respiratory diseases, p. 130-139. In: Horvath I and de Jongste JE (ed.): *Exhaled Biomarkers*. European Respiratory Society.

114. Mazzone PJ, Hammel J, Dweik R, Na J, Czich C, Laskowski D and Mekhail T (2007). Diagnosis of lung cancer by the analysis of exhaled breath with a colorimetric sensor array. Thorax. 62: 565-568. DOI: 10.1136/thx.2006.072892

115. Mazzone PJ, Wang XF, Xu Y, Mekhail T, Beukemann MC, Na J, Kemling JW, Suslick KS and Sasidhar M (2012). Exhaled breath analysis with a colorimetric sensor array for the identification and characterization of lung cancer. J Thorac Oncol. 7: 137-142. DOI: 10.1097/JTO.0b013e318233d80f

116. McAlpine MC, Agnew HD, Rohde RD, Blanco M, Ahmad H, Stuparu AD, Goddard WA, 3rd and Heath JR (2008). Peptide-nanowire hybrid materials for selective sensing of small molecules. J Am Chem Soc. 130: 9583-9589. DOI: 10.1021/ja802506d

117. McGarvey LJ and Shorten CV (2000). The effects of adsorption on the reusability of Tedlar air sampling bags. AIHA J. 61: 375-380. DOI: 10.1080/15298660008984546

118. McWilliams A, Beigi P, Srinidhi A, Lam S and MacAulay C (2015). Sex and Smoking Status Effects on the Early Detection of Early Lung Cancer in High-Risk Smokers using an Electronic Nose. IEEE Trans Biomed Eng. 2044-2054. DOI: 10.1109/tbme.2015.2409092

119. Merkes International. Bio-VOC breath sampler. Available from: https://www.markes.com/Products/Sampling-accessories/Sampling/BIO-VOC-breath-sampler.aspx. [last accessed: 1 June 2017].

120. Miekisch W, Kischkel S, Sawacki A, Liebau T, Mieth M and Schubert JK (2008). Impact of sampling procedures on the results of breath analysis. J Breath Res. 2: 026007. DOI: 10.1088/1752-7155/2/2/026007

121. Miekisch W, Schubert JK and Noeldge-Schomburg GF (2004). Diagnostic potential of breath analysis--focus on volatile organic compounds. Clin Chim Acta. 347: 25-39. DOI: 10.1016/j.cccn.2004.04.023

122. Misselwitz B, Pohl D, Fruhauf H, Fried M, Vavricka SR and Fox M (2013). Lactose malabsorption and intolerance: pathogenesis, diagnosis and treatment. United European Gastroenterol J. 1: 151-159. DOI: 10.1177/2050640613484463

123. Mochalski P, King J, Unterkofler K and Amann A (2013). Stability of selected volatile breath constituents in Tedlar, Kynar and Flexfilm sampling bags. Analyst. 138: 1405-1418. DOI: 10.1039/C2AN36193K

124. Mochalski P, Wzorek B, Sliwka I and Amann A (2009). Suitability of different polymer bags for storage of volatile sulphur compounds relevant to breath analysis. J Chromatogr B Analyt Technol Biomed Life Sci. 877: 189-196. DOI: 10.1016/j.jchromb.2008.12.003

125. Nardi-Agmon I and Peled N (2017). Exhaled breath analysis for the early detection of lung cancer: recent developments and future prospects. Lung Cancer (Auckl). 8: 31-38. DOI: 10.2147/LCTT.S104205

126. O'Neill HJ, Gordon SM, O'Neill MH, Gibbons RD and Szidon JP (1988). A computerized classification technique for screening for the presence of breath biomarkers in lung cancer. Clin Chem. 34: 1613-1618.

127. Ouyang DL, Chen JJ, Getzenberg RH and Schoen RE (2005). Noninvasive testing for colorectal cancer: a review. Am J Gastroenterol. 100: 1393-1403. DOI: 10.1111/j.1572-0241.2005.41427.x

128. Pan J, Xin L, Ma YF, Hu LH and Li ZS (2016). Colonoscopy Reduces Colorectal Cancer Incidence and Mortality in Patients With Non-Malignant Findings: A Meta-Analysis. Am J Gastroenterol. 111: 355-365. DOI: 10.1038/ajg.2015.418

129. Panico R, Powell WH and J-C. R (1993), A Guide to IUPAC Nomenclature of Organic Compounds: Recommendations 1993, p. In: (ed.): Blackwell Scientific publications.

130. Patterson SG, Bayer CW, Hendry RJ, Sellers N, Lee KS, Vidakovic B, Mizaikoff B and Gabram-Mendola SG (2011). Breath analysis by mass spectrometry: a new tool for breast cancer detection? Am Surg. 77: 747-751. Retrieved from http://www.ingentaconnect.com/contentone/sesc/tas

131. Pauling L, Robinson AB, Teranishi R and Cary P (1971). Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. Proc Natl Acad Sci U S A. 68: 2374-2376. Retrieved form http://www.pnas.org

132. Peled N, Hakim M, Bunn PA, Jr., Miller YE, Kennedy TC, Mattei J, Mitchell JD, Hirsch FR and Haick H (2012). Non-invasive breath analysis of pulmonary nodules. J Thorac Oncol. 7: 1528-1533. DOI: 10.1097/JTO.0b013e3182637d5f

133. Peng G, Hakim M, Broza YY, Billan S, Abdah-Bortnyak R, Kuten A, Tisch U and Haick H (2010). Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. Br J Cancer. 103: 542-551. DOI: 10.1038/sj.bjc.6605810

134. Peng G, Tisch U, Adams O, Hakim M, Shehada N, Broza YY, Billan S, Abdah-Bortnyak R, Kuten A and Haick H (2009). Diagnosing lung cancer in exhaled breath using gold nanoparticles. Nat Nanotechnol. 4: 669-673. DOI: 10.1038/nnano.2009.235

135. Peng G, Trock E and Haick H (2008). Detecting simulated patterns of lung cancer biomarkers by random network of single-walled carbon nanotubes coated with nonpolymeric organic materials. Nano Lett. 8: 3631-3635. DOI: 10.1021/nl801577u

136. Phillips M (1997). Method for the collection and assay of volatile organic compounds in breath. Anal Biochem. 247: 272-278. DOI: 10.1006/abio.1997.2069

137. Phillips M, Altorki N, Austin JH, Cameron RB, Cataneo RN, Greenberg J, Kloss R, Maxfield RA, Munawar MI, Pass HI, Rashid A, Rom WN and Schmitt P (2007). Prediction of lung cancer using volatile biomarkers in breath. Cancer Biomark. 3: 95-109. DOI: 10.3233/CBM-2007-3204

138. Phillips M, Altorki N, Austin JH, Cameron RB, Cataneo RN, Kloss R, Maxfield RA, Munawar MI, Pass HI, Rashid A, Rom WN, Schmitt P and Wai J (2008). Detection of lung cancer using weighted digital analysis of breath biomarkers. Clin Chim Acta. 393: 76-84. DOI: 10.1016/j.cca.2008.02.021

139. Phillips M, Bauer TL, Cataneo RN, Lebauer C, Mundada M, Pass HI, Ramakrishna N, Rom WN and Vallieres E (2015). Blinded Validation of Breath Biomarkers of Lung Cancer, a Potential Ancillary to Chest CT Screening. PLoS One. 10: e0142484. DOI: 10.1371/journal.pone.0142484

140. Phillips M, Beatty JD, Cataneo RN, Huston J, Kaplan PD, Lalisang RI, Lambin P, Lobbes MB, Mundada M, Pappas N and Patel U (2014). Rapid point-of-care breath test for biomarkers of breast cancer and abnormal mammograms. PLoS One. 9: e90226. DOI: 10.1371/journal.pone.0090226

141. Phillips M, Cataneo RN, Cummin AR, Gagliardi AJ, Gleeson K, Greenberg J, Maxfield RA and Rom WN (2003). Detection of lung cancer with volatile markers in the breath. Chest. 123: 2115-2123. DOI: 10.1378/chest.123.6.2115

142. Phillips M, Cataneo RN, Ditkoff BA, Fisher P, Greenberg J, Gunawardena R, Kwon CS, Rahbari-Oskoui F and Wong C (2003). Volatile markers of breast cancer in the breath. Breast J. 9: 184-191. DOI: 10.1046/j.1524-4741.2003.09427.x

143. Phillips M, Cataneo RN, Ditkoff BA, Fisher P, Greenberg J, Gunawardena R, Kwon CS, Tietje O and Wong C (2006). Prediction of breast cancer using volatile biomarkers in the breath. Breast Cancer Res Treat. 99: 19-21. DOI: 10.1007/s10549-006-9176-1

144. Phillips M, Cataneo RN, Saunders C, Hope P, Schmitt P and Wai J (2010). Volatile biomarkers in the breath of women with breast cancer. J Breath Res. 4: 026003. DOI: 10.1088/1752-7155/4/2/026003

145. Phillips M, Gleeson K, Hughes JM, Greenberg J, Cataneo RN, Baker L and McVay WP (1999). Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. Lancet. 353: 1930-1933. DOI: 10.1016/s0140-6736(98)07552-7

146. Poli D, Carbognani P, Corradi M, Goldoni M, Acampa O, Balbi B, Bianchi L, Rusca M and Mutti A (2005). Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study. Respir Res. 6: 71. DOI: 10.1186/1465-9921-6-71

147. Poli D, Goldoni M, Corradi M, Acampa O, Carbognani P, Internullo E, Casalini A and Mutti A (2010). Determination of aldehydes in exhaled breath of patients with lung cancer by means of on-fiber-derivatisation SPME-GC/MS. J Chromatogr B Analyt Technol Biomed Life Sci. 878: 2643-2651. DOI: 10.1016/j.jchromb.2010.01.022

148. Preti G, Labows JN, Kostelc JG, Aldinger S and Daniele R (1988). Analysis of lung air from patients with bronchogenic carcinoma and controls using gas chromatography-mass spectrometry. J Chromatogr. 432: 1-11. Retrieved from http://www.sciencedirect.com/

149. Prevost TC, Launoy G, Duffy SW and Chen HH (1998). Estimating sensitivity and sojourn time in screening for colorectal cancer: a comparison of statistical approaches. Am J Epidemiol. 148: 609-619. DOI: 10.1093/oxfordjournals.aje.a009687

150. Prince JT and Marcotte EM (2006). Chromatographic alignment of ESI-LC-MS proteomics data sets by ordered bijective interpolated warping. Anal Chem. 78: 6140-6152. DOI: 10.1021/ac0605344

151. Qin T, Liu H, Song Q, Song G, Wang HZ, Pan YY, Xiong FX, Gu KS, Sun GP and Chen ZD (2010). The screening of volatile markers for hepatocellular carcinoma. Cancer Epidemiol Biomarkers Prev. 19: 2247-2253. DOI: 10.1158/1055-9965.epi-10-0302

152. Queralto N, Berliner AN, Goldsmith B, Martino R, Rhodes P and Lim SH (2014). Detecting cancer by breath volatile organic compound analysis: a review of array-based sensors. J Breath Res. 8: 027112. DOI: 10.1088/1752-7155/8/2/027112

153. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM and Zwinderman AH (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 58: 982-990. DOI: 10.1016/j.jclinepi.2005.02.022

154. Rieder J, Lirk P, Ebenbichler C, Gruber G, Prazeller P, Lindinger W and Amann A (2001). Analysis of volatile organic compounds: possible applications in metabolic

disorders and cancer screening. Wien Klin Wochenschr. 113: 181-185. Retrieved from http://www.springer.com/medicine/journal/508

155. Risby TH and Solga SF (2006). Current status of clinical breath analysis. Appl Phys B-Lasers O. 85: 421-426. DOI: DOI 10.1007/s00340-006-2280-4

156. Robertson DJ, Lee JK, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, Kaltenbach T, Lieberman D, Levin TR and Rex DK (2017). Recommendations on Fecal Immunochemical Testing to Screen for Colorectal Neoplasia: A Consensus Statement by the US Multi-Society Task Force on Colorectal Cancer. Am J Gastroenterol. 112: 37-53. DOI: 10.1038/ajg.2016.492

157. Rudnicka J, Kowalkowski T, Ligor T and Buszewski B (2011). Determination of volatile organic compounds as biomarkers of lung cancer by SPME-GC-TOF/MS and chemometrics. J Chromatogr B Analyt Technol Biomed Life Sci. 879: 3360-3366. DOI: 10.1016/j.jchromb.2011.09.001

158. Rudnicka J, Walczak M, Kowalkowski T, Jezierski T and Buszewski B (2014). Determination of volatile organic compounds as potential markers of lung cancer by gas chromatography-mass spectrometry versus trained dogs. Sens Actuators B Chem. 202: 615-621. DOI: 10.1016/j.snb.2014.06.006

159. Sakumura Y, Koyama Y, Tokutake H, Hida T, Sato K, Itoh T, Akamatsu T and Shin W (2017). Diagnosis by Volatile Organic Compounds in Exhaled Breath from Lung Cancer Patients Using Support Vector Machine Algorithm. Sensors. 17: 287. DOI: 10.3390/s17020287

160. Santonico M, Lucantoni G, Pennazza G, Capuano R, Galluccio G, Roscioni C, La Delfa G, Consoli D, Martinelli E, Paolesse R, Di Natale C and D'Amico A (2012). In situ detection of lung cancer volatile fingerprints using bronchoscopic air-sampling. Lung Cancer. 77: 46-50. DOI: 10.1016/j.lungcan.2011.12.010

161. Schallschmidt K, Becker R, Jung C, Bremser W, Walles T, Neudecker J, Leschber G, Frese S and Nehls I (2016). Comparison of volatile organic compounds from lung cancer patients and healthy controls-challenges and limitations of an observational study. J Breath Res. 10: 046007. DOI: 10.1088/1752-7155/10/4/046007

162. Schmutzhard J, Rieder J, Deibl M, Schwentner IM, Schmid S, Lirk P, Abraham I and Gunkel AR (2008). Pilot study: volatile organic compounds as a diagnostic marker for head and neck tumors. Head Neck. 30: 743-749. DOI: 10.1002/hed.20779

163. Schreuders EH, Ruco A, Rabeneck L, Schoen RE, Sung JJ, Young GP and Kuipers EJ (2015). Colorectal cancer screening: a global overview of existing programmes. Gut. 64: 1637-1649. DOI: 10.1136/gutjnl-2014-309086

164. Schubert JK, Spittler KH, Braun G, Geiger K and Guttmann J (2001). CO(2)-controlled sampling of alveolar gas in mechanically ventilated patients. J Appl Physiol (1985). 90: 486-492. Retrieved from http://jap.physiology.org/

165. Sensigent Cyranose Electronic Nose. Available from: http://www.sensigent.com/products/cyranose.html. [last accessed: 22 January 2015].

166. Shaukat A, Mongin SJ, Geisser MS, Lederle FA, Bond JH, Mandel JS and Church TR (2013). Long-term mortality after screening for colorectal cancer. N Engl J Med. 369: 1106-1114. DOI: 10.1056/NEJMoa1300720

167. Shehada N, Bronstrup G, Funka K, Christiansen S, Leja M and Haick H (2015). Ultrasensitive Silicon Nanowire for Real-World Gas Sensing: Noninvasive Diagnosis of Cancer from Breath Volatolome. Nano Lett. 15: 1288-1295. DOI: 10.1021/nl504482t

168. Shen L, Shan YS, Hu HM, Price TJ, Sirohi B, Yeh KH, Yang YH, Sano T, Yang HK, Zhang X, Park SR, Fujii M, Kang YK and Chen LT (2013). Management of gastric cancer in Asia: resource-stratified guidelines. Lancet Oncol. 14: e535-e547. DOI: 10.1016/S1470-2045(13)70436-4

169. Shuster G, Gallimidi Z, Reiss AH, Dovgolevsky E, Billan S, Abdah-Bortnyak R, Kuten A, Engel A, Shiban A, Tisch U and Haick H (2011). Classification of breast cancer precursors through exhaled breath. Breast Cancer Res Treat. 126: 791-796. DOI: 10.1007/s10549-010-1317-x

170. Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RGS, Barzi A and Jemal A (2017). Colorectal cancer statistics, 2017. CA Cancer J Clin. 67: 177-193. DOI: 10.3322/caac.21395

171. Silanikove N, Leitner G and Merin U (2015). The Interrelationships between Lactose Intolerance and the Modern Dairy Industry: Global Perspectives in Evolutional and Historical Backgrounds. Nutrients. 7: 7312-7331. DOI: 10.3390/nu7095340

172. Smailyte G and Aleknaviciene B (2012). Incidence of prostate cancer in Lithuania after introduction of the Early Prostate Cancer Detection Programme. Public Health. 126: 1075-1077. DOI: 10.1016/j.puhe.2012.01.016

173. Smith CA, Tautenhahn R, Neumann S, Benton P, Conley C and Rainer J. LC/MS and GC/MS Data Analysis. Package 'xcms'. Available from: http://www.R-project.org. [last accessed: 2 May 2017].

174. Smolinska A, Hauschild AC, Fijten RR, Dallinga JW, Baumbach J and van Schooten FJ (2014). Current breathomics--a review on data pre-processing techniques and machine learning in metabolomics breath analysis. J Breath Res. 8: 027105. DOI: 10.1088/1752-7155/8/2/027105

175. Song G, Qin T, Liu H, Xu GB, Pan YY, Xiong FX, Gu KS, Sun GP and Chen ZD (2010). Quantitative breath analysis of volatile organic compounds of lung cancer patients. Lung Cancer. 67: 227-231. DOI: 10.1016/j.lungcan.2009.03.029

176. Sordo L, Barrio G, Bravo MJ, Villalbi JR, Espelt A, Neira M and Regidor E (2016). Estimating average alcohol consumption in the population using multiple sources: the case of Spain. Popul Health Metr. 14: 21. DOI: 10.1186/s12963-016-0090-4

177. Spanel P, Smith D, Holland TA, Al Singary W and Elder JB (1999). Analysis of formaldehyde in the headspace of urine from bladder and prostate cancer patients using selected ion flow tube mass spectrometry. Rapid Commun Mass Sp. 13: 1354-1359. DOI: Doi 10.1002/(Sici)1097-0231(19990730)13:14<1354::Aid-Rcm641>3.0.Co;2-J

178. Statistical Office of the European Communities (1990). EUROSTAT: Regional statistics.: Reference guide. Eurostat: Luxembourg.

179. Steeghs MM, Cristescu SM, Munnik P, Zanen P and Harren FJ (2007). An off-line breath sampling and analysis method suitable for large screening studies. Physiol Meas. 28: 503-514. DOI: 10.1088/0967-3334/28/5/005

180. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ and Kattan MW (2010). Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 21: 128-138. DOI: 10.1097/EDE.0b013e3181c30fb2

181. Stryker SJ, Wolff BG, Culp CE, Libbe SD, Ilstrup DM and MacCarty RL (1987). Natural history of untreated colonic polyps. Gastroenterology. 93: 1009-1013. DOI: Retrieved from http://www.gastrojournal.org

182. Szulejko JE, McCulloch M, Jackson J, McKee DL, Walker JC and Solouki T (2010). Evidence for Cancer Biomarkers in Exhaled Breath. IEEE Sens J. 10: 185-210. DOI: 10.1109/JSEN.2009.2035669

183. Taylor S and Pollard K (2009). Hypothesis Tests for Point-Mass Mixture Data with Application to 'Omics Data with Many Zero Values. Stat Appl Genet Mol. 8: Article 8. DOI: 10.2202/1544-6115.1425

184. Taylor S and Pollard K (2009). Hypothesis tests for point-mass mixture data with application to 'omics data with many zero values. Statistical applications in genetics and molecular biology. 8: Article 8. DOI: 10.2202/1544-6115.1425

185. Toth K, Sipos F, Kalmar A, Patai AV, Wichmann B, Stoehr R, Golcher H, Schellerer V, Tulassay Z and Molnar B (2012). Detection of methylated SEPT9 in plasma is a reliable screening method for both left- and right-sided colon cancers. PLoS One. 7: e46000. DOI: 10.1371/journal.pone.0046000

186. Ulanowska A, Kowalkowski T, Trawinska E and Buszewski B (2011). The application of statistical methods using VOCs to identify patients with lung cancer. J Breath Res. 5: 046008. DOI: 10.1088/1752-7155/5/4/046008

187. van den Velde S, Quirynen M, van Hee P and van Steenberghe D (2007). Differences between alveolar air and mouth air. Anal Chem. 79: 3425-3429. DOI: 10.1021/Ac062009a

188. Vijan S, Hwang I, Inadomi J, Wong RK, Choi JR, Napierkowski J, Koff JM and Pickhardt PJ (2007). The cost-effectiveness of CT colonography in screening for colorectal neoplasia. Am J Gastroenterol. 102: 380-390. DOI: 10.1111/j.1572-0241.2006.00970.x

189. Wang C, Ke C, Wang X, Chi C, Guo L, Luo S, Guo Z, Xu G, Zhang F and Li E (2014). Noninvasive detection of colorectal cancer by analysis of exhaled breath. Anal Bioanal Chem. 406: 4757-4763. DOI: 10.1007/s00216-014-7865-x

190. Wang C, Sun B, Guo L, Wang X, Ke C, Liu S, Zhao W, Luo S, Guo Z, Zhang Y, Xu G and Li E (2014). Volatile organic metabolites identify patients with breast cancer, cyclomastopathy, and mammary gland fibroma. Sci Rep. 4: 5383. DOI: 10.1038/srep05383

191. Wang D, Yu K, Wang Y, Hu Y, Zhao C, Wang L, Ying K and Wang P (2012). A Hybrid Electronic Noses' System Based on Mos-Saw Detection Units Intended for Lung Cancer Diagnosis. J Innov Opt Health Sci. 5: 1150006-1. DOI: 10.1142/S1793545811500064

192. Wang Y, Hu Y, Wang D, Yu K, Wang L, Zou Y, Zhao C, Zhang X, Wang P and Ying K (2012). The analysis of volatile organic compounds biomarkers for lung cancer in

exhaled breath, tissues and cell lines. Cancer Biomark. 11: 129-137. DOI: 10.3233/cbm-2012-00270

193. Warren JD, Xiong W, Bunker AM, Vaughn CP, Furtado LV, Roberts WL, Fang JC, Samowitz WS and Heichman KA (2011). Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. BMC Med. 9: 133. DOI: 10.1186/1741-7015-9-133

194. Webb-Robertson BJ, Wiberg HK, Matzke MM, Brown JN, Wang J, McDermott JE, Smith RD, Rodland KD, Metz TO, Pounds JG and Waters KM (2015). Review, evaluation, and discussion of the challenges of missing value imputation for mass spectrometry-based label-free global proteomics. J Proteome Res. 14: 1993-2001. DOI: 10.1021/pr501138h

195. Wehinger A, Schmid A, Mechtcheriakov S, Ledochowski M, Grabmer C, Gastl GA and Amann A (2007). Lung cancer detection by proton transfer reaction mass-spectrometric analysis of human breath gas. Int J Mass Spectrom. 265: 49-59. DOI: 10.1016/j.ijms.2007.05.012

196. Weiss G and Roesch T (2010). Potential of a new blood test for colorectal cancer screening—the Septin 9 gene biomarker. Eur Oncol Haematol. 6: 51-54. DOI: 10.17925/EOH.2010.06.1.51

197. Wenig P and Odermatt J (2010). OpenChrom: a cross-platform open source software for the mass spectrometric analysis of chromatographic data. BMC Bioinformatics. 11: 405. DOI: 10.1186/1471-2105-11-405

198. Westhoff M, Litterst P, Freitag L, Urfer W, Bader S and Baumbach JI (2009). Ion mobility spectrometry for the detection of volatile organic compounds in exhaled breath of patients with lung cancer: results of a pilot study. Thorax. 64: 744-748. DOI: 10.1136/thx.2008.099465

199. Williams H and Pembroke A (1989). Sniffer dogs in the melanoma clinic? Lancet. 1: 734. DOI: S0140-6736(89)92257-5

200. Wilson AD and Baietto M (2011). Advances in electronic-nose technologies developed for biomedical applications. Sensors. 11: 1105-1176. DOI: 10.3390/s110101105

201. Xu H, Wei Y, Zhu L, Huang J, Li Y, Liu F, Wang S and Liu S (2014). Bifunctional magnetic nanoparticles for analysis of aldehyde metabolites in exhaled breath of lung cancer patients. J Chromatogr A. 1324: 29-35. DOI: 10.1016/j.chroma.2013.11.041

202. Xu ZQ, Broza YY, Ionsecu R, Tisch U, Ding L, Liu H, Song Q, Pan YY, Xiong FX, Gu KS, Sun GP, Chen ZD, Leja M and Haick H (2013). A nanomaterial-based breath test for distinguishing gastric cancer from benign gastric conditions. Br J Cancer. 108: 941-950. DOI: 10.1038/bjc.2013.44

203. Yang ZX, Huang Y, Chen G, Guo ZP, Cheng SY and Huang SZ (2009). Ethanol gas sensor based on Al-doped ZnO nanomaterial with many gas diffusing channels. Sens Actuators B Chem. 140: 549-556. DOI: 10.1016/j.snb.2009.04.052

204. Yoruker EE, Holdenrieder S and Gezer U (2016). Blood-based biomarkers for diagnosis, prognosis and treatment of colorectal cancer. Clin Chim Acta. 455: 26-32. DOI: 10.1016/j.cca.2016.01.016

205. Yu K, Wang Y, Yu J and Wang P (2011). A Portable Electronic Nose Intended for Home Healthcare Based on a Mixed Sensor Array and Multiple Desorption Methods. Sens Lett. 9: 876-883. DOI: 10.1166/sl.2011.1635

206. Yuan Y (2013). A survey and evaluation of population-based screening for gastric cancer. Cancer Biol Med. 10: 72-80. DOI: 10.7497/j.issn.2095-3941.2013.02.002

207. Zhou X-H, Obuchowski NA and McClish DK (2011), Statistical Methods in Diagnostic Medicine, 2nd Edition. John Wiley & Sons, Inc., New York, p. 47-50.

208. Zou X, Zhou W, Lu Y, Shen C, Hu Z, Wang H, Jiang H and Chu Y (2016). Exhaled gases online measurements for esophageal cancer patients and healthy people by proton transfer reaction mass spectrometry. J Gastroenterol Hepatol. 31: 1837-1843. DOI: 10.1111/jgh.13380

209. Zou Y, Zhang X, Chen X, Hu Y, Ying K and Wang P (2014). Optimization of volatile markers of lung cancer to exclude interferences of non-malignant disease. Cancer Biomark. 14: 371-379. DOI: 10.3233/CBM-140418

LIST OF OWN PUBLICATIONS

Articles related to this dissertation:

1. **Krilaviciute A**, Heiss JA, Leja M, Kupcinskas J, Haick H, Brenner H (2015) Detection of cancer through exhaled breath: a systematic review. Oncotarget. 6: 38643-38657. DOI: 10.18632/oncotarget.5938
2. **Krilaviciute A**, Stock C, Brenner H. International variation in the prevalence of preclinical colorectal cancer: implications for predictive values of noninvasive screening tests and potential target populations for screening. Int J Cancer. DOI: 10.1002/ijc.30867 **[in press]**
3. **Krilaviciute A**, Stock C, Leja M, Brenner H. Potential for reduction of numbers needed to screen for gastric cancer by preselection of screening populations with breath testing. Gastric Cancer **[submitted]**
4. **Krilaviciute A**, Haitham A, Barash-Adams O, Khatib S, Brenner H, Leja M, Association analysis of VOCs and lifestyle factors. **[in preparation]**

Other articles:

1. **Krilaviciute A**., Smailyte G, Brenner H and Gondos A (2014) Cancer survival in Lithuania after the restoration of independence: rapid improvements, but persisting major gaps. Acta Oncol. 53: 1238-1244. DOI: 10.3109/0284186X.2014.888495
2. Smailyte G, Jasilionis D, Vincerzevskiene I, **Krilaviciute A**, Ambrozaitiene D, Stankuniene V and Shkolnikov VM (2015) Educational differences in incidence of cancer in Lithuania, 2001-2009: evidence from census-linked cancer registry data. Eur J Cancer Prev. 24: 261-266. DOI: 10.1097/CEJ.0000000000000036.
3. Jansen L, Castro FA, Gondos A, **Krilaviciute A**, Barnes B, Eberle A, Emrich K, Hentschel S, Holleczek B, Katalinic A, Brenner H, and G.C.S.W. Group (2015) Recent cancer survival in Germany: an analysis of common and less common cancers. Int J Cancer. 136: 2649-2458. DOI: 10.1002/ijc.29316
4. Castro FA, Jansen L, **Krilaviciute A**, Katalinic A, Pulte D, Sirri E, Ressing M, Holleczek B, Luttmann S, Brenner H, and G.C.S.W. Group (2015) Survival of patients

with gastric lymphoma in Germany and in the United States. J Gastroenterol Hepatol. 30: 1485-1491. DOI: 10.1111/jgh.12989

5. Gondos A, **Krilaviciute A**, Smailyte G, Ulys A and Brenner H (2015) Cancer surveillance using registry data: Results and recommendations for the Lithuanian national prostate cancer early detection programme. Eur J Cancer. 51: 1630-1637. DOI: 10.1016/j.ejca.2015.04.009
6. Eberle A, Jansen L, Castro FA, **Krilaviciute A**, Luttmann S, Emrich K, Holleczek B, Nennecke A, Katalinic A, Brenner H, and G.S.W. Group (2015) Lung cancer survival in Germany: A population-based analysis of 132,612 lung cancer patients. Lung Cancer. 90: 528-533. DOI: 10.1016/j.lungcan.2015.10.007
7. Pulte D, Jansen L, Castro FA, **Krilaviciute A**, Katalinic A, Barnes B, Ressing M, Holleczek B, Luttmann S, Brenner H, and G.C.S.W. Group (2016) Survival in patients with acute myeloblastic leukemia in Germany and the United States: Major differences in survival in young adults. Int J Cancer, 2016. 139: 1289-1296. DOI: 10.1002/ijc.30186
8. **Krilaviciute A**, Vincerzevskiene I and Smailyte G (2016) Basal cell skin cancer and the risk of second primary cancers: a cancer registry-based study in Lithuania. Ann Epidemiol, 2016. 26: 511-514. DOI: 10.1016/j.annepidem.2016.05.003
9. Winter A, Sirri E, Jansen L, Wawroschek F, Kieschke J, Castro FA, **Krilaviciute A**, Holleczek B, Emrich K, Waldmann A, Brenner H, and Association of Population-based Cancer Registries in Germany Cancer Survival Working Group (2017) Comparison of prostate cancer survival in Germany and the USA: can differences be attributed to differences in stage distributions? BJU Int. 119: 550-559. DOI: 10.1111/bju.13537

Poster presentations:

Krilaviciute A, Heiss JA, Leja M, Kupcinskas J, Haick H and Brenner H (2015). *Detection of cancer through exhaled breath: a systematic review*. IABR summit 2015, Vienna, Austria.

APPENDIX

Appendix 1. Reasons for excluding full papers from the systematic review on breath analysis for various cancers detection.

Improper study design: 1–Study on mice with results validated on people (3 cases and 3 controls); 2–Study on stomach tissue with results validated on people (3 cases and 10 controls); 3–Study with simulated data, no real patient data

1. Ebeler SE, Clifford AJ, Shibamoto T. Quantitative analysis by gas chromatography of volatile carbonyl compounds in expired air from mice and human. J Chromatogr B Biomed Sci Appl. 1997; 702(1-2): 211-5.
2. Ligor T, Szeliga J, Jackowski M, Buszewski B. Preliminary study of volatile organic compounds from breath and stomach tissue by means of solid phase microextraction and gas chromatography-mass spectrometry. J Breath Res. 2007; 1(1): 016001.
3. Peng G, Trock E, Haick H. Detecting simulated patterns of lung cancer biomarkers by random network of single-walled carbon nanotubes coated with nonpolymeric organic materials. Nano Lett. 2008; 8(11): 3631-5.

New method described, no statistical information about performance of this method provided

4. Na N, Liu H, Han J, Han F, Liu H, Ouyang J. Plasma-Assisted Cataluminescence Sensor Array for Gaseous Hydrocarbons Discrimination. Analytical Chemistry. 2012; 84(11): 4830-6.
5. Zhang G, Guo X, Wang S, Wang X, Zhou Y, Xu H. New graphene fiber coating for volatile organic compounds analysis. J Chromatogr B Analyt Technol Biomed Life Sci. 2014; 969: 128-31.

Missing required information: 6–No cancer cases; 7, 8–No controls; 9, 10–Missing information on number of cancer cases and controls

6. Silva LIB, Freitas AC, Rocha-Santos TAP, Pereira ME, Duarte AC. Breath analysis by optical fiber sensor for the determination of exhaled organic compounds with a view to diagnostics. Talanta. 2011; 83(5): 1586-94.
7. Chatterjee S, Castro M, Feller JF. An e-nose made of carbon nanotube based quantum resistive sensors for the detection of eighteen polar/nonpolar VOC biomarkers of lung cancer. Journal of Materials Chemistry B. 2013; 1(36): 4563-75.
8. Hou C, Lei J, Huo D, Song K, Li J, Luo X, et al. Discrimination of Lung Cancer Related Volatile Organic Compounds with a Colorimetric Sensor Array. Analytical Letters. 2013; 46(13): 2048-59.
9. Kumar S, Huang J, Abbassi-Ghadi N, Spanel P, Smith D, Hanna GB. Selected ion flow tube mass spectrometry analysis of exhaled breath for volatile organic compound profiling of esophago-gastric cancer. Anal Chem. 2013; 85(12): 6121-8.
10. Wu Y, Huo D, Hou C, Fa H, Yang M, Luo X. Colorimetric Artificial Nose for Identification of Breath Volatile Organic Compounds of Patients with Lung Cancer. Chemical Research in Chinese Universities. 2014; 30(4): 572-7.

No statistical information provided: 11–Visual representation of results, some cases were excluded without explanation; 12– Concentrations of volatile organic compounds were measured, but no statistical data provided; 13–Different sensors parameters were tested, but no data on overall classification model provided

11. Gaspar EM, Lucena AF, Duro da Costa J, Chaves das Neves H. Organic metabolites in exhaled human breath--a multivariate approach for identification of biomarkers in lung disorders. J Chromatogr A. 2009; 1216(14): 2749-56.
12. Yu H, Xu L, Wang P. Solid phase microextraction for analysis of alkanes and aromatic hydrocarbons in human breath. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences. 2005; 826(1-2): 69-74.
13. Tran VH, Chan HP, Thurston M, Jackson P, Lewis C, Yates D, et al. Breath Analysis of Lung Cancer Patients Using an Electronic Nose Detection System. Ieee Sensors Journal. 2010; 10(9): 1514-8.

Other reasons: 14–Different masses (compounds) were measured, but not identified; 15–Conference paper, full paper published elsewhere

14. Schmutzhard J, Rieder J, Deibl M, Schwentner IM, Schmid S, Lirk P, et al. Pilot study: volatile organic compounds as a diagnostic marker for head and neck tumors. Head Neck. 2008; 30(6): 743-9.
15. Chen X, Cao M, Hao Y, Li Y, Wang P, Ying K, et al. A Non-invasive detection of lung cancer combined virtual gas sensors array with imaging recognition technique. Conf Proc IEEE Eng Med Biol Soc. 2005; 6: 5873-6.

Appendix 2. List of the studies included in the systematic review on exhaled breath analysis for cancer detection.

First author	Year	Country	Cancer site	Reference
Altomare	2013	Italy	Colorectal cancer	(4)
Amal	2013	Latvia	Gastric cancer	(5)
Amal	2015	Latvia	Gastric cancer	(7)
Amal	2015	China	Ovarian cancer	(8)
Bajtarevic	2009	Austria	Lung cancer	(16)
Bousamra	2014	The U.S.	Lung cancer	(30)
Broza	2013	Israel	Lung cancer	(36)
Buszewski	2012	Poland	Lung cancer	(38)
Chapman	2012	Australia	Malignant mesothelioma	(39)
Chen	2005	China	Lung cancer	(40)
Dragonieri	2009	The Netherlands	Lung cancer	(53)
D'Amico	2010	Italy	Lung cancer	(49)
de Genaro	2010	Italy	Malignant mesothelioma	(50)
Di *Natale*	2003	Italy	Lung cancer	(51)
Dragonieri	2012	Italy	Malignant mesothelioma	(54)
Filipiak	2014	Austria	Lung cancer	(58)
Fu	2014	The U.S.	Lung cancer	(61)
Fuchs	2010	Germany	Lung cancer	(62)
Garcia	2014	Spain	Head and neck cancer	(63)
Gordon	1985	The U.S.	Lung cancer	(66)
Gruber	2014	Israel	Head and neck cancer	(67)
Guo	2015	China	Thyroid caner	(68)
Hakim	2011	Israel	Lung cancer, head and neck cancer	(71)
Handa	2014	Japan	Lung cancer	(72)
Hauschild	2012	Germany	Lung cancer	(74)
Hietanen	1994	The United Kingdom	Breast cancer	(77)
Hubers	2014	The Netherlands	Lung cancer	(81)
Kischkel	2010	Germany	Lung cancer	(89)
Kumar	2015	The United Kingdom	Gastric cancer	(96)
Leunis	2014	The Netherlands	Head and neck cancer	(105)
Li	2014	China	Breast cancer	(107)
Ligor	2009	Austria	Lung cancer	(108)
Ma	2014	China	Lung cancer	(110)
Machado	2005	The U.S.	Lung cancer	(111)
Mangler	2012	Germany	Breast cancer	(112)
Mazzone	2007	The U.S.	Lung cancer	(114)
Mazzone	2012	The U.S.	Lung cancer	(115)
McWilliams	2015	Canada	Lung cancer	(118)

(Continues on the next page)

(Continued)

First author	Year	Study population	Cancer site	Reference
Patterson	2011	The U.S.	Breast cancer	(130)
Peled	2012	The U.S.	Lung cancer	(132)
Peng	2009	Israel	Lung cancer	(134)
Peng	2010	Israel	Colorectal, prostate, breast and lung cancer	(133)
Phillips	1999	The U.S.	Lung cancer	(145)
Phillips	2003	The U.S. / the United Kingdom	Lung cancer	(141)
Phillips	2003	The U.S.	Breast cancer	(142)
Phillips	2006	The U.S.	Breast cancer	(143)
Phillips	2007	The U.S.	Lung cancer	(137)
Phillips	2008	The U.S.	Lung cancer	(138)
Phillips	2010	The U.S.	Breast cancer	(144)
Phillips	2014	The U.S. / the Netherlands	Breast cancer	(140)
P*oli*	2005	Italy	Lung cancer	(146)
Poli	2010	Italy	Lung cancer	(147)
Preti	1988	The U.S.	Lung cancer	(148)
Qin	2010	China	Liver cancer	(151)
Rieder	2001	Austria	Heamatological and gynecological cancers	(154)
Rudnicka	2011	Poland	Lung cancer	(157)
Rudnicka	2014	Poland	Lung cancer	(158)
Santonico	2012	Italy	Lung cancer	(160)
Shehada	2014	Latvia	Gastric cancer	(167)
Shuster	2011	Israel	Breast cancer	(169)
Song	2010	China	Lung cancer	(175)
Steeghs	2007	The Netherlands	Lung cancer	(179)
Ulanowska	2011	Poland	Lung cancer	(186)
Wang *C*	2014	China	Colorectal cancer	(189)
Wang *C*	2014	China	Breast cancer	(190)
Wang *D*	2012	China	Lung cancer	(191)
Wang *Y*	2012	China	Lung cancer	(192)
Wehinger	2007	Austria	Lung cancer	(195)
Westhoff	2009	Germany	Lung cancer	(198)
Xu *H*	2014	China	Lung cancer	(201)
Xu *Z*	2013	China	Gastric cancer	(202)
Yu	2011	China	Lung cancer	(205)
Zou	2014	China	Lung cancer	(209)

Appendix 3. Study population characteristics of studies included in the systematic review on breath analysis for cancer detection.

Ref	Idx[a]	Cases			Controls			Comments
		Sex[b]	Age[c]	Smoking[d]	Sex[b]	Age[c]	Smoking[d]	
(4)		20 / 17 7 / 8	63 ±10 67 ±11	NA NA	13 / 28 6 / 4	47 ±12 56 ±10	NA NA	Training set Validation set
(5)		28 / 9	57 ±2	NA	30 / 31	57 ±2	NA	
(7)		56 / 42	63 ±13	29 / NA / NA	102 / 223 34 / 19	59 ±14 53 ±15	45 / NA / NA 23 / NA / NA	Cn: OLGIM stages 0-IV Cn: gastric ulcer
(8)		0 / 48	51 ±11	0 / 0 / 48	0 / 48 0 / 86 0 / 134	48 ±9 40 ±13 43[WA]	0 / 0 / 48 0 / 0 / 86 0 / 0 / 134	Cn: healthy Cn: benign conditions Cn: all
(17, 108)		41 / 24	63 (37-84)	28 / 31 / 6	15 / 16	38 (21-87)	7 / 2 / 22	
(30)		107	66 ±10	44 / 51 / 12	40	51 ±15	12 / 14 / 14	Cn: benign conditions
(36)	×	7 / 5	65 ±7	5 / 5 / 2	3 / 2	64 ±7	2 / 2 / 1	Cn: benign conditions
(38)		29	NA	NA	44	NA	NA	
(39)		18 / 2	69 ±10	0 / 12 / 8	34 / 8 18 / 0	67 ±14 71[WA]	0 / 12 / 30 0 / 12 / 6	Cn: healthy Cn: lung diseases
(40)		20 5	NA NA	NA NA	22 5	NA NA	NA NA	Training set Validation set
(49)		28	62 ±6	0 / 17 / 11	36 28	na (50-70) 61 ±7	0 / 0 / 36 0 / 17 / 11	Cn: healthy Cn: lung diseases
(50, 54)		11 / 2	61 ±12	0 / 5 / 8	5 / 8 9 / 4	52 ±16 67 ±10	0 / 0 / 13 0 / 4 / 9	Cn: healthy Cn: exposed to asbestos
(51)		35 9	NA NA	NA NA	18	NA	NA	Cn: normal Cn: after surgery
(53)		10 / 0	66 ±9	2 / 7 / 1	4 / 6 8 / 2	58 ±8 61 ±6	0 / 0 / 10 6 / 4 / 0	Cn: healthy Cn: Chronic obstructive pulmonary disease
(58)		25 / 11	63 ±7	0 / NA / NA	12 / 16	52 ±17	0 / NA / NA	
(61)		97	NA	NA	32	NA	NA	Cn: benign conditions
(62)		11 / 1	68 ±9	0 / 12 / 0	10 / 14	36 ±12	12 / 0 / 12	
(63)		10 / 1	61 ±11	NA	6 / 4	45 ±15	5 / 0 / 5	
(66)		11 / 1	61 ±10	NA / NA / 0	8 / 1	? (25-70)	2 / NA / NA	
(67)		19 / 3	62 ±12	13 / NA / NA	6 / 14 14 / 7	50 ±12 55 ±14	5 / NA / NA 12 / NA / NA	Cn: healthy Cn: benign conditions
(68)		11 / 53	48[WA]	13 / NA / NA	6 / 26	40 ±13	4 / NA / NA	
(71)		19 / 3 22 / 3	60 ±9 66 ±8	NA	17 / 23	45 ±13	NA	Cs: head and neck cancer Cs: lung cancer
(72)		31 / 19	68 ±10	NA / NA / 17	25 / 14	32 ±8	NA / NA / 32	
(74)		54	NA	NA	35	NA	NA	
(77)		0 / 20	61 (37-85)	NA	NA	NA	NA	Cn: benign conditions
(81)		12 / 8 10 / 8	65 ±8 64 ±8	7 / 12 / 0 10 / 8 / 0	22 / 9 4 / 4	65 ±9 53 ±7	12 / 19 / 0 4 / 4 / 0	Training set Validation set
(89)		23 / 8	68 ±?	0 / 29 / 2	19 / 43	34[WA]	35 / 0 / 27	

(Continues on the next page)

(Continued)

Ref	Idx[a]	Cases			Controls			Comments
		Sex[b]	Age[c]	Smoking[d]	Sex[b]	Age[c]	Smoking[d]	
(96)		64 / 17	62 (53-71)	14 / 32 / 35	89 / 50	64 (51-72)	24 / 32 / 73	
(105)		25 / 11	59 ±NA	36 / 0 / 0	10 / 13	48 ±NA	23 / 0 / 0	
(107)		0 / 22	NA	0 / 0 / 22	0 / 24	NA	0 / 0 / 24	
(110)		10 / 3	63 ±7	5 / 0 / 8	16 / 9	35[WA]	8 / 0 / 17	
(111)		10 / 4 10 / 4	64 ±3 61 ±13	2 / 12 / 0 NA	18 / 27 31 / 31	45[WA] 45[WA]	6 / 17 / 22 NA / NA / ≥30	Training set Validation set
(112)		0 / 10	65 ±NA	NA	0 / 10	59 ±NA	NA	
(114)		24 / 25	65 ±NA	13 / 35 / 1	44 / 50	56[WA]	5 / 48 / 41	
(115)		49 / 43	69 ±NA	25 / 58 / 9	65 / 72	59 ±NA	28 / 71 / 35	
(118)		12 / 13	67 ±6	9 / 16 / 0	86 / 80	63 ±7	87 / 79 / 0	
(130)		0 / 20	53 ±NA	NA	0 / 20	55 ±NA	NA	
(132)	×	31 / 22	65 ±7	19 / 26 / 8	15 / 4	61 ±7	6 / 11 / 2	Cn: benign conditions
(134)		40	na (28-60)	0 / 15 / 25	56	NA (28-60)	17 / 0 / 39	
(133)		18 26 0 / 22 30	NA NA NA NA	3 / 3 / 12 10 / 5 / 11 1 / 1 / 19 12 / 5 / 8	22	NA	4 / 0 / 17	Cs: prostate cancer Cs: colorectal cancer Cs: breast cancer Cs: colorectal cancer
(145)	×	34 / 26	67 ±13	NA / NA / 5	29 / 19	61 ±13	NA / NA / 12	Cn: abnormal chest x-rays
(141)	×	48 / 19	68 ±10	NA / NA / 3	16 / 25 41 / 50	70 ±13 58 ±14	NA / NA / 18 NA	Cn: healthy Cn: abnormal chest x-rays
(137, 138)		96 / 97	66 ±11	33 / 134 / 24	106 / 105	68 ±6	78 / 133 / 0	
(142, 143)	×	0 / 51	61 ±12	12 / 6 / 31	0 / 42 0 / 50	64 ±18 NA	3 / 11 / 20 NA	Cn: healthy Cn: abnormal mammogram
(144)		0 / 54	55 ±7	2 / 0 / 38	0 / 204	55 ±11	26 / 0 / 177	
(140)	×	0 / 35	NA	NA	0 / 93 0 / 79	NA NA	NA NA	Cn: healthy Cn: abnormal mammogram
(146)		28 / 8	67 (?)	2 / 28 / 6	27 / 23 30 / 5 18 / 7	56 (?) 54 (?) 70 (?)	0 / 0 / 50 35 / 0 / 0 1 / 21 / 3	Cn: non-smokers Cn: smokers Cn: Chronic obstructive pulmonary disease
(147)		28 / 12	68 ±10	21 / 12 / 7	17 / 21	49 ±15	0 / 10 / 28	
(148)		7 / 3	66 ±7	4 / 6 / 0	4 / 4 1 / 7	60 ±4 28 ±6	4 / 1 / 3 3 / 1 / 4	Cn: age-matched Cn: younger
(151)		26 / 4	53 ±12	8 / 0 / 22	24 / 12 18 / 9	49 ±11 52 ±11	7 / 0 / 29 3 / 2 / 22	Cn: healthy Cn: hepatocirrhosis
(154)		16	NA	NA	100	NA	NA	
(157)		17 / 6	NA (51-78)	21 / 0 / 2	10 / 20	36[WA]	6 / NA / NA	
(158)		76 / 31	51[WA]	NA	31 / 90	49[WA]	NA	
(160)		16 / 4	67 ±8	8 / 7 / 5	8 / 2	65 ±8	3 / 4 / 3	Cn: benign conditions

(Continues on the next page)

(Continued)

Ref	Idx[a]	Cases			Controls			Comments
		Sex[b]	Age[c]	Smoking[d]	Sex[b]	Age[c]	Smoking[d]	
(167)		19 / 11	60 ±10	9 / 0 / 19	24 / 53	55 ±16	14 / 0 / 63	
(169)		0 / 13	NA	3 / 1 / 9	0 / 7 0 / 16	NA NA	2 / 0 / 5 4 / 0 / 10	Cn: healthy Cn: cancer precursors
(175)		34 / 9	58 ±8	0 / 21 / 22	34 / 7	48 ±7	0 / 0 / 41	
(179)		11	56 ±5	11 / 0 / 0	57 / 0	60 ±5	57 / 0 / 0	
(186)		99 / 38	NA (38-86)	NA	40 / 103	NA (20-58)	41 / 0 / 102	
(189)		13 / 7	58 ±14	5 / NA / NA	8 / 12	50 ±9	8 / NA / NA	
(190)		0 / 39	53 ±11	2 / NA / NA	0 / 91	42[WA]	1 / NA / NA	
(191)		35 / 12	NA (32-80)	24 / 0 / 23	14 / 28	NA (30-75)	6 / 0 / 36	
(192)		24 / 61	62 ±10	47 / 0 / 38	36 / 52 26 / 44 62 / 96	45 ±9 54 ±13 49[WA]	29 / 0 / 59 23 / 0 / 47 52 / 0 / 106	Cn: healthy Cn: benign conditions Cn: all
(195)		13 / 4	62 ±11	9 / 5 / 4	83 / 87	41 ±13	60 / 11 / 95	
(198)		24 / 8	65 ±10	7 / 17 / 6	39 / 15	46 ±12	12 / 0 / 42	
(201)		6	NA	NA	8	NA	NA	
(202)	×	28 / 9	58 ±9	15 / NA / NA	23 / 9 30 / 31 53 / 40	51 ±14 51 ±9 51[WA]	14 / NA / NA 13 / NA / NA 27 / NA / NA	Cn: gastric ulcer Cn: less severe conditions Cn: all
(205)		9	NA	7 / NA / NA	9	NA	5 / NA / NA	
(209)		58 / 21 58	63 ±8 NA	15 / 40 / 24 NA	63 / 29 20	57[WA] NA	21 / 45 / 26 NA	Training set Validation set

Cn, controls; Cs, cases; NA, data not provided; OLGIM - operative link on gastric intestinal metaplasia assessment.

[a] Studies that recruited patients before determining final diagnosis.

[b] Number of males / females or total number of participants if sex-specific data were not provided.

[c] Median age (range) or mean age ±standard deviation.

[d] Number of smokers / ex-smokers / non-smokers

[WA] Weighted average.

Appendix 4. Differences in proportions of age, sex and smoking distributions between cases and controls, restrictions applied before sampling, analysis time and inclusion criteria for all studies.

ref	Sex / age / smoking[a]	Restrictions before sampling (collection time)	Room air[b]	Analysis time after collection	Treat-ment[c]	Exclusion/inclusion criteria	Comments
(4)	A: N / N / NA B: N / N / NA	3 h no food / drinks; rest 10 min; 5 min lung washout	NA	At once	No	Cs: no asthma, severe COPD, unstable diabetes, other cancer. Cn: no inflammatory bowel disease or diverticulitis	A: training B: validation
(5)	N / Y / NA	no food / drinks; 2 h no smoking / alcohol; 3 min lung washout (morning)	Yes	≤3 months	NA	Cn: no grade III–IV atrophic gastritis, no patients following stomach resections	
(8)	A: Y / Y / Y B: Y / N / Y C: Y / Y / Y	2 h no food / coffee / alcohol; no cosmetics; 3 min lung washout	Yes	≤4 months	NA	A: no cancer, no chronic disease, no autoimmune diseases	A: healthy B: benign C = A + B
(7)	A: N / Y / Y B: Y / N / N	no food / drinks; 3 h no smoking; 3 min lung washout; (morning)	Yes	≤6 months	No	A, B: no past stomach surgery B: no dysplasia	A: OLGIM 0-IV B: peptic ulcer
(16, 108)	N / N / N	1 h no food; rest 5min (any time)	Yes	3-6 h	Yes	NA	
(30)	A: NA / N / N B: NA / N / N	(morning)	NA	At once	No	A: no active pulmonary disease	A: healthy B: benign
(36)	Y / Y / Y	1 h no food / coffee / smoking; 3 min lung washout	NA	≤6 h	No	Cs: hospitalized for benign conditions	
(38)	NA / NA / NA	NA	Yes	NA	NA	NA	
(39)	A: Y / Y / Y B: Y / Y / Y	1.5 h no food / drinks; 20 min rest; mouth wash with distillated water	NA	At once	NA	A: no recent respiratory tract infection, no acute exacerbation of any underlying respiratory disease in 4 weeks, no current uncontrolled medical conditions	A: healthy B: other diseases
(40)	A: NA / NA / NA B: NA / NA / NA	2 h no food	NA	At once	Yes	Cn: bronchitis (N=7)	A: training B: validation
(49)	A: NA / NA / Y B: NA / Y / Y	no food / drinks (morning)	NA	At once	No	Patients: no current therapy A: no other disease B: COPD (N=16), bronchitis (N=5), other diseases (N=7)	A: healthy B: other diseases
(50) (54)	A: N / Y / Y B: N / Y / Y	3 h no food / drinks; 5 min lung washout	Yes	At once	No	Cs: no other pulmonary or cardiovascular abnormalities. A: no history of respiratory tract infection in 4 weeks, no asthma, COPD, systemic diseases (e.g., diabetes), no cancer, no drugs	A: healthy B: exposed to asbestos
(51)	A: NA / NA / NA B: NA / NA / NA	no food (morning)	NA	At once	No	Cs: no therapy A: no apparent disease, no drugs	A: controls B: cases after surgery
(53)	A: N / Y / N B: N / Y / N	no smoking; 2 h no food / drinks; 5 min lung washout	Yes	At once	No	Cs: no respiratory tract infections in 4 weeks, no systematic diseases / other cancer A: no diseases B: no asthma, pulmonary / cardiovascular diseases	A: healthy B: COPD

(Continues on the next page)

(Continued)

ref	Sex / age / smoking[a]	Restrictions before sampling (collection time)	Room air[b]	Analysis time after collection	Treat-ment[c]	Exclusion/inclusion criteria	Comments
(58)	N / N / Y	2 h no food; rest 10 min (morning)	Yes	≤5 h	NA	NA	
(61)	NA / NA / NA	NA	NA	At once	No	NA	
(62)	N / N / N	rest 10 min	NA	≤6 h	No	Cs: no cancer, no chronic disease	
(66)	Y / NA / N	5 min lung washout	NA	NA	No	Cs: no comorbidities Cn: no pulmonary / systematic disease	
(71)	NA / NA / NA	lung washout	Yes	≤3 days	No	NA	
(63)	N / N / NA	(morning)	Yes	-≤12 h (if possible)	NA	NA	
(67)	A: N / N / N B: N / Y / N	12 h no food / drinks / alcohol; 3-5 min lung washout, (morning)	Yes	na	No	A: no previous cancer, active infectious disease, present antibiotic treatment, pregnancy or lactation	A: healthy B: benign
(68)	Y / Y / Y	no food / drinks (morning)	Yes	≤3 h	NA	All: no pregnancy, lactation, no congenital, chronic inflammatory, acute or infectious disease in 2 weeks, no family history of mental illness	
(72)	Y / N / N	NA	NA	At once	No	NA	
(74)	A: NA / NA / NA B: NA / NA / NA	NA	NA	At once	NA	NA	A: healthy B: COPD
(77)	Y / Y[d] / Y[d]	NA	Yes	NA	NA	Cs: some comorbidities (N=5) Cn: hospitalized for benign conditions	
(81)	A: N / Y / Y B: Y / N / Y	2 h no food / drinks / smoking; 5 min lung washout	Yes	At once	No	Cn: no therapy Cs: no cancer, majority had COPD	A: training B: validation
(96)	N / Y / Y	6 h no food / drinks (morning)	Yes	≤1 h	No	All: no liver disease, small bowel/colonic pathology, other cancers, no acute infection. Some had comorbidities Cs: no squamous cell carcinoma of the upper gastrointestinal tract Cn: Barrett's metaplasia (N=16), benign condition (N=62)	
(89)	A: N / N / N B: N / N / Y	1 h no food / drinks; rest 10min	Yes	≤6 h	No	A, B: no COPD	A: healthy smokers B: healthy non-smokers
(105)	N / N / Y	8 h no food (morning)	NA	At once	No	Cn: visiting hospital with some conditions	
(107)	A: Y / NA / Y B: Y / NA / Y	rest 10 min	Yes	≤12 h	No	A: no cancer, no breast disease	A: healthy B: benign

(Continues on the next page)

(Continued)

ref	Sex / age / smoking[a]	Restrictions before sampling (collection time)	Room air[b]	Analysis time after collection	Treat-ment[c]	Exclusion/inclusion criteria	Comments
(110)	N / N / Y	24 h no spicy food; 10 h no smoking / alcohol; 2 h no brushing; 30 min no food / drinks; stay 30 min in ventilated room (before lunch)	Yes	≤24 h	Yes	Cn: no history of severe COPD, asthma, neurological disorder, Wilson's disease, diabetes, no sedatives or narcotics within 48 h	
(111)	Y / Y / N	NA	Yes	At once	Yes	Cn: no lung conditions, cardiopulmonary symptoms	
(112)	Y / Y / NA	8 h no food / drinks; no tooth brushing (morning)	Yes	NA	No	Cn: no diseases	
(114)	Y / Y / N	NA	Yes	At once	Yes	Controls: no lung conditions, no cardiopulmonary symptoms	
(115)	Y / Y / Y	No restrictions	Yes	At once	No	Cn: no cancer in 5 years, not requiring continuous supplemental oxygen, no under long-term immunosuppressive therapies. All had COPD or family lung cancer history	
(118)	Y / Y / N	NA	Yes	At once	No	All: comorbidities present Cs: therapy for breast cancer (N=1); previous cancer diagnosis (N=4), Cn: “High risk smokers”	
(130)	Y / Y / NA	2 h no food / drinks; VOCs related to food, smoking, cosmetics were excluded	NA	NA	No	NA	
(132)	N / Y / Y	3 min lung washout	NA	NA	No	Cn: heart disease or COPD present	
(134)	NA / NA / N	12 h no alcohol; 1 h no coffee; 5 min lung washout	NA	≤2 days	No	Cn: no restrictions, asthma (N=3), asthma+sinusitis (N=2)	
(133)	NA / NA / N[e]	12 h no coffee / alcohol; 3-5 min lung washout (morning)	NA	≤4 days	No	Cn: no restrictions	
(145)	Y / Y / N	no food / drinks (morning)	Yes	At once	No	Cn: no cancer	
(141)	A: N / Y / N B: N / Y / NA	NA	Yes	NA	No	B: no cancer, chronic disease	A: healthy B: abnormal chest x-rays
(137, 138)	Y / Y / N	NA	Yes	NA	No	Cn: no cancer	
(147)	A: NA / NA / N B: NA / NA / N C: N / N / N	NA	NA	≤2 h	No	Cn: no pulmonary disease, no previous cancer	A: cases smokers B: cases non- and ex-smokers C = A + B

(Continues on the next page)

(Continued)

ref	Sex / age / smoking[a]	Restrictions before sampling (collection time)	Room air[b]	Analysis time after collection	Treat-ment[c]	Exclusion/inclusion criteria	Comments
(148)	A: N / Y / Y B: N / N / Y	VOCs related to food and drugs were checked	NA	≤1 week	No	A: no chronic/acute pulmonary disease; no changes on chest roentgenogram; no industrial dust exposure; no drugs. B - no diseases, no drugs	A: controls (age-matched) B: controls (younger)
(146)	A: N / N / Y B: Y / N / N C: Y / Y / Y	rest 1 h; B: 1 h no smoking	Yes	~30min	No	All controls: no chronic bronchitis A: no pulmonary symptoms, no history of pulmonary diseases, abnormal lung spirometry results	A: controls non-smokers B: controls smokers C: COPD
(142, 143)	A: Y / Y / N B: Y / Y / NA	no food / drinks (morning)	Yes	NA	No	A, B: no cancer A: no chronic disease	A: healthy B abnormal mammogram
(144)	Y / Y / Y	no restrictions	Yes	NA	No	Cs: no therapy Controls: no cancer	
(140)	Y / NA / NA	NA	Yes	At once	No	Cn: no cancer, palpable breast mass present	
(151)	N / Y / Y	no food/drinks (morning)	Yes	At once	No	Cs: no therapy Controls: no chronic diseases	
(154)	NA / NA / NA	2 h no food / smoking / gum chewing; rest 15 min (morning)	Yes	NA	NA	Cn: systemic inflammatory response syndrome present	
(157)	N / NA / N	NA	Yes	NA	NA	NA	
(158)	N / Y / NA	NA	Yes	3-4 h	NA	NA	
(160)	Y / Y / Y	NA	NA	At once	Yes	Cn: benign conditions; under therapy (N=2)	
(167)	N / Y / N	2 h no food / smoking / alcohol; 3 min lung washout (morning)	NA	≤3 months	NA	Cn: no exclusions; gastric ulcer or other gastric diseases present	
(169)	A: Y / NA na / Y B: Y / NA / Y	NA	NA	NA	No	Cn: no exclusions, all volunteers	A: healthy B: precursors
(175)	Y / N / Y	no food / drinks (morning)	Yes	≤6 h	No	NA	
(179)	NA / Y / Y	no restrictions	Yes	~54 h	No	NA	
(186)	N / NA / Y	NA	Yes	At once	NA	Cn: no restrictions	
(189)	N / Y / N	8 h no food / drinks (morning)	Yes	≤3 h	No	Cs: no pregnancy, lactation; congenital disease; family history of mental illness; no current chronic inflammatory disease; symptoms of an acute disease in 2 weeks; no history of infectious disease. Cn: no history of cancer, no infectious disease.	

(Continues on the next page)

(Continued)

ref	Sex / age / smoking[a]	Restrictions before sampling (collection time)	Room air[b]	Analysis time after collection	Treat-ment[c]	Exclusion/inclusion criteria	Comments
(190)	A: Y / Y / Y B: Y / N / Y C: Y / N / Y	no food / drinks (morning)	Yes	≤3 h	No	A: not currently breast feeding or pregnant, no congenital disease, cancer, COPD, asthma, tuberculosis, other pulmonary diseases; chronic inflammatory or infectious disease; no manifestation of any acute disease symptoms during 2 weeks	A: healthy B: cyclo-mastopathy C: mammary gland fibroma
(191)	N / NA / N	no high-fat dinner the night before; 12 h no food / smoking; gargle water before breath sampling; (morning)	NA	At once	No	Cn: no respiratory symptoms, no diabetes, bronchitis, peptic ulcer, oxy hepatitis or coronary heart disease	
(192)	A : N / N / N B: Y / Y / N C: N / N / N	12 h no food / smoking; stay 30 min in ventilated room (morning)	Yes	NA	No	NA	A: healthy B: benign C = A+B
(195)	N / N / N	NA	Yes	≤12 h	No	NA	
(198)	Y / N / Y	NA	Yes	At once	No	Cn: no cancer, no disease	
(201)	NA / NA / NA	NA	NA	NA	NA	NA	
(202)	A: Y / Y / Y B: N / Y / N C: N / Y / N	24 h no heavy physical activity; 12 h no food / smoking / alcohol; 1 h rest; lung washout (morning)	Yes	≤4 months	No	A, B: no medication affecting gastric acid secretion and/or antibiotics in 1 month	A: gastric ulcer B: less severe conditions C = A+B
(205)	NA / NA / N	no food; stay 10 min in ventilated room (morning)	NA	At once	NA	NA	
(209)	A: Y / Y / Y B: NA / NA / NA	12 h no food; rinse out mouth with distilled water (morning)	Yes	At once	No	NA	A: training B: validation

COPD, chronic obstructive pulmonary disease N, no; NA, data not provided; Y, yes

[a] differences in proportions of age, sex and smoking distributions between cases and controls (“yes” if difference ≤10 units),

[b] indication whether room air was analyzed

[c] indication whether cancer patients were undergoing the treatment.

[d] as stated by authors.

[e] comparison of controls to lung cancer cases.

Appendix 5. Breath analysis technique, analyzed breath part, breath collection system or storage container and classifier.

Analysis technique	Analyzed breath	Storage container	Statistical analysis method
Studies which used electronic nose (N=19)			
Cyranose 320	Vital capacity	Mylar bag	DFA (118)
Cyranose 320	NA	Mylar bag	SVM (111)
Cyranose 320	Vital capacity	Rapak bag	Linear canonical DA (39)
Cyranose 320	Vital capacity	Tedlar bag	Linear canonical DA (53); canonical DA (54); Multinomial linear RA (81)
Colorimetric sensors	Tidal	No storing of samples	Multinomial linear RA (115); Random forest (114)
DiagNose	Whole	Tedlar bag	Multinomial linear RA (105)
LibraNose	Vital capacity	Sterile disposable bag	Partial least square DA (51)
MOS sensors	Alveolar	Tedlar bag	PCA (205)
MOS-SAW sensors	NA	Tedlar bag	Artificial NN (191)
NANOSE	Alveolar	Mylar bag	DFA (36); SVM (169)
QMS sensors	Alveolar	Tedlar bag	Partial least square DA (49, 160)
SAW sensors	No info	Tedlar bag	ANN (40)
TPS-SiNW FET sensors	Alveolar	Tenax sorption tubes	DFA (167)
Studies which used electronic noses and gas chromatography-mass spectrometry (N=6)			
NANOSE, GC-MS	Alveolar	Tenax sorption tubes	DFA (8, 67, 202)
NANOSE, SPME/GC-MS	Alveolar	Mylar bag	SVM (71)
NANOSE, SPME/GC-MS	Alveolar	Tenax sorption tubes	DFA (132)
NANOSE, TD-GC-MS	Alveolar	Tenax sorption tubes	DFA (7)
Studies which used gas chromatography-mass spectrometry (N=36)			
SPME/GC-MS	Alveolar	bio-VOC breath sampler	DA (147)
SPME/GC-MS	Alveolar	Glass vials	No prediction model[a] (68); Partial least square DA (189, 190)
SPME/GC-MS	Alveolar	Mylar bag	No prediction model[a] (133, 134)
SPME/GC-MS	Alveolar	Sealed headspace vial	No prediction model[a] (62, 89)
SPME/GC-MS	Alveolar	Tedlar bag	DA (186); Linear DA (192); Fisher's DA (107); Artificial NN (158); No prediction model[a] (38, 209)
SPME/GC-MS	Mixed alveolar	Tedlar bag	No prediction model[a] (16) (108)
SPME/GC-MS	NA	Tedlar bag	No prediction model[a] (63)
SPME/GC-MS	Vital capacity	Tedlar bag	No prediction model[a] (175)
SPME/GC-TOF-MS	NA	Tedlar bag	DFA (157)
SPME/TD-GC-MS	Alveolar	bio-VOC breath sampler	Multinomial linear RA (146)
SPME/TD-GC-MS	Vital capacity	Tedlar bag	Fisher's linear DA (151)
TD-GC-MS	Alveolar	Tedlar bag	No prediction model[a] (58);

(Continues on the next page)

(Continued)

Analysis technique	Analyzed breath	Storage container	Statistical analysis method
TD-GC-MS	Vital capacity	Tedlar bag	DFA (50); Probabilistic NN (4)
TD-GC-MS	Alveolar	Teflon sampling bag	Linear DA / Quadratic DA / SVM (130);
TD-GC-MS	Tidal	Teflon sampling bag	Linear DA (66)
TD-GC-MS	Alveolar	Tenax sorption tubes	No prediction modela (112, 148)
TD-GC-MS	Alveolar	Portable electrical device[b]	DA (141, 142, 145); Fuzzy logic (137, 143); Weighted digital analysis (138, 144)
TD-GC-MS	Alveolar	Tenax sorption tubes	No prediction model[a] (5)
Studies which used other techniques (N=13)			
Carbotrap/Carbosieve SIII-TD-GC	Alveolar	Vacu-sampler can	No prediction model[a] (77)
FT-ICR-MS	Vital capacity	Tedlar bag	Custom rule[c] (30, 61)
GC-SAW	Alveolar	Portable electrical device[b]	Weighted digital analysis (140)
MCC/IMS	Tidal[d]	No storing of samples	Decision Tree (72);
MCC/IMS	Alveolar	No storing of samples	Random forest (74); Lin. DA (198)
MSPE	Tidal	RTube collection system	No prediction model[a] (201)
PTR-MS	Vital capacity	No storing of samples	No prediction model[a] (154)
PTR-MS	Alveolar	Tedlar bag	Logistic RA (179);
PTR-MS	Tidal	Tedlar bag	Fisher's quadratic DA (195)
SIFT-MS	Mixed alveolar	Nalophan bag	Logistic RA (96)
SPME/GCxGC	Alveolar	Tedlar bag	No prediction model[a] (110)

D(F)A, discriminant (factor) analysis; FT-ICR, Fourier transform ion cyclotron resonance; GC, gas chromatography; MCC/IMS, multi capillary column/ion mobility spectrometry; MOS, metal oxide semiconductor; MS, mass spectrometry; MSPE, magnetic solid-phase extraction; NA, data not provided; NN, neural networks; PTR, proton transfer reaction; QMS, quartz microbalance; RA, regression analysis; SAW, surface acoustic wave; SIFT, selected ion flow tube; SPME, solid phase microextraction; SVM, support vector machine; TD, thermal desorption; TPS-SiNW-FET, trichloro-(phenethyl)silane-silicon nanowire-field effect transistor; TOF, time of flight.

[a] Statistics for two groups comparison provided only.

[b] A device designed specifically for collection and storage of exhaled breath (136).

[c] Levels of at least 2 out of 4 selected volatile organic compounds are elevated.

[d] dead space air excluded

Appendix 6. Volatile organic compounds reported by at least 3 independent studies on cancer detection through exhaled breath analysis.

Volatile organic compound	Studies that measured concentration	Studies that used compound for building a classification model
Ethenylbenzene	Gastric cancer (5); liver cancer (151); ovarian cancer (8)	Lung cancer (40, 134, 145, 146, 157); malignant mesothelioma (50)
Heptanal	Lung cancer (72, 201); breast cancer (107); gastric cancer (96)	Lung cancer (40, 145-147); breast cancer (143)
Nonanal	Lung cancer (62, 72, 201); breast cancer (107); ovarian cancer (8); gastric cancer (5, 96);	Lung cancer (147); colorectal cancer (4)
Propan-2-ol	Lung cancer (38, 58, 89, 158)	Lung cancer (138, 157, 195); breast cancer (143)
Hexanal	Lung cancer (58, 62, 72); breast cancer (107); gastric cancer (96)	Lung cancer (40, 145, 147)
Butan-2-one	Head and neck cancer (63); ovarian cancer (8); lung cancer (30, 38, 61)	Lung cancer (17, 66, 108)
2-methylbuta-1,3-diene	Gastric cancer (202); lung cancer (110)	Lung cancer (17, 40, 145, 146); breast cancer (144)
Propan-1-ol	Lung cancer(38, 89, 110)	Lung cancer (17, 66, 108, 137)
Dodecane	Lung cancer (58, 72); colorectal cancer (189)	Lung cancer (133); breast cancer (144); malignant mesothelioma (50)
Pentane	Lung cancer (110); breast cancer (77)	Lung cancer (108, 141, 146, 186)
Undecane	Head and neck cancer (67)	Lung cancer (17, 40, 145) breast cancer (144) malignant mesothelioma (50)
Decane	Liver cancer (151);	Lung cancer (40, 145, 146) malignant mesothelioma (50)
Octanal	Lung cancer (58, 62); breast cancer (107); gastric cancer (96)	Lung cancer (147)
Pentanal	Lung cancer (62, 201); gastric cancer (96)	Lung cancer (17, 147)
3-hydroxybutan-2-one	Lung cancer (30, 61, 175); liver cancer (151);	Lung cancer (17)
Toluene		Lung cancer (133, 134, 146); malignant mesothelioma (50); prostate cancer (133)
2,4-dimethylheptane	Lung cancer (158)	Lung cancer (134, 145); malignant mesothelioma (50)
2-methylpentane	Lung cancer (58)	Lung cancer (108, 146); colorectal cancer (4)
4-methyloctane		Lung cancer (134, 141, 186); colorectal cancer (4)
Benzene		Lung cancer (40, 145, 146, 186)
Cyclohexanone	Lung cancer (72); colorectal cancer (189); thyroid cancer (68)	Breast cancer (190)
Decanal	Lung cancer (58); gastric cancer (96); ovarian cancer (8)	Colorectal cancer (4)
Ethanol	Lung cancer (58); head and neck cancer(63, 67)	Lung cancer (134)
Ethylbenzene	Lung cancer (38, 72)	Lung cancer (146, 157)
2,2-dimethyldecane	Colorectal cancer (189); thyroid cancer (68)	Head and neck cancer (71); prostate cancer (133)
(methylsulfanyl)methane	Lung cancer (89, 158)	Lung cancer (186)
1,2,4-trimethylbenzene		Lung cancer (40, 145, 146)

(Continues on the next page)

(Continued)

1,4-xylene		**Lung cancer (145); colorectal cancer (4); prostate cancer (133)**
1-phenylethan-1-one		Lung cancer (17); breast cancer (143); malignant mesothelioma (50)
2,6,11-trimethyldodecane	Lung cancer (192, 209)	Breast cancer (144)
2,6-di-tert-butylcyclohexa-2,5-diene-1,4-dione	Lung cancer (192)	Lung cancer (137); breast cancer (144)
2-methylbutane		Lung cancer (17, 186); colorectal cancer (4)
3-methylhexane	Lung cancer (58); breast cancer (112)	Lung cancer (141)
3-methylpentane		Lung cancer (108, 186); colorectal cancer (4)
5-(2-methylpropyl)nonane	Lung cancer (192, 209)	Breast cancer (133)
6-methylhept-5-en-2-one	Lung cancer (58); gastric cancer (5, 202)	
benzaldehyde		Lung cancer (17, 108); malignant mesothelioma (50)
Butanal	Lung cancer (38); gastric cancer (96)	Lung cancer (147)
Cyclohexane		Lung cancer (145); colorectal cancer (4); malignant mesothelioma (50)
Methanol	Lung cancer (58, 110)	Lung cancer (17)
Methylcyclopentane		Lung cancer (40, 145); colorectal cancer (4)
Prop-2-enenitrile	Head and neck cancer (67); gastric cancer(7, 202)	
Propan-2-one	Lung cancer (110)	Lung cancer (17, 66)
Propanal	Lung cancer (58, 89)	Lung cancer (147)
Tridecane	Lung cancer (192, 209)	Breast cancer (144)

Appendix 7. Annual transition rates between preclinical to clinical colorectal and gastric cancer.

Cancer site	Gender	Age group	Transition rate		Mean sojourn time	
			Annual (%)	95% CI	Years	95% CI
Colorectal cancer[a]	Men	50-54[b]	18.1	(16.7, 19.5)	5.5	(5.1, 6.0)
		55–59	18.1	(16.7, 19.5)	5.5	(5.1, 6.0)
		60–64	19.2	(18.1, 20.3)	5.2	(4.9, 5.5)
		65–69	21.3	(20.3, 22.4)	4.7	(4.5, 4.9)
		70–74	20.6	(19.5, 21.7)	4.9	(4.6, 5.1)
Colorectal cancer[a]	Women	50-54[a]	21.3	(19.5, 23.4)	4.7	(4.3, 5.1)
		55–59	21.3	(19.5, 23.4)	4.7	(4.3, 5.1)
		60–64	22.5	(20.9, 24.2)	4.5	(4.1, 4.8)
		65–69	21.9	(20.6, 23.3)	4.6	(4.3, 4.8)
		70–74	20.8	(19.4, 22.2)	4.8	(4.5, 5.1)
Gastric cancer[c]	Men and women	50-59	31.44		3.18	
		60-69	26.72		3.74	
		70-74[d]	26.72		3.74	

CI, confidence intervals.

[a] derived from the German screening colonoscopy registry (32);

[b] not estimated, same as for the age group 55-59;

[c] derived from a two-rounds endoscopic screening in Korea (14);

[d] not estimated, same as for the age group 60-69.

Appendix 8. List of countries in each geographical region.

Continent	Region	Countries[a]
Europe	Northern Europe	Denmark, Estonia, Finland, Iceland, Ireland, Latvia, Lithuania, Norway, Sweden, the United Kingdom
	Western Europe	Austria, Belgium, France, Germany, Luxembourg, the Netherlands, Switzerland
	Southern Europe	Albania, Bosnia and Herzegovina, Croatia, Cyprus, Greece, Italy, Macedonia, Malta, Montenegro, Portugal, Serbia, Slovenia, Spain
	Central and Eastern Europe	Belarus, Bulgaria, Czech Republic, Hungary, Republic of Moldova, Poland, Romania, Russian Federation, Slovakia, Ukraine
Americas	Northern America	Canada, the United States of America
	Central America	Belize, Costa Rica, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama
	Southern America	Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, French Guiana, Guyana, Paraguay, Peru, Suriname, Uruguay, Venezuela
Asia	Western Asia	Armenia, Azerbaijan, Bahrain, State of Palestine, Georgia, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Turkey, United Arab Emirates, Yemen
	Southern-Central Asia	Afghanistan, Bangladesh, Bhutan, India, Islamic Republic of Iran, Kazakhstan, Kyrgyzstan, Maldives, Nepal, Pakistan, Sri Lanka, Tajikistan, Turkmenistan, Uzbekistan
	Southern-Eastern Asia	Brunei, Cambodia, Indonesia, Lao, Malaysia, Myanmar, Philippines, Singapore, Thailand, Timor-Leste, Viet Nam
	Eastern Asia	China, Japan, North Korea, Mongolia, South Korea
Africa	Africa	All countries in the continent
Australia	Australia	Australia

[a] as defined in GLOBOCAN 2012 database (57).

Appendix 9. Proportion of men and women in 5-year-old age groups in different regions and countries calculated from GLOBOCAN 12 data.

Country or region	Men					Women				
	50-54	55-59	60-64	65-69	70-74	50-54	55-59	60-64	65-69	70-74
Northern Europe	24.5	21.6	21.6	18.6	13.8	23.6	21.0	21.4	18.9	15.0
Denmark	23.2	20.9	22.1	19.9	13.9	22.4	20.5	21.9	20.2	15.0
Estonia	26.5	24.4	20.4	15.5	13.3	22.5	22.4	20.5	17.2	17.4
Finland	22.8	22.4	23.9	18.5	12.4	21.5	21.7	23.7	19.0	14.1
Iceland	27.7	24.8	21.1	15.7	10.8	27.4	24.3	20.5	16.0	11.7
Ireland	26.1	23.4	21.2	17.0	12.3	25.9	23.1	20.9	17.0	13.2
Latvia	27.4	24.2	18.5	15.9	13.9	23.0	21.8	18.6	18.2	18.5
Lithuania	29.9	23.8	18.2	15.1	12.9	25.4	21.7	18.5	17.4	17.0
Norway	24.7	22.2	23.2	18.5	11.4	23.8	21.7	22.9	18.8	12.8
Sweden	22.0	20.5	22.0	20.7	14.8	21.3	20.2	22.0	20.9	15.7
The United Kingdom	24.7	21.3	21.4	18.5	14.1	24.0	20.9	21.3	18.8	15.0
Western Europe	25.5	22.5	20.5	16.7	14.8	24.3	22.0	20.3	17.0	16.4
Austria	27.1	21.7	18.8	17.8	14.6	25.3	21.0	18.7	18.6	16.4
Belgium	25.7	23.4	21.4	16.4	13.1	24.6	22.6	21.1	16.9	14.8
France	24.4	23.2	22.9	17.1	12.5	23.7	22.9	22.6	17.1	13.6
Germany	26.1	22.0	18.6	16.1	17.3	24.4	21.4	18.5	16.5	19.2
Luxembourg	28.5	24.1	20.1	15.3	12.0	27.3	23.4	19.6	15.8	13.9
The Netherlands	25.1	22.5	22.1	17.8	12.4	24.5	22.2	21.8	17.9	13.5
Switzerland	25.9	22.3	21.1	17.7	13.0	24.6	21.6	20.9	18.2	14.7
Southern Europe	25.4	22.4	20.6	17.1	14.4	24.0	21.6	20.4	17.8	16.2
Albania	30.6	24.4	17.3	15.0	12.7	30.0	23.7	16.7	15.7	13.9
Bosnia and Herzegovina	27.7	25.8	20.1	13.6	12.8	26.5	24.6	20.1	14.5	14.4
Croatia	25.4	24.9	21.7	15.1	12.9	23.2	22.9	21.1	16.5	16.2
Cyprus	27.2	24.3	20.3	15.9	12.3	25.6	23.3	20.4	17.2	13.5
Greece	24.7	22.9	21.0	16.6	14.8	23.5	21.8	20.6	17.2	17.0
Italy	24.4	21.0	20.7	18.1	15.7	23.0	20.5	20.4	18.7	17.4
Malta	24.7	23.9	24.2	15.9	11.4	23.2	23.0	23.9	16.7	13.3
Macedonia	27.3	25.7	20.4	15.0	11.7	25.1	24.3	20.6	16.3	13.6
Montenegro	27.5	26.4	21.2	13.3	11.6	25.0	24.1	21.3	15.0	14.6
Portugal	25.1	22.7	20.5	17.2	14.6	23.4	21.7	20.4	18.1	16.4
Serbia	24.7	26.4	22.3	14.5	12.1	22.8	25.2	22.2	15.5	14.4
Slovenia	25.5	25.5	21.7	15.2	12.2	24.2	23.6	20.8	16.3	15.1
Spain	26.8	22.4	20.1	17.2	13.6	25.2	21.7	20.1	17.8	15.2
Central and Eastern Europe	28.8	26.8	20.8	12.6	11.1	25.0	24.9	20.9	14.1	15.1
Belarus	32.0	26.8	19.2	11.7	10.3	27.3	24.5	19.6	13.7	14.9
Bulgaria	23.9	23.2	22.3	17.7	12.9	21.2	21.7	22.5	19.2	15.5
Czech Republic	22.7	25.0	23.9	17.7	10.7	20.5	23.5	24.0	19.1	12.9
Hungary	23.2	25.7	22.0	16.8	12.3	20.2	23.8	21.8	18.4	15.8
Moldova	29.2	27.3	20.9	12.8	9.6	27.4	26.3	19.9	13.4	12.9
Poland	26.9	28.1	22.0	13.2	9.8	24.1	26.6	22.2	14.5	12.6

(Continues on the next page)

(Continued)

Country or region	Men					Women				
	50-54	55-59	60-64	65-69	70-74	50-54	55-59	60-64	65-69	70-74
Romania	24.0	26.2	21.5	15.1	13.2	21.5	24.7	21.5	16.3	16.0
Slovakia	27.7	27.2	21.3	14.2	9.6	24.4	25.3	21.3	16.1	12.8
Ukraine	28.7	25.4	19.6	12.9	13.3	24.5	23.3	19.7	14.8	17.7
Northern America	27.7	24.5	20.8	15.9	11.0	26.5	24.0	20.9	16.5	12.1
the U.S.	27.7	24.6	20.9	15.8	11.0	26.5	24.0	20.9	16.5	12.1
Central America	30.7	25.3	18.9	14.3	10.8	30.5	25.4	18.8	14.2	11.1
South America	30.8	25.1	19.6	14.2	10.3	29.5	24.6	19.6	14.9	11.4
Brazil	31.6	25.3	19.3	13.8	10.1	30.3	24.8	19.5	14.4	11.1
South-Central Asia	32.1	26.4	19.0	13.1	9.4	30.8	25.6	19.2	13.9	10.5
India	31.8	26.9	19.4	12.9	9.0	30.2	26.0	19.5	13.9	10.4
South-Eastern Asia	33.3	26.6	18.3	12.5	9.2	31.7	25.6	18.5	13.6	10.7
Eastern Asia	27.1	26.3	21.4	14.5	10.7	26.3	25.4	21.4	15.2	11.8
China	27.8	27.2	21.4	13.7	9.9	27.2	26.4	21.4	14.2	10.7
Japan	19.1	20.1	23.5	21.0	16.3	18.0	19.2	23.1	21.7	18.0
South Korea	31.7	25.0	18.2	14.1	11.0	29.8	24.0	18.0	14.9	13.3
Western Asia	33.3	26.7	18.9	12.2	8.9	30.8	25.8	19.3	13.4	10.7
Africa	31.4	25.3	19.6	14.2	9.6	30.1	24.9	19.8	14.8	10.5
Morocco	32.6	27.3	18.7	12.3	9.1	32.3	25.4	18.1	13.0	11.2
Australia	25.9	23.0	21.5	17.3	12.3	25.7	23.0	21.4	17.1	12.8

Appendix 10. Age-specific and overall age-adjusted prevalence of potentially detectable colorectal cancer cases in Europe.

Country or region	Men							Women						
	50-54	55-59	60-64	65-69	70-74	Overall[a]	Rank	50-54	55-59	60-64	65-69	70-74	Overall[a]	Rank
Northern Europe	0.26	0.49	0.77	1.07	1.51	0.74		0.18	0.30	0.43	0.66	0.94	0.46	
Denmark	0.34	0.64	0.98	1.38	1.94	0.97	6	0.27	0.44	0.63	0.95	1.33	0.68	1
Estonia	0.22	0.48	0.78	1.14	1.58	0.72	26	0.18	0.29	0.40	0.60	0.89	0.45	17
Finland	0.20	0.39	0.61	0.83	1.15	0.58	34	0.16	0.25	0.34	0.49	0.68	0.36	33
Iceland	0.15	0.39	0.74	0.90	1.23	0.57	35	0.26	0.39	0.49	0.64	0.93	0.48	14
Ireland	0.32	0.59	0.90	1.27	1.81	0.85	14	0.20	0.32	0.47	0.72	1.07	0.49	13
Latvia	0.18	0.40	0.66	0.94	1.34	0.60	32	0.14	0.26	0.39	0.60	0.82	0.42	24
Lithuania	0.20	0.40	0.68	0.99	1.43	0.61	30	0.14	0.25	0.37	0.54	0.74	0.38	31
Norway	0.28	0.55	0.88	1.27	1.83	0.84	17	0.24	0.42	0.62	0.95	1.36	0.64	2
Sweden	0.21	0.42	0.67	0.98	1.41	0.69	28	0.19	0.32	0.47	0.73	1.04	0.52	8
The United Kingdom	0.26	0.50	0.76	1.05	1.49	0.74	24	0.17	0.28	0.40	0.62	0.89	0.44	18
Western Europe	0.32	0.58	0.87	1.17	1.60	0.82		0.20	0.31	0.42	0.61	0.85	0.44	
Austria	0.26	0.49	0.75	1.01	1.39	0.70	27	0.16	0.25	0.34	0.48	0.67	0.36	34
Belgium	0.38	0.67	1.02	1.30	1.72	0.91	10	0.23	0.36	0.52	0.74	1.01	0.52	7
France	0.31	0.52	0.77	1.07	1.48	0.74	25	0.20	0.29	0.40	0.62	0.88	0.43	21
Germany	0.32	0.60	0.90	1.18	1.61	0.85	15	0.19	0.29	0.39	0.55	0.78	0.42	25
Luxembourg	0.29	0.48	1.16	1.10	2.05	0.84	16	0.18	0.24	0.34	0.47	0.67	0.34	36
The Netherlands	0.34	0.68	1.06	1.46	2.03	0.98	5	0.26	0.43	0.62	0.93	1.29	0.64	3
Switzerland	0.29	0.55	0.83	1.07	1.44	0.75	20	0.20	0.31	0.42	0.59	0.80	0.43	22
Southern Europe	0.35	0.62	0.92	1.16	1.53	0.84		0.22	0.33	0.44	0.60	0.79	0.44	
Albania	0.10	0.15	0.22	0.27	0.31	0.19	39	0.09	0.11	0.14	0.18	0.23	0.14	39
Bosnia and Herzegovina	0.20	0.34	0.52	0.58	0.63	0.41	37	0.15	0.19	0.22	0.35	0.43	0.24	37
Croatia	0.31	0.62	0.98	1.36	1.89	0.90	11	0.19	0.31	0.44	0.65	0.91	0.46	15
Cyprus	0.23	0.39	0.60	0.77	1.10	0.54	36	0.20	0.29	0.36	0.52	0.75	0.38	30
Greece	0.12	0.21	0.31	0.42	0.62	0.31	38	0.08	0.13	0.17	0.24	0.35	0.18	38
Cyprus	0.23	0.39	0.60	0.77	1.10	0.54	36	0.20	0.29	0.36	0.52	0.75	0.38	30
Greece	0.12	0.21	0.31	0.42	0.62	0.31	38	0.08	0.13	0.17	0.24	0.35	0.18	38
Italy	0.36	0.67	1.00	1.25	1.62	0.92	7	0.26	0.40	0.53	0.69	0.88	0.53	6
Malta	0.35	0.62	0.85	1.12	1.49	0.79	18	0.24	0.40	0.45	0.74	0.87	0.49	12
Macedonia	0.30	0.51	0.69	0.81	0.98	0.59	33	0.16	0.27	0.38	0.59	0.81	0.39	29
Montenegro	0.55	0.71	0.65	0.93	1.21	0.74	22	0.28	0.29	0.33	0.52	0.60	0.37	32
Portugal	0.38	0.68	0.97	1.22	1.59	0.89	12	0.22	0.32	0.42	0.57	0.76	0.43	19
Serbia	0.46	0.77	1.05	1.24	1.48	0.91	8	0.23	0.35	0.45	0.59	0.73	0.43	20
Slovenia	0.35	0.70	1.10	1.57	2.17	1.01	4	0.21	0.36	0.51	0.74	1.02	0.52	9
Spain	0.39	0.68	0.99	1.27	1.75	0.91	9	0.21	0.31	0.40	0.58	0.81	0.43	23
Central-Eastern Europe	0.27	0.53	0.83	1.12	1.42	0.69		0.19	0.31	0.43	0.60	0.76	0.42	
Belarus	0.25	0.50	0.76	0.99	1.28	0.61	31	0.17	0.29	0.42	0.62	0.79	0.40	26
Bulgaria	0.35	0.66	0.97	1.23	1.58	0.88	13	0.22	0.37	0.50	0.68	0.86	0.50	11

(Continues on the next page)

(Continued)

Country or region	Men							Women						
	50-54	55-59	60-64	65-69	70-74	Overall[a]	Rank	50-54	55-59	60-64	65-69	70-74	Overall[a]	Rank
Czech Republic	0.39	0.80	1.25	1.67	2.27	1.13	3	0.19	0.33	0.49	0.73	1.01	0.50	10
Hungary	0.57	1.02	1.45	1.73	2.17	1.27	1	0.27	0.42	0.57	0.78	1.03	0.58	4
Moldova	0.27	0.55	0.98	1.30	1.56	0.75	19	0.21	0.36	0.50	0.67	0.82	0.45	16
Poland	0.29	0.58	0.88	1.16	1.55	0.74	23	0.16	0.25	0.35	0.51	0.72	0.35	35
Romania	0.33	0.60	0.85	1.05	1.32	0.75	21	0.18	0.28	0.38	0.54	0.70	0.39	28
Slovakia	0.45	0.86	1.40	1.94	2.71	1.19	2	0.23	0.36	0.52	0.79	1.14	0.53	5
Ukraine	0.25	0.49	0.78	0.99	1.20	0.64	29	0.19	0.32	0.42	0.54	0.65	0.40	27

[a] Age-adjusted estimates (%) weighted by country-specific underlying population age structure.

Appendix 11. Age-specific and overall positive and negative predictive values for various screening tests for colorectal cancer detection in Australia.

Screening test (sensitivity [%] / specificity [%]) or actual test	Positive predictive value [%] / Negative predictive value [%]					
	50-54	55-59	60-64	65-69	70-74	Overall[a]
Men						
Prevalence (%)[a]	0.36	0.66	0.99	1.30	1.77	0.90
70 / 70	0.8 / 99.8	1.5 / 99.7	2.3 / 99.6	3.0 / 99.4	4.0 / 99.2	2.1 / 99.6
70 / 80	1.3 / 99.9	2.3 / 99.8	3.4 / 99.6	4.4 / 99.5	5.9 / 99.3	3.1 / 99.7
70 / 90	2.5 / 99.9	4.5 / 99.8	6.5 / 99.7	8.5 / 99.6	11.2 / 99.4	5.9 / 99.7
80 / 70	1.0 / 99.9	1.7 / 99.8	2.6 / 99.7	3.4 / 99.6	4.6 / 99.5	2.4 / 99.7
80 / 80	1.4 / 99.9	2.6 / 99.8	3.8 / 99.8	5.0 / 99.7	6.7 / 99.6	3.5 / 99.8
80 / 90	2.8 / 99.9	5.1 / 99.9	7.4 / 99.8	9.6 / 99.7	12.6 / 99.6	6.7 / 99.8
90 / 70[b]	1.1 / 99.9	2.0 / 99.9	2.9 / 99.9	3.8 / 99.8	5.1 / 99.7	2.6 / 99.9
90 / 80	1.6 / 100	2.9 / 99.9	4.3 / 99.9	5.6 / 99.8	7.5 / 99.8	3.9 / 99.9
90 / 90	3.2 / 100	5.7 / 99.9	8.3 / 99.9	10.6 / 99.9	14.0 / 99.8	7.5 / 99.9
95 / 95	6.4 / 100	11.2 / 100	16.0 / 99.9	20.1 / 99.9	25.5 / 99.9	14.3 / 100
Breast test 1[c]	4.9 / 99.9	8.6 / 99.9	12.4 / 99.8	15.8 / 99.8	20.3 / 99.7	11.2 / 99.9
FIT[d]	4.6 / 99.9	8.1 / 99.9	11.6 / 99.8	14.8 / 99.7	19.2 / 99.6	10.5 / 99.8
Cologuard[e]	2.4 / 100	4.4 / 99.9	6.4 / 99.9	8.3 / 99.9	11.0 / 99.8	5.8 / 99.9
Epi proColon[f]	2.0 / 99.8	3.6 / 99.6	5.4 / 99.4	7.0 / 99.3	9.3 / 99.0	4.9 / 99.5
Women						
Prevalence (%)[a]	0.23	0.38	0.54	0.82	1.16	0.55
70 / 70	0.5 / 99.9	0.9 / 99.8	1.3 / 99.8	1.9 / 99.6	2.7 / 99.5	1.3 / 99.8
70 / 80	0.8 / 99.9	1.3 / 99.9	1.9 / 99.8	2.8 / 99.7	4.0 / 99.6	1.9 / 99.8
70 / 90	1.6 / 99.9	2.6 / 99.9	3.7 / 99.8	5.5 / 99.7	7.6 / 99.6	3.7 / 99.8
80 / 70	0.6 / 99.9	1.0 / 99.9	1.4 / 99.8	2.2 / 99.8	3.0 / 99.7	1.5 / 99.8
80 / 80	0.9 / 99.9	1.5 / 99.9	2.1 / 99.9	3.2 / 99.8	4.5 / 99.7	2.2 / 99.9
80 / 90	1.8 / 99.9	2.9 / 99.9	4.2 / 99.9	6.2 / 99.8	8.6 / 99.7	4.2 / 99.9
90 / 70[b]	0.7 / 100	1.1 / 99.9	1.6 / 99.9	2.4 / 99.9	3.4 / 99.8	1.6 / 99.9
90 / 80	1.0 / 100	1.7 / 100	2.4 / 99.9	3.6 / 99.9	5.0 / 99.9	2.4 / 99.9
90 / 90	2.1 / 100	3.3 / 100	4.7 / 99.9	6.9 / 99.9	9.6 / 99.9	4.7 / 99.9
95 / 95	4.3 / 100	6.7 / 100	9.4 / 100	13.6 / 100	18.3 / 99.9	9.3 / 100
Breast test 1[c]	3.2 / 100	5.1 / 99.9	7.2 / 99.9	10.5 / 99.9	14.3 / 99.8	7.2 / 99.9
FIT[d]	3.0 / 99.9	4.8 / 99.9	6.7 / 99.9	9.8 / 99.8	13.4 / 99.7	6.7 / 99.9
Cologuard[e]	1.6 / 100	2.5 / 100	3.6 / 100	5.4 / 99.9	7.5 / 99.9	3.7 / 100
Epi proColon[f]	1.3 / 99.9	2.1 / 99.8	3.0 / 99.7	4.5 / 99.5	6.3 / 99.3	3.0 / 99.7

[a] Weighted sum of age-specific estimates using region-specific relative population size weights (weights can be found in Appendix 9).

[b] Breath test 2 by Altomare *et al.*, sensitivity 90%, specificity 70% (4).

[c] Breath test 1 by Amal *et al*, sensitivity 85%, specificity 94% (6).

[d] Fecal immunochemical test (FIT), sensitivity 79% and specificity 94% (102).

[e] Cologuard, sensitivity 92.3% and specificity 86.6% (82).

[f] Epi proColon, sensitivity 48.2% and specificity 91.5% (43).

Appendix 12. Age-specific and overall positive and negative predictive values for various screening tests for colorectal cancer detection in Germany.

Screening test (sensitivity [%] / specificity [%]) or actual test	Positive predictive value [%] / Negative predictive value [%]					
	50-54	55-59	60-64	65-69	70-74	Overall[a]
Men						
Prevalence (%)[a]	0.32	0.60	0.90	1.18	1.61	0.85
70 / 70	0.7 / 99.9	1.4 / 99.7	2.1 / 99.6	2.7 / 99.5	3.7 / 99.3	2.0 / 99.6
70 / 80	1.1 / 99.9	2.1 / 99.8	3.1 / 99.7	4.0 / 99.6	5.4 / 99.4	2.9 / 99.7
70 / 90	2.2 / 99.9	4.0 / 99.8	6.0 / 99.7	7.7 / 99.6	10.3 / 99.5	5.6 / 99.7
80 / 70	0.8 / 99.9	1.6 / 99.8	2.4 / 99.7	3.1 / 99.7	4.2 / 99.5	2.2 / 99.8
80 / 80	1.3 / 99.9	2.3 / 99.8	3.5 / 99.8	4.6 / 99.7	6.1 / 99.6	3.3 / 99.8
80 / 90	2.5 / 99.9	4.6 / 99.9	6.7 / 99.8	8.7 / 99.7	11.6 / 99.6	6.3 / 99.8
90 / 70[b]	0.9 / 100	1.8 / 99.9	2.6 / 99.9	3.5 / 99.8	4.7 / 99.8	2.5 / 99.9
90 / 80	1.4 / 100	2.6 / 99.9	3.9 / 99.9	5.1 / 99.9	6.9 / 99.8	3.7 / 99.9
90 / 90	2.8 / 100	5.1 / 99.9	7.5 / 99.9	9.7 / 99.9	12.8 / 99.8	7.0 / 99.9
95 / 95	5.7 / 100	10.3 / 100	14.7 / 100	18.5 / 99.9	23.7 / 99.9	13.5 / 100
Breast test 1[c]	4.3 / 99.9	7.9 / 99.9	11.4 / 99.9	14.5 / 99.8	18.8 / 99.7	10.5 / 99.9
FIT[d]	4.0 / 99.9	7.3 / 99.9	10.6 / 99.8	13.6 / 99.7	17.7 / 99.6	9.9 / 99.8
Cologuard[e]	2.2 / 100	4.0 / 99.9	5.9 / 99.9	7.6 / 99.9	10.1 / 99.9	5.5 / 99.9
Epi proColon[f]	1.8 / 99.8	3.3 / 99.7	4.9 / 99.5	6.3 / 99.3	8.5 / 99.1	4.6 / 99.5
Women						
Prevalence (%)[a]	0.19	0.29	0.39	0.55	0.78	0.42
70 / 70	0.4 / 99.9	0.7 / 99.9	0.9 / 99.8	1.3 / 99.8	1.8 / 99.7	1.0 / 99.8
70 / 80	0.7 / 99.9	1.0 / 99.9	1.3 / 99.9	1.9 / 99.8	2.7 / 99.7	1.4 / 99.8
70 / 90	1.3 / 99.9	2.0 / 99.9	2.7 / 99.9	3.7 / 99.8	5.2 / 99.7	2.8 / 99.9
80 / 70	0.5 / 99.9	0.8 / 99.9	1.0 / 99.9	1.4 / 99.8	2.0 / 99.8	1.1 / 99.9
80 / 80	0.8 / 100	1.1 / 99.9	1.5 / 99.9	2.2 / 99.9	3.0 / 99.8	1.6 / 99.9
80 / 90	1.5 / 100	2.2 / 99.9	3.0 / 99.9	4.2 / 99.9	5.9 / 99.8	3.2 / 99.9
90 / 70[b]	0.6 / 100	0.9 / 100	1.2 / 99.9	1.6 / 99.9	2.3 / 99.9	1.2 / 99.9
90 / 80	0.9 / 100	1.3 / 100	1.7 / 100	2.4 / 99.9	3.4 / 99.9	1.9 / 99.9
90 / 90	1.7 / 100	2.5 / 100	3.4 / 100	4.7 / 99.9	6.6 / 99.9	3.6 / 100
95 / 95	3.5 / 100	5.2 / 100	6.9 / 100	9.5 / 100	12.9 / 100	7.3 / 100
Breast test 1[c]	2.6 / 100	3.9 / 100	5.2 / 99.9	7.2 / 99.9	10.0 / 99.9	5.6 / 99.9
FIT[d]	2.5 / 100	3.6 / 99.9	4.9 / 99.9	6.7 / 99.9	9.3 / 99.8	5.2 / 99.9
Cologuard[e]	1.3 / 100	1.9 / 100	2.6 / 100	3.6 / 100	5.1 / 99.9	2.8 / 100
Epi proColon[f]	1.1 / 99.9	1.6 / 99.8	2.2 / 99.8	3.0 / 99.7	4.2 / 99.6	2.3 / 99.8

[a] Weighted sum of age-specific estimates using region-specific relative population size weights (weights can be found in Appendix 9).

[b] Breath test 2 by Altomare *et al.*, sensitivity 90%, specificity 70% (4).

[c] Breath test 1 by Amal *et al*, sensitivity 85%, specificity 94% (6).

[d] Fecal immunochemical test (FIT), sensitivity 79% and specificity 94% (102).

[e] Cologuard, sensitivity 92.3% and specificity 86.6% (82).

[f] Epi proColon, sensitivity 48.2% and specificity 91.5% (43).

Appendix 13. Age-specific and overall positive and negative predictive values for various screening tests for colorectal cancer detection in the U.S.

Screening test (sensitivity [%] / specificity [%]) or actual test	Positive predictive value [%] / Negative predictive value [%]					
	50-54	55-59	60-64	65-69	70-74	Overall[a]
Men						
Prevalence (%)[a]	0.31	0.45	0.59	0.70	0.92	0.53
70 / 70	0.7 / 99.9	1.0 / 99.8	1.4 / 99.7	1.6 / 99.7	2.1 / 99.6	1.2 / 99.8
70 / 80	1.1 / 99.9	1.6 / 99.8	2.0 / 99.8	2.4 / 99.7	3.1 / 99.7	1.8 / 99.8
70 / 90	2.1 / 99.9	3.1 / 99.8	4.0 / 99.8	4.7 / 99.8	6.1 / 99.7	3.6 / 99.8
80 / 70	0.8 / 99.9	1.2 / 99.9	1.5 / 99.8	1.9 / 99.8	2.4 / 99.7	1.4 / 99.8
80 / 80	1.2 / 99.9	1.8 / 99.9	2.3 / 99.9	2.8 / 99.8	3.6 / 99.8	2.1 / 99.9
80 / 90	2.4 / 99.9	3.5 / 99.9	4.5 / 99.9	5.4 / 99.8	6.9 / 99.8	4.1 / 99.9
90 / 70[b]	0.9 / 100	1.3 / 99.9	1.7 / 99.9	2.1 / 99.9	2.7 / 99.9	1.6 / 99.9
90 / 80	1.4 / 100	2.0 / 99.9	2.6 / 99.9	3.1 / 99.9	4.0 / 99.9	2.3 / 99.9
90 / 90	2.7 / 100	3.9 / 99.9	5.0 / 99.9	6.0 / 99.9	7.7 / 99.9	4.6 / 99.9
95 / 95	5.6 / 100	7.9 / 100	10.1 / 100	11.8 / 100	14.9 / 100	9.1 / 100
Breast test 1[c]	4.2 / 100	6.0 / 99.9	7.7 / 99.9	9.1 / 99.9	11.6 / 99.9	7.0 / 99.9
FIT[d]	3.9 / 99.9	5.6 / 99.9	7.2 / 99.9	8.5 / 99.8	10.8 / 99.8	6.5 / 99.9
Cologuard[e]	2.1 / 100	3.0 / 100	3.9 / 99.9	4.6 / 99.9	6.0 / 99.9	3.5 / 100
Epi proColon[f]	1.7 / 99.8	2.5 / 99.7	3.2 / 99.7	3.9 / 99.6	5.0 / 99.5	2.9 / 99.7
Women						
Prevalence (%)[a]	0.20	0.27	0.34	0.48	0.67	0.35
70 / 70	0.5 / 99.9	0.6 / 99.9	0.8 / 99.9	1.1 / 99.8	1.5 / 99.7	0.8 / 99.8
70 / 80	0.7 / 99.9	0.9 / 99.9	1.2 / 99.9	1.7 / 99.8	2.3 / 99.7	1.2 / 99.9
70 / 90	1.4 / 99.9	1.9 / 99.9	2.4 / 99.9	3.3 / 99.8	4.5 / 99.8	2.4 / 99.9
80 / 70	0.5 / 99.9	0.7 / 99.9	0.9 / 99.9	1.3 / 99.9	1.8 / 99.8	0.9 / 99.9
80 / 80	0.8 / 99.9	1.1 / 99.9	1.4 / 99.9	1.9 / 99.9	2.6 / 99.8	1.4 / 99.9
80 / 90	1.6 / 100	2.1 / 99.9	2.7 / 99.9	3.7 / 99.9	5.1 / 99.9	2.7 / 99.9
90 / 70[b]	0.6 / 100	0.8 / 100	1.0 / 100	1.4 / 99.9	2.0 / 99.9	1.0 / 99.9
90 / 80	0.9 / 100	1.2 / 100	1.5 / 100	2.1 / 99.9	2.9 / 99.9	1.6 / 100
90 / 90	1.8 / 100	2.4 / 100	3.0 / 100	4.2 / 99.9	5.7 / 99.9	3.1 / 100
95 / 95	3.7 / 100	4.9 / 100	6.1 / 100	8.5 / 100	11.3 / 100	6.2 / 100
Breast test 1[c]	2.8 / 100	3.7 / 100	4.7 / 99.9	6.5 / 99.9	8.7 / 99.9	4.7 / 99.9
FIT[d]	2.6 / 100	3.4 / 99.9	4.3 / 99.9	6.0 / 99.9	8.1 / 99.8	4.4 / 99.9
Cologuard[e]	1.4 / 100	1.8 / 100	2.3 / 100	3.2 / 100	4.4 / 99.9	2.4 / 100
Epi proColon[f]	1.1 / 99.9	1.5 / 99.8	1.9 / 99.8	2.7 / 99.7	3.7 / 99.6	1.9 / 99.8

[a] Weighted sum of age-specific estimates using region-specific relative population size weights (weights can be found in Appendix 9).

[b] Breath test 2 by Altomare *et al.*, sensitivity 90%, specificity 70% (4).

[c] Breath test 1 by Amal *et al*, sensitivity 85%, specificity 94% (6).

[d] Fecal immunochemical test (FIT), sensitivity 79% and specificity 94% (102).

[e] Cologuard, sensitivity 92.3% and specificity 86.6% (82).

[f] Epi proColon, sensitivity 48.2% and specificity 91.5% (43).

Appendix 14. Age-specific and overall positive and negative predictive values for various screening tests for colorectal cancer detection in Brazil.

Screening test (sensitivity [%] / specificity [%]) or actual test	Positive predictive value [%] / Negative predictive value [%]					
	50-54	55-59	60-64	65-69	70-74	Overall[a]
Men						
Prevalence (%)[a]	0.16	0.25	0.35	0.44	0.63	0.30
70 / 70	0.4 / 99.9	0.6 / 99.9	0.8 / 99.9	1.0 / 99.8	1.5 / 99.7	0.7 / 99.9
70 / 80	0.6 / 99.9	0.9 / 99.9	1.2 / 99.9	1.5 / 99.8	2.2 / 99.8	1.1 / 99.9
70 / 90	1.1 / 99.9	1.7 / 99.9	2.4 / 99.9	3.0 / 99.9	4.2 / 99.8	2.1 / 99.9
80 / 70	0.4 / 100	0.7 / 99.9	0.9 / 99.9	1.2 / 99.9	1.7 / 99.8	0.8 / 99.9
80 / 80	0.6 / 100	1.0 / 99.9	1.4 / 99.9	1.8 / 99.9	2.5 / 99.8	1.2 / 99.9
80 / 90	1.3 / 100	1.9 / 99.9	2.7 / 99.9	3.4 / 99.9	4.8 / 99.9	2.4 / 99.9
90 / 70[b]	0.5 / 100	0.7 / 100	1.0 / 100	1.3 / 99.9	1.9 / 99.9	0.9 / 100
90 / 80	0.7 / 100	1.1 / 100	1.5 / 100	2.0 / 99.9	2.8 / 99.9	1.3 / 100
90 / 90	1.4 / 100	2.2 / 100	3.0 / 100	3.9 / 100	5.4 / 99.9	2.7 / 100
95 / 95	2.9 / 100	4.5 / 100	6.2 / 100	7.8 / 100	10.7 / 100	5.4 / 100
Breast test 1[c]	2.2 / 100	3.4 / 100	4.7 / 99.9	6.0 / 99.9	8.2 / 99.9	4.1 / 100
FIT[d]	2.1 / 100	3.1 / 99.9	4.4 / 99.9	5.6 / 99.9	7.7 / 99.9	3.8 / 99.9
Cologuard[e]	1.1 / 100	1.7 / 100	2.3 / 100	3.0 / 100	4.2 / 99.9	2.0 / 100
Epi proColon[f]	0.9 / 99.9	1.4 / 99.9	1.9 / 99.8	2.5 / 99.7	3.5 / 99.6	1.7 / 99.8
Women						
Prevalence (%)[a]	0.12	0.17	0.22	0.35	0.53	0.23
70 / 70	0.3 / 99.9	0.4 / 99.9	0.5 / 99.9	0.8 / 99.9	1.2 / 99.8	0.5 / 99.9
70 / 80	0.4 / 100	0.6 / 99.9	0.8 / 99.9	1.2 / 99.9	1.8 / 99.8	0.8 / 99.9
70 / 90	0.9 / 100	1.2 / 99.9	1.6 / 99.9	2.4 / 99.9	3.6 / 99.8	1.6 / 99.9
80 / 70	0.3 / 100	0.4 / 100	0.6 / 99.9	0.9 / 99.9	1.4 / 99.8	0.6 / 99.9
80 / 80	0.5 / 100	0.7 / 100	0.9 / 99.9	1.4 / 99.9	2.1 / 99.9	0.9 / 99.9
80 / 90	1.0 / 100	1.3 / 100	1.8 / 99.9	2.7 / 99.9	4.1 / 99.9	1.8 / 99.9
90 / 70[b]	0.4 / 100	0.5 / 100	0.7 / 100	1.0 / 100	1.6 / 99.9	0.7 / 100
90 / 80	0.6 / 100	0.7 / 100	1.0 / 100	1.5 / 100	2.3 / 99.9	1.0 / 100
90 / 90	1.1 / 100	1.5 / 100	2.0 / 100	3.0 / 100	4.6 / 99.9	2.0 / 100
95 / 95	2.3 / 100	3.1 / 100	4.1 / 100	6.2 / 100	9.2 / 100	4.2 / 100
Breast test 1[c]	1.7 / 100	2.3 / 100	3.1 / 100	4.7 / 99.9	7.0 / 99.9	3.2 / 100
FIT[d]	1.6 / 100	2.2 / 100	2.9 / 99.9	4.4 / 99.9	6.6 / 99.9	2.9 / 99.9
Cologuard[e]	0.8 / 100	1.1 / 100	1.5 / 100	2.3 / 100	3.5 / 100	1.6 / 100
Epi proColon[f]	0.7 / 99.9	0.9 / 99.9	1.3 / 99.9	1.9 / 99.8	2.9 / 99.7	1.3 / 99.9

[a] Weighted sum of age-specific estimates using region-specific relative population size weights (weights can be found in Appendix 9).

[b] Breath test 2 by Altomare *et al.*, sensitivity 90%, specificity 70% (4).

[c] Breath test 1 by Amal *et al*, sensitivity 85%, specificity 94% (6).

[d] Fecal immunochemical test (FIT), sensitivity 79% and specificity 94% (102).

[e] Cologuard, sensitivity 92.3% and specificity 86.6% (82).

[f] Epi proColon, sensitivity 48.2% and specificity 91.5% (43).

Appendix 15. Age-specific and overall positive and negative predictive values for various screening tests for colorectal cancer detection in Morocco.

Screening test (sensitivity [%] / specificity [%]) or actual test	Positive predictive value [%] / Negative predictive value [%]					
	50-54	55-59	60-64	65-69	70-74	Overall[a]
Men						
Prevalence (%)[a]	0.11	0.18	0.25	0.27	0.3	0.19
70 / 70	0.3 / 100	0.4 / 99.9	0.6 / 99.9	0.6 / 99.9	0.7 / 99.9	0.5 / 99.9
70 / 80	0.4 / 100	0.6 / 99.9	0.9 / 99.9	1.0 / 99.9	1.0 / 99.9	0.7 / 99.9
70 / 90	0.8 / 100	1.2 / 99.9	1.7 / 99.9	1.9 / 99.9	2.0 / 99.9	1.3 / 99.9
80 / 70	0.3 / 100	0.5 / 99.9	0.7 / 99.9	0.7 / 99.9	0.8 / 99.9	0.5 / 99.9
80 / 80	0.4 / 100	0.7 / 100	1.0 / 99.9	1.1 / 99.9	1.2 / 99.9	0.8 / 100
80 / 90	0.9 / 100	1.4 / 100	2.0 / 99.9	2.2 / 99.9	2.3 / 99.9	1.5 / 100
90 / 70[b]	0.3 / 100	0.5 / 100	0.8 / 100	0.8 / 100	0.9 / 100	0.6 / 100
90 / 80	0.5 / 100	0.8 / 100	1.1 / 100	1.2 / 100	1.3 / 100	0.9 / 100
90 / 90	1.0 / 100	1.6 / 100	2.2 / 100	2.4 / 100	2.6 / 100	1.7 / 100
95 / 95	2.1 / 100	3.3 / 100	4.6 / 100	5.0 / 100	5.3 / 100	3.5 / 100
Breast test 1[c]	1.6 / 100	2.5 / 100	3.5 / 100	3.8 / 100	4.0 / 100	2.7 / 100
FIT[d]	1.5 / 100	2.3 / 100	3.2 / 99.9	3.5 / 99.9	3.8 / 99.9	2.5 / 100
Cologuard[e]	0.8 / 100	1.2 / 100	1.7 / 100	1.9 / 100	2.0 / 100	1.3 / 100
Epi proColon[f]	0.6 / 99.9	1.0 / 99.9	1.4 / 99.9	1.5 / 99.8	1.7 / 99.8	1.1 / 99.9
Women						
Prevalence (%)[a]	0.09	0.13	0.14	0.16	0.17	0.13
70 / 70	0.2 / 100	0.3 / 99.9	0.3 / 99.9	0.4 / 99.9	0.4 / 99.9	0.3 / 99.9
70 / 80	0.3 / 100	0.4 / 100	0.5 / 99.9	0.6 / 99.9	0.6 / 99.9	0.4 / 100
70 / 90	0.6 / 100	0.9 / 100	1.0 / 100	1.1 / 99.9	1.2 / 99.9	0.9 / 100
80 / 70	0.2 / 100	0.3 / 100	0.4 / 100	0.4 / 100	0.5 / 100	0.3 / 100
80 / 80	0.4 / 100	0.5 / 100	0.6 / 100	0.6 / 100	0.7 / 100	0.5 / 100
80 / 90	0.7 / 100	1.0 / 100	1.1 / 100	1.3 / 100	1.3 / 100	1.0 / 100
90 / 70[b]	0.3 / 100	0.4 / 100	0.4 / 100	0.5 / 100	0.5 / 100	0.4 / 100
90 / 80	0.4 / 100	0.6 / 100	0.6 / 100	0.7 / 100	0.8 / 100	0.6 / 100
90 / 90	0.8 / 100	1.1 / 100	1.3 / 100	1.4 / 100	1.5 / 100	1.1 / 100
95 / 95	1.7 / 100	2.4 / 100	2.6 / 100	3.0 / 100	3.1 / 100	2.4 / 100
Breast test 1[c]	1.3 / 100	1.8 / 100	2.0 / 100	2.2 / 100	2.3 / 100	1.8 / 100
FIT[d]	1.2 / 100	1.6 / 100	1.8 / 100	2.1 / 100	2.2 / 100	1.7 / 100
Cologuard[e]	0.6 / 100	0.9 / 100	1.0 / 100	1.1 / 100	1.2 / 100	0.9 / 100
Epi proColon[f]	0.5 / 99.9	0.7 / 99.9	0.8 / 99.9	0.9 / 99.9	1.0 / 99.9	0.7 / 99.9

[a] Weighted sum of age-specific estimates using region-specific relative population size weights (weights can be found in Appendix 9).

[b] Breath test 2 by Altomare *et al.*, sensitivity 90%, specificity 70% (4).

[c] Breath test 1 by Amal *et al*, sensitivity 85%, specificity 94% (6).

[d] Fecal immunochemical test (FIT), sensitivity 79% and specificity 94% (102).

[e] Cologuard, sensitivity 92.3% and specificity 86.6% (82).

[f] Epi proColon, sensitivity 48.2% and specificity 91.5% (43).

Appendix 16. Age-specific and overall positive and negative predictive values for various screening tests for colorectal cancer detection in India.

Screening test (sensitivity [%] / specificity [%]) or actual test	Positive predictive value [%] / Negative predictive value [%]					
	50-54	55-59	60-64	65-69	70-74	Overall[a]
Men						
Prevalence (%)[a]	0.08	0.12	0.15	0.18	0.23	0.13
70 / 70	0.2 / 100	0.3 / 99.9	0.4 / 99.9	0.4 / 99.9	0.5 / 99.9	0.3 / 99.9
70 / 80	0.3 / 100	0.4 / 100	0.5 / 99.9	0.6 / 99.9	0.8 / 99.9	0.5 / 100
70 / 90	0.5 / 100	0.8 / 100	1.0 / 99.9	1.2 / 99.9	1.6 / 99.9	0.9 / 100
80 / 70	0.2 / 100	0.3 / 100	0.4 / 100	0.5 / 99.9	0.6 / 99.9	0.3 / 100
80 / 80	0.3 / 100	0.5 / 100	0.6 / 100	0.7 / 100	0.9 / 99.9	0.5 / 100
80 / 90	0.6 / 100	0.9 / 100	1.2 / 100	1.4 / 100	1.8 / 99.9	1.0 / 100
90 / 70[b]	0.2 / 100	0.4 / 100	0.5 / 100	0.5 / 100	0.7 / 100	0.4 / 100
90 / 80	0.4 / 100	0.5 / 100	0.7 / 100	0.8 / 100	1.0 / 100	0.6 / 100
90 / 90	0.7 / 100	1.0 / 100	1.3 / 100	1.6 / 100	2.0 / 100	1.1 / 100
95 / 95	1.5 / 100	2.2 / 100	2.8 / 100	3.2 / 100	4.2 / 100	2.4 / 100
Breast test 1[c]	1.1 / 100	1.6 / 100	2.1 / 100	2.4 / 100	3.1 / 100	1.8 / 100
FIT[d]	1.0 / 100	1.5 / 100	1.9 / 100	2.3 / 100	2.9 / 99.9	1.7 / 100
Cologuard[e]	0.5 / 100	0.8 / 100	1.0 / 100	1.2 / 100	1.6 / 100	0.9 / 100
Epi proColon[f]	0.4 / 100	0.7 / 99.9	0.8 / 99.9	1.0 / 99.9	1.3 / 99.9	0.7 / 99.9
Women						
Prevalence (%)[a]	0.05	0.07	0.08	0.11	0.14	0.08
70 / 70	0.1 / 100	0.2 / 100	0.2 / 100	0.2 / 100	0.3 / 99.9	0.2 / 100
70 / 80	0.2 / 100	0.2 / 100	0.3 / 100	0.4 / 100	0.5 / 99.9	0.3 / 100
70 / 90	0.4 / 100	0.5 / 100	0.6 / 100	0.7 / 100	0.9 / 100	0.5 / 100
80 / 70	0.1 / 100	0.2 / 100	0.2 / 100	0.3 / 100	0.4 / 100	0.2 / 100
80 / 80	0.2 / 100	0.3 / 100	0.3 / 100	0.4 / 100	0.5 / 100	0.3 / 100
80 / 90	0.4 / 100	0.6 / 100	0.7 / 100	0.8 / 100	1.1 / 100	0.6 / 100
90 / 70[b]	0.2 / 100	0.2 / 100	0.2 / 100	0.3 / 100	0.4 / 100	0.2 / 100
90 / 80	0.2 / 100	0.3 / 100	0.4 / 100	0.5 / 100	0.6 / 100	0.4 / 100
90 / 90	0.5 / 100	0.6 / 100	0.7 / 100	0.9 / 100	1.2 / 100	0.7 / 100
95 / 95	1.0 / 100	1.3 / 100	1.5 / 100	2.0 / 100	2.5 / 100	1.5 / 100
Breast test 1[c]	0.7 / 100	1.0 / 100	1.2 / 100	1.5 / 100	1.9 / 100	1.1 / 100
FIT[d]	0.7 / 100	0.9 / 100	1.1 / 100	1.4 / 100	1.8 / 100	1.0 / 100
Cologuard[e]	0.4 / 100	0.5 / 100	0.6 / 100	0.7 / 100	0.9 / 100	0.5 / 100
Epi proColon[f]	0.3 / 100	0.4 / 100	0.5 / 100	0.6 / 99.9	0.8 / 99.9	0.4 / 100

[a] Weighted sum of age-specific estimates using region-specific relative population size weights (weights can be found in Appendix 9).

[b] Breath test 2 by Altomare *et al.*, sensitivity 90%, specificity 70% (4).

[c] Breath test 1 by Amal *et al*, sensitivity 85%, specificity 94% (6).

[d] Fecal immunochemical test (FIT), sensitivity 79% and specificity 94% (102).

[e] Cologuard, sensitivity 92.3% and specificity 86.6% (82).

[f] Epi proColon, sensitivity 48.2% and specificity 91.5% (43).

Appendix 17. Study description of the studies included in the meta-analysis for the estimation of breath tests performance for gastric cancer detection.

	Studies included in meta-analysis			
First author, year	Amal, 2016	Kumar, 2015	Shehada, 2015	Xu, 2013
Reference	(7)	(96)	(167)	(202)
Country	Latvia	the United Kingdom	Latvia	China
Cancer group	gastric cancer	gastric cancer + oesophageal cancer	gastric cancer	gastric cancer
Controls	OLGIM 0–IV	HC+ Barrett's metaplasia + benign conditions	HC + early intestinal metaplasia, gastric ulcer	HC + early intestinal metaplasia, gastric ulcer
Method	sensor arrays: gold nanoparticles and single-wall carbon nanotubes	selected ion flow tube mass spectrometry	sensor arrays: silicon nanowires	sensor arrays: nanomaterial-based sensor
True positives	22	20	5[a]	33
False negatives	8	3	2[a]	4
False positives	2	9	2[a]	9
True negatives	93	39	17[a]	84
Sensitivity [95% CI]	73% [56-86]	87% [68-95]	71% [36-92]	89% [75-96]
Specificity [95% CI]	98% [93-99]	81% [68-90]	89% [69-97]	90% [83-95]

CI, confidence intervals, HC, healthy controls; OLGIM, operative link on intestinal metaplasia assessment; TN

[a] values calculated from sensitivity and specificity.

Appendix 18. Negative predictive value of breath tests for gastric cancer detection, by age and sex.

Population	NPV for men (%)				NPV for women (%)			
	50-59	60-69	70-74	Overall[a]	50-59	60-69	70-74	Overall[a]
Northern Europe	99.99	99.98	99.96	99.98	100	99.99	99.98	99.99
Western Europe	99.99	99.97	99.95	99.98	99.99	99.99	99.98	99.99
Southern Europe	99.99	99.96	99.93	99.97	99.99	99.98	99.97	99.99
Central and Eastern Europe	99.97	99.93	99.89	99.95	99.99	99.97	99.95	99.98
Northern America	99.99	99.98	99.97	99.99	100	99.99	99.99	99.99
Central America	99.99	99.97	99.94	99.98	99.99	99.98	99.96	99.98
South America	99.98	99.95	99.92	99.97	99.99	99.98	99.96	99.98
Western Asia	99.98	99.96	99.94	99.97	99.99	99.98	99.96	99.98
South-Central Asia	99.99	99.97	99.96	99.98	99.99	99.99	99.98	99.99
South-Eastern Asia	99.99	99.97	99.96	99.98	99.99	99.99	99.98	99.99
Eastern Asia	99.96	99.88	99.81	99.91	99.98	99.96	99.93	99.97
Africa	99.99	99.99	99.98	99.99	100	99.99	99.99	99.99
Australia	99.99	99.98	99.96	99.98	100	99.99	99.98	99.99
Korea	99.91	99.78	99.70	99.84	99.97	99.93	99.90	99.94
Japan	99.95	99.83	99.71	99.86	99.98	99.95	99.92	99.96
China	99.96	99.89	99.83	99.92	99.99	99.96	99.94	99.97
Albania	99.97	99.92	99.87	99.94	99.98	99.95	99.92	99.96
Belarus	99.96	99.89	99.85	99.93	99.98	99.96	99.94	99.97
India	99.99	99.97	99.96	99.98	99.99	99.99	99.99	99.99
The United Kingdom	99.99	99.98	99.96	99.98	100	99.99	99.98	99.99
The U.S.	99.99	99.98	99.97	99.99	100	99.99	99.99	99.99

NPV, negative predictive value.

[a] Weighted sum of age-specific estimates using region- and country-specific relative population size weights (weights can be found in Appendix 9).

Appendix 19. Age-specific predictive values for tests with various levels of sensitivity and specificity, by country and sex.

Country	PPV / NPV[a] in men (%)				PPV / NPV[a] in women (%)			
	50-59	60-69	70-74	Overall[b]	50-59	60-69	70-74	Overall[b]
70% sensitivity and 70% specificity								
Korea	1.2 / 99.8	2.7 / 99.5	3.6 / 99.3	1.9 / 99.6	0.4 / 99.9	0.9 / 99.8	1.3 / 99.8	0.7 / 99.9
Japan	0.6 / 99.9	2.1 / 99.6	3.5 / 99.3	1.8 / 99.7	0.3 / 100	0.6 / 99.9	1.0 / 99.8	0.5 / 99.9
China	0.5 / 99.9	1.3 / 99.8	2.0 / 99.6	0.9 / 99.8	0.2 / 100	0.5 / 99.9	0.8 / 99.9	0.4 / 99.9
Albania	0.4 / 99.9	1.0 / 99.8	1.6 / 99.7	0.8 / 99.9	0.2 / 100	0.6 / 99.9	1.0 / 99.8	0.5 / 99.9
Belarus	0.5 / 99.9	1.3 / 99.8	1.9 / 99.7	0.9 / 99.8	0.2 / 100	0.5 / 99.9	0.7 / 99.9	0.4 / 99.9
India	0.2 / 100	0.4 / 99.9	0.5 / 99.9	0.3 / 100	0.1 / 100	0.1 / 100	0.2 / 100	0.1 / 100
The UK	0.1 / 100	0.3 / 100	0.5 / 99.9	0.2 / 100	0.0 / 100	0.1 / 100	0.2 / 100	0.1 / 100
The U.S.	0.1 / 100	0.2 / 100	0.4 / 99.9	0.2 / 100	0.0 / 100	0.1 / 100	0.2 / 100	0.1 / 100
70% sensitivity and 80% specificity								
Korea	1.7 / 99.8	4.0 / 99.6	5.4 / 99.4	2.9 / 99.7	0.6 / 99.9	1.4 / 99.9	1.9 / 99.8	1.0 / 99.9
Japan	0.9 / 99.9	3.1 / 99.7	5.2 / 99.4	2.6 / 99.7	0.4 / 100	0.9 / 99.9	1.5 / 99.8	0.8 / 99.9
China	0.8 / 99.9	1.9 / 99.8	3.0 / 99.7	1.4 / 99.8	0.3 / 100	0.7 / 99.9	1.2 / 99.9	0.5 / 99.9
Albania	0.6 / 99.9	1.4 / 99.8	2.4 / 99.7	1.1 / 99.9	0.3 / 100	0.9 / 99.9	1.5 / 99.8	0.7 / 99.9
Belarus	0.8 / 99.9	2.0 / 99.8	2.8 / 99.7	1.4 / 99.9	0.3 / 100	0.8 / 99.9	1.1 / 99.9	0.6 / 99.9
India	0.2 / 100	0.6 / 99.9	0.7 / 99.9	0.4 / 100	0.1 / 100	0.2 / 100	0.3 / 100	0.2 / 100
The UK	0.1 / 100	0.4 / 100	0.8 / 99.9	0.3 / 100	0.1 / 100	0.2 / 100	0.3 / 100	0.1 / 100
The U.S.	0.1 / 100	0.3 / 100	0.5 / 99.9	0.2 / 100	0.1 / 100	0.1 / 100	0.2 / 100	0.1 / 100
70% sensitivity and 90% specificity								
Korea	3.4 / 99.8	7.8 / 99.6	10.2 / 99.5	5.5 / 99.7	1.2 / 99.9	2.7 / 99.9	3.7 / 99.8	2.0 / 99.9
Japan	1.9 / 99.9	6.0 / 99.7	9.9 / 99.5	5.0 / 99.7	0.8 / 100	1.7 / 99.9	3.0 / 99.9	1.6 / 99.9
China	1.6 / 99.9	3.8 / 99.8	5.9 / 99.7	2.8 / 99.9	0.6 / 100	1.4 / 99.9	2.3 / 99.9	1.0 / 99.9
Albania	1.3 / 99.9	2.8 / 99.9	4.7 / 99.8	2.2 / 99.9	0.7 / 100	1.9 / 99.9	2.9 / 99.9	1.4 / 99.9
Belarus	1.6 / 99.9	3.9 / 99.8	5.4 / 99.7	2.7 / 99.9	0.6 / 100	1.5 / 99.9	2.2 / 99.9	1.1 / 99.9
India	0.5 / 100	1.1 / 99.9	1.4 / 99.9	0.8 / 100	0.2 / 100	0.4 / 100	0.5 / 100	0.3 / 100
The UK	0.2 / 100	0.8 / 100	1.5 / 99.9	0.6 / 100	0.1 / 100	0.3 / 100	0.6 / 100	0.3 / 100
The U.S.	0.2 / 100	0.7 / 100	1.0 / 99.9	0.5 / 100	0.1 / 100	0.3 / 100	0.5 / 100	0.2 / 100
80% sensitivity and 70% specificity								
Korea	1.3 / 99.9	3.1 / 99.7	4.1 / 99.5	2.2 / 99.8	0.5 / 100	1.0 / 99.9	1.5 / 99.8	0.8 / 99.9
Japan	0.7 / 99.9	2.4 / 99.7	4.0 / 99.6	2.0 / 99.8	0.3 / 100	0.7 / 99.9	1.2 / 99.9	0.6 / 99.9
China	0.6 / 99.9	1.5 / 99.8	2.3 / 99.7	1.1 / 99.9	0.2 / 100	0.5 / 99.9	0.9 / 99.9	0.4 / 100
Albania	0.5 / 99.9	1.1 / 99.9	1.8 / 99.8	0.9 / 99.9	0.3 / 100	0.7 / 99.9	1.1 / 99.9	0.5 / 99.9
Belarus	0.6 / 99.9	1.5 / 99.8	2.1 / 99.8	1.0 / 99.9	0.2 / 100	0.6 / 99.9	0.9 / 99.9	0.4 / 100
India	0.2 / 100	0.4 / 100	0.5 / 99.9	0.3 / 100	0.1 / 100	0.2 / 100	0.2 / 100	0.1 / 100
The UK	0.1 / 100	0.3 / 100	0.6 / 99.9	0.2 / 100	0.0 / 100	0.1 / 100	0.2 / 100	0.1 / 100
The U.S.	0.1 / 100	0.3 / 100	0.4 / 100	0.2 / 100	0.0 / 100	0.1 / 100	0.2 / 100	0.1 / 100
80% sensitivity and 80% specificity								
Korea	2.0 / 99.9	4.6 / 99.7	6.1 / 99.6	3.3 / 99.8	0.7 / 100	1.5 / 99.9	2.2 / 99.9	1.2 / 99.9
Japan	1.1 / 99.9	3.5 / 99.8	5.9 / 99.6	3.0 / 99.8	0.5 / 100	1.0 / 99.9	1.8 / 99.9	0.9 / 99.9
China	0.9 / 99.9	2.2 / 99.9	3.4 / 99.8	1.6 / 99.9	0.3 / 100	0.8 / 99.9	1.4 / 99.9	0.6 / 100
Albania	0.7 / 100	1.6 / 99.9	2.7 / 99.8	1.3 / 99.9	0.4 / 100	1.1 / 99.9	1.7 / 99.9	0.8 / 100

(Continues on the next page)

(Continued)

Country	PPV / NPV[a] in men				PPV / NPV[a] in women			
	50-59	60-69	70-74	Overall[b]	50-59	60-69	70-74	Overall[b]
Belarus	0.9 / 99.9	2.3 / 99.9	3.2 / 99.8	1.6 / 99.9	0.3 / 100	0.9 / 99.9	1.3 / 99.9	0.7 / 100
India	0.3 / 100	0.7 / 100	0.8 / 99.9	0.4 / 100	0.1 / 100	0.2 / 100	0.3 / 100	0.2 / 100
the UK	0.1 / 100	0.4 / 100	0.9 / 99.9	0.4 / 100	0.1 / 100	0.2 / 100	0.4 / 100	0.2 / 100
The U.S.	0.1 / 100	0.4 / 100	0.6 / 100	0.3 / 100	0.1 / 100	0.2 / 100	0.3 / 100	0.1 / 100
80% sensitivity and 90% specificity								
Korea	3.8 / 99.9	8.8 / 99.7	11.5 / 99.6	6.3 / 99.8	1.4 / 100	3.1 / 99.9	4.3 / 99.9	2.3 / 99.9
Japan	2.1 / 99.9	6.8 / 99.8	11.1 / 99.7	5.7 / 99.8	0.9 / 100	2.0 / 99.9	3.5 / 99.9	1.8 / 99.9
China	1.8 / 100	4.3 / 99.9	6.7 / 99.8	3.2 / 99.9	0.6 / 100	1.6 / 100	2.7 / 99.9	1.2 / 100
Albania	1.4 / 100	3.2 / 99.9	5.3 / 99.8	2.5 / 99.9	0.7 / 100	2.1 / 99.9	3.3 / 99.9	1.5 / 100
Belarus	1.8 / 100	4.4 / 99.9	6.2 / 99.8	3.0 / 99.9	0.7 / 100	1.7 / 100	2.5 / 99.9	1.3 / 100
India	0.5 / 100	1.3 / 100	1.6 / 100	0.9 / 100	0.3 / 100	0.5 / 100	0.6 / 100	0.4 / 100
The UK	0.3 / 100	0.9 / 100	1.7 / 100	0.7 / 100	0.1 / 100	0.4 / 100	0.7 / 100	0.3 / 100
The U.S.	0.3 / 100	0.8 / 100	1.2 / 100	0.6 / 100	0.1 / 100	0.3 / 100	0.5 / 100	0.3 / 100
90% sensitivity and 70% specificity								
Korea	1.5 / 99.9	3.5 / 99.8	4.6 / 99.8	2.5 / 99.9	0.5 / 100	1.2 / 99.9	1.6 / 99.9	0.9 / 100
Japan	0.8 / 100	2.7 / 99.9	4.5 / 99.8	2.2 / 99.9	0.3 / 100	0.7 / 100	1.3 / 99.9	0.7 / 100
China	0.7 / 100	1.7 / 99.9	2.6 / 99.9	1.2 / 99.9	0.2 / 100	0.6 / 100	1.0 / 100	0.5 / 100
Albania	0.5 / 100	1.2 / 99.9	2.1 / 99.9	1.0 / 100	0.3 / 100	0.8 / 100	1.3 / 99.9	0.6 / 100
Belarus	0.7 / 100	1.7 / 99.9	2.4 / 99.9	1.2 / 99.9	0.3 / 100	0.6 / 100	1.0 / 100	0.5 / 100
India	0.2 / 100	0.5 / 100	0.6 / 100	0.3 / 100	0.1 / 100	0.2 / 100	0.2 / 100	0.1 / 100
The UK	0.1 / 100	0.3 / 100	0.7 / 100	0.3 / 100	0.0 / 100	0.1 / 100	0.3 / 100	0.1 / 100
The U.S.	0.1 / 100	0.3 / 100	0.5 / 100	0.2 / 100	0.1 / 100	0.1 / 100	0.2 / 100	0.1 / 100
90% sensitivity and 80% specificity								
Korea	2.2 / 99.9	5.1 / 99.8	6.8 / 99.8	3.7 / 99.9	0.8 / 100	1.7 / 100	2.4 / 99.9	1.3 / 100
Japan	1.2 / 100	4.0 / 99.9	6.6 / 99.8	3.3 / 99.9	0.5 / 100	1.1 / 100	2.0 / 99.9	1.0 / 100
China	1.0 / 100	2.5 / 99.9	3.9 / 99.9	1.8 / 99.9	0.4 / 100	0.9 / 100	1.5 / 100	0.7 / 100
Albania	0.8 / 100	1.8 / 99.9	3.1 / 99.9	1.4 / 100	0.4 / 100	1.2 / 100	1.9 / 99.9	0.9 / 100
Belarus	1.0 / 100	2.5 / 99.9	3.6 / 99.9	1.7 / 100	0.4 / 100	1.0 / 100	1.4 / 100	0.7 / 100
India	0.3 / 100	0.7 / 100	0.9 / 100	0.5 / 100	0.2 / 100	0.3 / 100	0.3 / 100	0.2 / 100
The UK	0.1 / 100	0.5 / 100	1.0 / 100	0.4 / 100	0.1 / 100	0.2 / 100	0.4 / 100	0.2 / 100
The U.S.	0.2 / 100	0.4 / 100	0.7 / 100	0.3 / 100	0.1 / 100	0.2 / 100	0.3 / 100	0.1 / 100
90% sensitivity and 90% specificity								
Korea	4.3 / 99.9	9.8 / 99.9	12.7 / 99.8	7.0 / 99.9	1.5 / 100	3.4 / 100	4.8 / 99.9	2.6 / 100
Japan	2.4 / 100	7.6 / 99.9	12.3 / 99.8	6.3 / 99.9	1.0 / 100	2.2 / 100	3.9 / 100	2.1 / 100
China	2.0 / 100	4.9 / 99.9	7.4 / 99.9	3.5 / 100	0.7 / 100	1.8 / 100	3.0 / 100	1.3 / 100
Albania	1.6 / 100	3.6 / 100	6.0 / 99.9	2.8 / 100	0.8 / 100	2.4 / 100	3.7 / 100	1.7 / 100
Belarus	2.0 / 100	5.0 / 99.9	6.9 / 99.9	3.4 / 100	0.8 / 100	1.9 / 100	2.8 / 100	1.5 / 100
India	0.6 / 100	1.5 / 100	1.8 / 100	1.0 / 100	0.3 / 100	0.5 / 100	0.6 / 100	0.4 / 100
The UK	0.3 / 100	1.0 / 100	1.9 / 100	0.8 / 100	0.1 / 100	0.4 / 100	0.8 / 100	0.4 / 100
The U.S.	0.3 / 100	0.8 / 100	1.3 / 100	0.6 / 100	0.2 / 100	0.4 / 100	0.6 / 100	0.3 / 100

The UK, The United Kingdom, The U.S., The United States of America.

[a] NPVs >99.95% are listed as 100%.

[b] Weighted sum of age-specific estimates using region- and country-specific relative population size weights (weights can be found in Appendix 9).

CURRICULUM VITAE

Personal data		
Name:	Agne Krilaviciute	
Date of Birth:	21 October 1988	
Place of Birth:	Vilnius, Lithuania	
Marital status:	Single	
Nationality:	Lithuanian	
Education (school)		
1995-1999		primary school
1999-2007		high school
13 July 2007		Brandos atestatas (equivalent to Abitur)
Education (university)		
2007-2011	Bachelor of Informatics	Mathematics and Informatics Faculty, Vilnius University
2011-2013	Master of Informatics	Mathematics and Informatics Faculty, Vilnius University
2013-now	PhD	Division of Clinical Epidemiology and Aging Research, German Cancer Research Center
Professional Experiences		
2011-2013	Senior specialist	Cancer Registry, National Cancer Institute (Lithuania)
2012-2013	IT specialist	Biobank, National Cancer Institute (Lithuania)

ACKNOWLEDGMENT

Firstly, I would like to express my sincere gratitude to my supervisor, Prof. Hermann Brenner, for offering me the opportunity to learn and grow, for his patience, constructive criticism and continuous support throughout my PhD study. I am privileged to have an opportunity to be a part of his team.

I am heartily thankful for Prof. Marcis Leja for offering me the opportunity to use his data for my PhD study. I am sincerely grateful for his continuous support and interest in my work, advices and moral support. My thanks also go to Ieva Lašina and Sergejs Paršutins for their immediate help whenever I needed it.

I thank my boyfriend for his support and advices through all the stages of my work.

Finally, I would like to thank my family for supporting me and believing in me. In particular, my thanks go to my brother who spent countless hours listening to me and encouraging me to continue my study.