≋CHEST®

# Prospective Detection of Early Lung Cancer in Patients With COPD in Regular Care by Electronic Nose Analysis of Exhaled Breath

Check for updates

*Rianne de Vries, PhD; Niloufar Farzan, PhD; Timon Fabius, PhD; Frans H. C. De Jongh, PhD; Patrick M. C. Jak;*
*Eric G. Haarman, MD; Erik Snoey; Johannes C. C. M. In 'T Veen, MD, PhD; Yennece W. F. Dagelet;*
*Anke-Hilse Maitland-Van Der Zee, PhD; Annelies Lucas, MD, PHD; Michel M. Van Den Heuvel, MD, PhD;*
*Marguerite Wolf-Lansdorf; Mirte Muller, MD; Paul Baas, MD, PhD; and Peter J. Sterk, MD, PhD*

**BACKGROUND:** Patients with COPD are at high risk of lung cancer developing, but no validated predictive biomarkers have been reported to identify these patients. Molecular profiling of exhaled breath by electronic nose (eNose) technology may qualify for early detection of lung cancer in patients with COPD.

**RESEARCH QUESTION:** Can eNose technology be used for prospective detection of early lung cancer in patients with COPD?

**STUDY DESIGN AND METHODS:** BreathCloud is a real-world multicenter prospective follow-up study using diagnostic and monitoring visits in day-to-day clinical care of patients with a standardized diagnosis of asthma, COPD, or lung cancer. Breath profiles were collected at inclusion in duplicate by a metal-oxide semiconductor eNose positioned at the rear end of a pneumotachograph (SpiroNose; Breathomix). All patients with COPD were managed according to standard clinical care, and the incidence of clinically diagnosed lung cancer was prospectively monitored for 2 years. Data analysis involved advanced signal processing, ambient air correction, and statistics based on principal component (PC) analysis, linear discriminant analysis, and receiver operating characteristic analysis.

**RESULTS:** Exhaled breath data from 682 patients with COPD and 211 patients with lung cancer were available. Thirty-seven patients with COPD (5.4%) demonstrated clinically manifest lung cancer within 2 years after inclusion. Principal components 1, 2, and 3 were significantly different between patients with COPD and those with lung cancer in both training and validation sets with areas under the receiver operating characteristic curve of 0.89 (95% CI, 0.83-0.95) and 0.86 (95% CI, 0.81-0.89). The same three PCs showed significant differences ($P < .01$) at baseline between patients with COPD who did and did not subsequently demonstrate lung cancer within 2 years, with a cross-validation value of 87% and an area under the receiver operating characteristic curve of 0.90 (95% CI, 0.84-0.95).

**INTERPRETATION:** Exhaled breath analysis by eNose identified patients with COPD in whom lung cancer became clinically manifest within 2 years after inclusion. These results show that eNose assessment may detect early stages of lung cancer in patients with COPD.

CHEST 2023; 164(5):1315-1324

**KEY WORDS:** breath test; breathomics; COPD; early detection; eNose; lung cancer

---

## Take-home Points

**Study Question:** Can electronic nose (eNose) technology be used for prospective detection of early lung cancer in patients with COPD?

**Results:** The eNose was able to discriminate patients with COPD who subsequently received a lung cancer diagnosis from those who did not receive such a diagnosis with 87% accuracy, an area under the receiver operating characteristic curve of 0.90 (95% CI, 0.84-0.95), 86% sensitivity, and 89% specificity.

**Interpretation:** These results show that eNose assessment may detect early stages of lung cancer in patients with COPD and therefore may be of value in screening this risk group.

Patients with COPD are at higher risk of lung cancer developing, with studies showing a relative risk of twofold to fourfold compared with the general population.[1] Although several biomarkers are candidates for lung cancer discovery, such as autoantibodies, complement fragments, microRNA, circulating DNA, DNA methylation, RNA and protein profiling, and metabolomics,[2] no validated biomarkers have been discovered that can identify patients with COPD who are at a higher risk of lung cancer developing.

During the past decade, screening studies using low-dose CT imaging have shown significant reductions (up to 39% in women) in lung cancer mortality rates among high-risk individuals, who were defined based on tobacco use history (status, pack-years, quit time) and age (50-75 years).[3-5] Interestingly, a recently published study showed that patients with COPD are at higher risk of lung cancer regardless of tobacco use history.[6]

Because the potential benefits of screening might be exceeded by the increased risk of death inherent to COPD and its associated comorbidities, concerns exist about the inclusion of these patients in lung cancer screening programs.[7,8] In addition, concerns regarding increased overdiagnosis by low-dose CT imaging and invasive follow-up investigations limit its applicability to patients with COPD.[9] Therefore, an urgent need exists for an accurate and noninvasive test that can be implemented at the point of care. Specifically, a test that can refine the selection procedure of high-risk individuals for further follow-up screening tests is likely to prevail in clinical practice.

Molecular profiling of exhaled breath by electronic nose (eNose) technology may qualify for the early detection of lung cancer.[2] eNose technology is an appealing noninvasive approach that applies advanced pattern recognition algorithms for analysis of the mixture of volatile organic compounds (VOCs) in exhaled breath.[10,11] The thousands of VOCs present in exhaled breath reflect the metabolic processes occurring in the host both locally in the airways and systemically.[12] Comprehensive analysis of these VOC patterns (breathomics) provides opportunities for noninvasive biomarker discovery in lung cancer.[10,13,14] By using eNoses, differences in exhaled VOC patterns of individuals with COPD, individuals with lung cancer, and healthy individuals already have been demonstrated by multiple research groups.[11,15-19] Therefore, we hypothesized that metabolic and molecular changes that occur in early stage asymptomatic lung cancer can be detected from exhaled breath using an eNose.

This study aimed to determine the diagnostic accuracy of exhaled breath analysis by eNose for (1) the discrimination between patients with COPD and those with lung cancer in a training and validation set and (2) the prospective prediction of early lung cancer in COPD. By using this stepwise approach and transparent reporting, the study follows the recommendations of the Standards for Reporting of Diagnostic Accuracy Studies guidelines[20] and the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis statement.[21]

## Study Design and Methods
### Study Population

BreathCloud is a real-world multicenter observational study in healthy control participants and participants with a suspected or established diagnosis of asthma, COPD, or lung cancer.[22] All patients who

Institute (M. W.-L., M. M., and P. B.), Amsterdam; Breathomix B.V. (R. d. V., N. F., and Y. W. F. D.), Leiden, Medisch Spectrum Twente (T. F. and F. H. C. D. J.), Enschede; Franciscus Gasthuis & Vlietland (E. S. and J. C. C. M. I. T. V.), Rotterdam, Diagnostiek voor U (A. L.), Eindhoven; and Radboud University Medical Center (M. M. V. D. H.), Nijmegen, The Netherlands.

**CORRESPONDENCE TO:** Rianne de Vries, PhD; email: riannedevries1@gmail.com

visited the lung function departments for diagnostic and monitoring assessments in day-to-day care were recruited sequentially. The presently reported data include results from 682 patients with COPD and 211 patients with lung cancer who were included between May 2017 and November 2018. Patients with COPD were characterized according to the Global Initiative for Chronic Obstructive Lung Disease criteria[23] and an established medical diagnosis of lung cancer was based on current guidelines.[24,25] Exclusion criteria for participating in this study were the recent (< 12 h) intake of alcohol or if patients were not willing or able to participate. No further restrictions (eg, eating, drinking, tobacco use, or medication use) on participation were made to increase the applicability of breath analysis by eNose in clinical practice.

The ethics board of all participating centers concluded in writing that Dutch legislation on human participation in research was not considered to be applicable, given the noninvasive and minimally bothering nature of this study that merely added exhaled breath analysis by eNose to standard diagnostic procedures.[22,26,27] Despite the waiver that was provided by the ethics review board, the purpose of adding the eNose to routine diagnostics was explained to the patients, all of whom gave their oral consent.

### Study Design and Measurements

The study had an observational prospective follow-up design. At baseline all patients with COPD underwent eNose assessment. The patients with COPD were treated according to standard care, and the subsequent incidence of clinically diagnosed lung cancer was assessed prospectively based on documented clinical records by following up the patients prospectively for 2 years. An established medical diagnosis of lung cancer was confirmed with CT scan imaging and was based on current guidelines.[24,25]

Clinical assessment of patients with COPD was performed using the Clinical COPD Questionnaire.[28] In addition, the personal best $FEV_1$ % predicted after bronchodilation was used from data collected up to 12 months before inclusion. Other clinical data were collected for routine clinical care and subsequently were handled by complying with the Wet Bescherming Persoonsgegevens (Dutch Personal Data Protection Act).

### Exhaled Breath Analysis

Exhaled breath measurements were performed in duplicate using a technically and clinically validated eNose, the SpiroNose.[22,26,27,29] The SpiroNose consists of seven different cross-reactive metal-oxide semiconductor sensors (sensors 1-7) for sampling of exhaled air. Another set of the same sensors sampled ambient air for background correction (Fig 1). SpiroNose measurements comprise five tidal breaths followed by an inspiratory capacity maneuver to total lung capacity and a 5-s breath hold followed by a slow expiration (< 0.4 L/s) to residual volume. The raw SpiroNose sensor signals were sent in real time to an online analysis platform for automated data analysis. An in-depth description of the measurement setup and the verification of the sensor stability is published elsewhere.[22,26]

### Data Processing

The processing of the eNose sensor deflections was carried out automatically using the standard eNose software as was published previously.[22,27] Signal processing included signal detrending, filtering, ambient air correction, automatic peak detection, and parameter selection. From each sensor signal, two variables were determined: (1) the highest sensor peak normalized to the most stable sensor, sensor 2, to minimize inter array differences and (2) the ratio between the sensor peak and the breath hold point.[22,26] The sensor peak ratios and peak to breath hold ratios were used for statistical analysis.
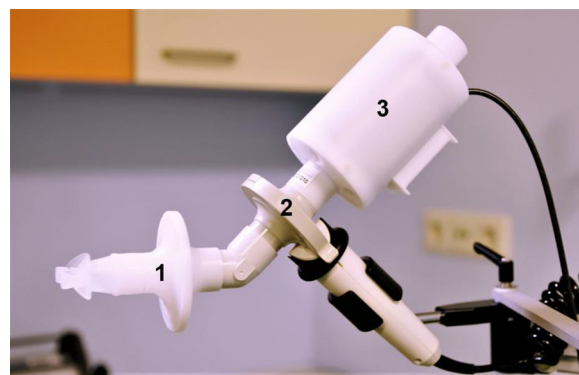


Figure 1 – *Photograph showing the SpiroNose measurement setup: (1) mouthpiece, nose clamp, and bacteria filter; (2) spirometer; and (3) SpiroNose.*

### Statistical Analysis

A principal component analysis (PCA) was performed to merge the variables of interest into a multivariate component. According to the Kaiser criterion, all principal components (PCs) with an eigenvalue of > 1 were retained.[30] The processed sensor variables, the original sensor peaks, and the peak to breath hold ratios were restructured to four PCs that captured 78.4% of the variance within the dataset (PC 1, 39.8%; PC 2, 19.5%; PC 3, 11.1%; and PC 4, 8.0%). PCs were constructed for all participants (training and validation set) based on eNose data from participants within the training set (e-Table 1). Reducing the dimensionality of the data before machine learning is preferred to reduce the risk of overfitting.[31,32]

### COPD vs Lung Cancer

Analyses were performed in both a training and validation set defined by random split analysis (2:1), as recommended for metabolomics experiments.[33] The obtained PCs were compared between groups using independent $t$ tests. The PCs that discriminated ($P < .05$) between groups were selected for further analysis. The $t$ tests were validated internally by 1,000 bootstrap iterations. Subsequently, linear discriminant analysis was performed using the selected variables. The aim of linear discriminant analysis is to maximize the separability of the defined categories and represents a relatively simple and portable algorithm. The latter is recommended for eNose research[31] and often is used in clinical studies performing exhaled breath analysis with the SpiroNose.[26,27] Based on the differentiating PCs, a discriminant function was calculated that best distinguished between groups. The accuracy of this model was defined as the percentage of correctly classified patients in the training set. Leave-one-out cross-validation was used to calculate the cross-validated accuracy value (percentage). The discriminant scores were used to construct a receiver operating characteristic curve, including the area under the receiver operating characteristic curve (AUC) and corresponding 95% CI. Finally, the discriminant function obtained from the training set was examined in the independent validation set and compared mutually based on AUC curves.

As a first step in the assessment of early lung cancer, the training and validation sets were combined and the accuracy, sensitivity, and specificity for the discrimination between COPD and lung cancer were assessed across the entire study population. A PCA plot was created to visualize the data. Finally, the influence of comorbid COPD in the lung cancer group on the accuracy of distinguishing lung cancer from COPD was assessed by removing all patients with a double diagnosis from the analysis (e-Appendix 1, e-Fig 1, e-Table 2).

### Other Classification Algorithms

To avoid drawing conclusions based on a single classification model, we used three additional and powerful machine learning techniques to classify the eNose data: gradient boosting machine, adaptive least absolute shrinkage and selection operator, and sparse partial least squares discriminant analysis. These machine learning techniques all have been used before in metabolomics research[34-36] and provide an estimate of the robustness of the statistical performance across different models (e-Appendix 1, e-Tables 3, 4). To contribute to the standardization of the SpiroNose's methods, we chose to present the linear discriminant analysis results throughout herein.

### Prospective Early Lung Cancer Detection

Similar analyses including PCA, independent *t* tests, linear discriminant analysis, and receiver operating characteristic curve

analysis were used to determine the diagnostic accuracy of eNose analysis for the discrimination between patients with COPD who did and did not receive a clinical diagnosis of lung cancer within 2 years after inclusion. In addition, sensitivity, specificity, a positive likelihood ratio (LR), and a negative LR were calculated. A PCA plot was created to visualize the data. A duplicate of this plot was created, in which the COPD group that did receive a clinical diagnosis of lung cancer within 2 years was divided into the different lung cancer stages (e-Fig 2).

### Clinical and Biological Determinants on the eNose Data

A multiple linear regression was calculated to predict eNose data based on the clinical metadata. More information on the analysis and results is presented in e-Tables 5-7.

## Results

### COPD vs Lung Cancer

In total, 893 patients (682 patients with COPD and 211 patients with lung cancer) were included in this study, of whom 596 patients (455 patients with COPD and 141 patients with lung cancer) were included for training of the results and 297 patients (227 patients with COPD and 70 patients with lung cancer) for validation of the results (Table 1). In the training set, only pack-years of tobacco use and $FEV_1$ showed significant differences ($P < .05$) between patients with COPD and those with lung cancer. No significant differences were found between these groups in baseline characteristics in the validation set. In addition, the baseline characteristics of patients in the training and validation set were similar, except for significantly different COPD staging ($P < .05$) and lung cancer pathologic features ($P < .05$) (Table 1).

In the training set, PC 1 ($P = .039$), PC 2 ($P < .001$), and PC 3 ($P < .001$) were significantly different between patients with COPD and those with lung cancer. Subsequent discriminant analysis showed a cross-validated accuracy value of 88%. The AUC ± 95% CI after internal cross-validation reached 0.89 (95% CI, 0.83-0.95). The ability to discriminate between COPD and lung cancer, with PC 1, PC 2, and PC 3 as input for the model obtained from the training set, was confirmed in an independent validation set. Breathprints of patients with COPD and those with lung cancer in the validation set were distinguished with a cross-validated accuracy value of 83% and an AUC of 0.86 (95% CI, 0.81-0.89). The results of the other classification algorithms in the validation set were similar and are presented in e-Tables 3 and 4.

For completeness and as a first step toward assessment of early lung cancer, we combined both

the training and validation sets. PC 2 ($P < .001$) and PC 3 ($P < .001$) were significantly different between COPD and lung cancer. Groups were distinguished with a cross-validated accuracy of 84% and the AUC reached 0.86 (95% CI, 0.82-0.94) (Fig 2). When excluding all patients with lung cancer with comorbid COPD from the analysis (e-Table 2, e-Fig 1), a cross-validated accuracy of 90% and AUC of 0.95 (95% CI, 0.92-0.97) were reached.

### Prospective Early Lung Cancer Detection

Of the total 682 patients with COPD, 37 patients (5.4%) demonstrated clinically manifest lung cancer within 2 years after inclusion (Table 2). PC 1 ($P = .002$), PC 2 ($P < .001$), and PC 3 ($P < .001$) showed a significant difference at baseline between the patients with COPD who did and did not receive a diagnosis of lung cancer within 2 years, with a cross-validated accuracy value of 87%. The AUC after internal cross-validation reached 0.90 (95% CI, 0.84-0.95). Thirty-three of these 37 patients with COPD (89%) were classified by eNose as having lung cancer already at baseline (Fig 3). The identification of patients with COPD who demonstrated clinically manifest lung cancer resulted in 86% sensitivity, 89% specificity, a positive LR of 7.80, and a negative LR of 0.15.

### Lung Cancer Stage

The COPD group that did receive a clinical diagnosis of lung cancer was divided into early and advanced stage lung cancer (e-Fig 2). PC 3 ($P < .001$) was significantly different between early (stages I and II) and advanced (stages III and IV) lung cancer. Stages were distinguished with a cross-validated accuracy of 88% and an AUC that reached 0.93 (95% CI, 0.87-0.98).

TABLE 1 ] Baseline Characteristics of Patients With COPD and Those With Lung Cancer From Both the Training and Validation Sets

| Variable | Training Set | | Validation Set | |
|---|---|---|---|---|
| | COPD (n = 455) | Lung Cancer (n = 141) | COPD (n = 227) | Lung Cancer (n = 70) |
| Male sex | 191 (42) | 75 (53) | 91 (40) | 33 (47) |
| Age, y | 64.3 ± 9.4 | 63.0 ± 9.9 | 63.1 ± 8.1 | 64.6 ± 10.9 |
| BMI, kg/m$^2$ | 26.2 ± 5.4 | 25.2 ± 4.8 | 27.9 ± 6.8 | 26.1 ± 4.4 |
| Tobacco use history | | | | |
| Never | 3 (0) | 14 (10) | 1 (0) | 7 (10) |
| Former | 334 (73) | 92 (65) | 174 (77) | 48 (69) |
| Current | 118 (26) | 35 (25) | 52 (23) | 15 (21) |
| Pack-y of tobacco use | 37.6 (31.1-55.8)[a] | 29.4 (25.9-47.3)[a] | 39.1 (31.9-53.8) | 35.2 (29.3-45.3) |
| White race | 398 (87) | 127 (90) | 205 (90) | 22 (97) |
| FEV$_1$, % predicted | 63.7 ± 18.4[a] | 72.0 ± 20.1[a] | 68.7 ± 23.5 | 69.4 ± 18.9 |
| COPD staging, GOLD classification | 455 (100) | 82 (58) | 227 (100) | 34 (49) |
| I | 41 (9) | 20 (14) | 18 (8) | 5 (7) |
| II | 245 (54)[b] | 39 (28) | 93 (41)[b] | 23 (32) |
| III | 129 (28)[b] | 22 (16) | 89 (39)[b] | 6 (9) |
| IV | 40 (9) | 1 (0) | 27 (12) | 0 (0) |
| Pathologic findings | | | | |
| SCLC | NA | 17 (12) | NA | 11 (16) |
| NSCLC | NA | 124 (88) | NA | 59 (84) |
| Adenocarcinoma | NA | 87 (62) | NA | 35 (50) |
| Squamous cell carcinoma | NA | 22 (16)[b] | NA | 18 (26)[b] |
| Large cell carcinoma | NA | 15 (11) | NA | 6 (9) |
| Stage | | | | |
| I | NA | 17 | NA | 5 |
| II | NA | 31 | NA | 16 |
| III | NA | 48 | NA | 29 |
| IV | NA | 45 | NA | 20 |

Data are presented as No. (%), No., mean ± SD, or median (interquartile range). GOLD = Global Initiative for Chronic Obstructive Lung Disease; NA = not applicable; NSCLC = non-small cell lung cancer; SCLC = small cell lung cancer.
[a]Significant difference between COPD and lung cancer group (P < .05).
[b]Significant difference between training and validation set (P < .05).

## Discussion

This study showed that exhaled breath analysis using eNose is able to distinguish between lung cancer and COPD. These results were confirmed in an independent validation set. Additionally, of the 37 patients with COPD who received a clinical diagnosis of lung cancer within 2 years after inclusion in the study, 33 were classified correctly as having lung cancer by eNose already at baseline. The eNose was able to discriminate patients with COPD who subsequently received a lung cancer diagnosis from those who did not with 87% accuracy.

To our knowledge, this is the first prospective discovery study of early lung cancer among patients with COPD by using eNose. Distinct exhaled VOC patterns of lung cancer vs healthy individuals long have been demonstrated by researchers worldwide, including favorable accuracies using various eNose devices.[11,15-19,37,38] In the past few years, however, several research groups have taken further steps to assess the potential of exhaled breath analysis for discrimination between lung cancer and COPD, two closely associated diseases. The present study showed accurate discrimination of lung cancer and COPD, and more so when double diagnoses were excluded. Dragonieri et al[15] also showed significant differences in breath profiles of participants with non-small cell lung cancer, participants with COPD, and healthy control
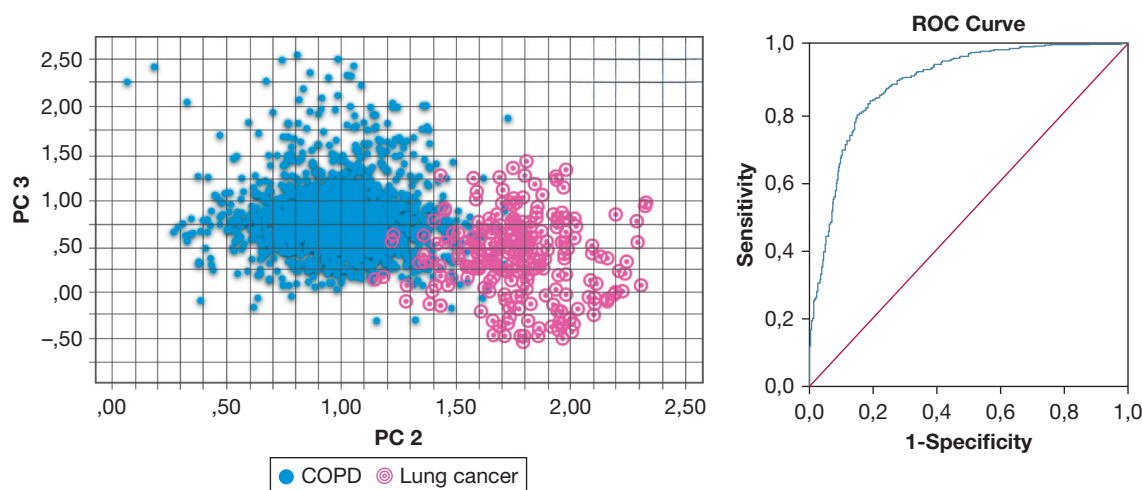
Figure 2 – A, Two-dimensional plot showing the discrimination of breath profiles between patients with COPD and those with lung cancer (combined training and validation set) along PC 2 and PC 3 (PCs with the highest discriminative value). B, Graph showing the ROC curve with line of identity for the identification of lung cancer of 0.86 (95% CI, 0.81-0.89). PC = principal component; ROC = receiver operating characteristic.

participants in a small eNose study. Similar to our results, they were able to distinguish between the groups with 85% accuracy. Another study using gas chromatography-mass spectrometry showed that levels of several VOCs, including isoprene and methyl pentane, were significantly higher in non-small cell lung cancer compared with COPD.[17] Increased levels of isoprene in the breath of patients with lung cancer have been reported in numerous studies.[39-41] One of the important conclusions of the study by Poli et al[17] is that none of the VOCs alone was able to distinguish between the two groups adequately. In fact, the highest accuracy (83%; with 72% sensitivity and 94% specificity) was achieved when combining 13 VOCs. Indeed, considering the complexity of the two diseases regarding their underlying cellular and molecular pathways, difficulties in identifying highly informative single biomarkers are not surprising. Therefore, technologies such as eNose that capture the complete mixture of VOCs combined with pattern recognition algorithms are highly valuable in identifying composite biomarkers, providing better insights into the actual condition of patients.[11]

In the current study, we used a technically and clinically validated eNose linked to an online cloud solution to perform exhaled breath measurements.[22,26] It should be noted that eNoses can differ significantly in number and type of sensors as well as algorithms used for signal processing, statistical analysis, or both.[42] In a study using eNose technology, 11 of 23 patients with COPD (47%) were identified as having lung cancer and only one of those patients received a clinical diagnosis of lung cancer within 56 months.[16] As concluded by the authors, a lack

of sufficient sample size for adequate training of the eNose could be one of the reasons for suboptimal distinction. The present study, using a bigger sample size, distinct signal processing, and dissimilar statistical analysis using in total four classification models, seemed to confirm this.

The present study was a spin-out of the multicenter BreathCloud project. A potential bias could be that patients with COPD were recruited mostly from secondary care centers and had received a diagnosis clinically according to the Global Initiative for Chronic Obstructive Lung Disease criteria. In the Netherlands, patients with mild to moderate COPD are managed mainly at primary care centers by general practitioners, whereas patients with more severe forms of the disease are managed at secondary care centers. The presence of moderate or severe COPD is shown to be associated with a higher risk of lung cancer,[6] which could explain the relatively high lung cancer incidence in the present study population (5%) compared with that of previous reports (approximately 1%).[43,44]

Tobacco use is known to influence the composition of exhaled VOCs, and therefore is considered to be a potential confounder in exhaled breath analysis.[45] Although pack-years of tobacco use drove the prediction of PC 2 (not PC 1 or PC 3) significantly, tobacco use status did not. It could very well be that hematologic parameters are affected with long-term tobacco use and that this could affect exhaled breath profiles.[46] Although the median pack-years of tobacco use in patients with COPD were higher than in patients with lung cancer in

**TABLE 2 ]** Baseline Characteristics of Patients With COPD Who Did and Did Not Demonstrate Clinically Manifest Lung Cancer Within 2 Years After Inclusion

| Variable | Clinically Manifest Lung Cancer Did Not Develop (n = 645) | Clinically Manifest Lung Cancer Did Develop (n = 37) |
|---|---|---|
| Male sex | 268 (42) | 14 (38) |
| Age, y | 63.7 ± 14.2 | 64.5 ± 16.9 |
| BMI, kg/m$^2$ | 26.9 ± 5.0 | 25.7 ± 6.2 |
| Tobacco use history | | |
| Never | 4 (0) | 0 (0) |
| Former | 483 (75) | 25 (68) |
| Current | 158 (24) | 12 (32) |
| Pack-y of tobacco use | 38.2 (31.2-55.4) | 36.1 (30.3-53.0) |
| White race | 572 (89) | 31 (84) |
| FEV$_1$, % predicted | 65.0 ± 20.1 | 69.4 ± 18.9 |
| COPD staging, GOLD classification | | |
| I | 59 (9) | 0 (0) |
| II | 327 (51)[a] | 11 (30)[a] |
| III | 202 (31) | 16 (43) |
| IV | 57 (9)[a] | 10 (27)[a] |
| Pathologic findings | | |
| SCLC | NA | 5 (14) |
| NSCLC | NA | 32 (86) |
| Adenocarcinoma | NA | 17 (46) |
| Squamous cell carcinoma | NA | 14 (38) |
| Large cell carcinoma | NA | 1 (3) |
| Staging | | |
| I | NA | 5 |
| II | NA | 12 |
| III | NA | 13 |
| IV | NA | 7 |

Data are presented as No. (%), No., mean ± SD, or median (interquartile range). GOLD = Global Initiative for Chronic Obstructive Lung Disease; NA = not applicable; NSCLC = non-small cell lung cancer; SCLC = small cell lung cancer.
[a]Significant difference between COPD and early lung cancer group (P < .05).

the training set, no significant differences were found between the two groups in the validation set. For this reason and because pack-years of tobacco use did not contribute significantly to the prediction of PC 1 or PC 3, tobacco use is unlikely to have influenced our findings, consistent with a previous report.[47] Furthermore, our analysis demonstrated no significant differences regarding tobacco use status or pack-years of tobacco use in patients with COPD who received a lung cancer diagnosis compared with those who did not.

Furthermore, age, BMI, ethnicity, lung cancer staging, and lung cancer pathologic features also added significantly to the prediction of one or two PCs (eNose data). However, these baseline characteristics in the training and validation set were similar for patients with

COPD and those with lung cancer and were similar between the training and validation set, except for lung cancer pathologic features. The validation set contained slightly more patients with squamous cell carcinoma, which may explain in part the small difference in results between the training and validation set.

One of the main strengths of our study is its prospective, real-world design and the use of an independent validation set and multiple classification algorithms for distinguishing between lung cancer and COPD that confirmed the results of the training set. Another strength of the study is its use of a technically validated eNose. The extensive technical validation of the eNose was performed regarding the development of signal processing, ambient air correction, and data analysis
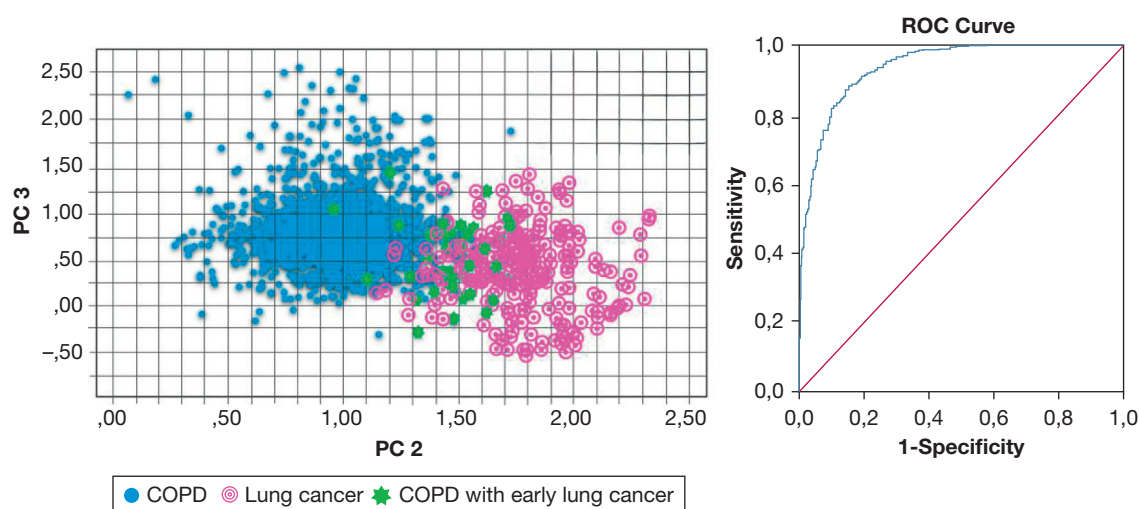
Figure 3 – A, Two-dimensional plot showing the discrimination of breath profiles between patients with COPD and those with COPD who demonstrated clinically manifest lung cancer within 2 years after inclusion in the study (COPD with early lung cancer) along PC 2 and PC 3. The lung cancer breath profiles are plotted as a reference. B, Graph showing ROC curve with line of identity for the identification of patients with COPD with early lung cancer of 0.90 (95% CI, 0.84-0.95). PC = principal component; ROC = receiver operating characteristic.

techniques as well as identifying the optimal measurement maneuver (eg, flow, volume) and influence of environmental factors (eg, humidity, temperature) on breath measurements.[27] Our study certainly also has limitations. The real-world design lacked CT scan imaging at baseline. Therefore, it is unclear whether a tumor already was present at baseline. None of the patients with COPD showed symptoms suggestive of lung cancer at that time. Still, it cannot be excluded that, for some patients with COPD with a subsequent diagnosis of lung cancer, a baseline CT scan might have detected the existing tumor. However, evidence that imaging techniques exhibit high false-positive rates is good, while also showing that these methods have difficulty in distinguishing between benign and malignant tumors.[3,48-50]

## Interpretation

The data suggest that cancer-related cellular and metabolic processes already were present at the time of exhaled breath measurement despite the absence of any symptoms suggestive of lung cancer. Therefore, the obtained breath profiles in these patients seem to reflect the composition of VOCs that arise from both COPD and lung cancer. Baseline breath profiles of patients with COPD who subsequently demonstrated clinically manifest lung cancer were comparable with those in whom COPD and lung cancer coexist. However, because eNoses analyze the complete mixture of VOCs, it remains to be elaborated whether VOCs arise from

cancer cells, the host's immunologic or inflammatory responses, or the tumor microbiome.[51]

Interestingly, the VOC pattern associated with early development of lung cancer in COPD did not match to the pattern related to lung cancer stages: the former was mainly captured by PC 2 and the latter by PC 3 (e-Fig 2). This suggests that early identification of upcoming clinically manifest lung cancer in patients with COPD by eNose analysis is not driven by VOCs that are associated predominantly with a particular stage of the disease.

What could be the implications of this study? According to the Global Initiative for Chronic Obstructive Lung Disease report, lung cancer is the main cause of death among patients with COPD. The eligibility criteria of current lung cancer screening programs are limited to age and tobacco use history, and they show high false-positive rates.[3] It is becoming more evident that COPD is a risk factor for lung cancer. Recently, the HUNCHEST screening program reported higher relative risk for malignant tumor entity in patients with COPD compared with patients without COPD (relative risk, 1.85; 95% CI, 0.85-4.07).[52] Moreover, COPD has been shown to be an independent risk factor for lung cancer regardless of the tobacco use status of the patient because a large-scale study demonstrated that those with COPD who have never used tobacco have a 2.6-times higher risk of lung cancer developing compared with those without COPD who have never used tobacco.[6] Therefore, a large number of lung cancer cases may not be detected because of the exclusion of current and

former tobacco users with COPD from screening programs,[6,53] whereas the number of missed cases might be even higher when considering patients with COPD who do not use tobacco. Therefore, optimizing the screening criteria for lung cancer by including patients with COPD seems to be recommended. Early detection not only results in potentially curative intervention, but it also can decrease the risk of life-threatening complications of treatment. Our data suggest that eNose technology can capture lung cancer-specific VOC patterns in its early development, which provides a window for potentially curable interventions.

In conclusion, exhaled breath analysis by eNose examination identified patients with COPD in whom lung cancer subsequently became clinically manifested within 2 years after inclusion. These results show that eNose assessment may provide a novel means for early identification of patients with COPD at risk of malignancy and may improve patient outcomes by identifying those who will benefit most from further diagnostic procedures and early intervention.

## Financial/Nonfinancial Disclosures

The authors have reported to *CHEST* the following: R. d. V. receives personal fees and has a substantial interest in the start-up company Breathomix BV. N. F. and Y. W. F. D. receive personal fees from the start-up company Breathomix BV. P. J. S. is scientific adviser and has an officially nonsubstantial interest in the start-up company Breathomix BV. None declared (T. F., F. H. C. D. J., P. M. C. J., E. G. H., E. S., J. C. C. M. I. T. V., A.-H. M.-V. D. Z., A. L., M. M. V. D. H., M. W.-L., M. M., P. B.).

## References

1. de Torres JP, Marín JM, Casanova C, et al. Lung cancer in patients with chronic obstructive pulmonary disease—incidence and predicting factors. *Am J Respir Crit Care Med.* 2011;184(8):913-919.

2. Seijo LM, Peled N, Ajona D, et al. Biomarkers in lung cancer screening: achievements, promises, and challenges. *J Thorac Oncol.* 2019;14(3):343-357.

3. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med.* 2011;365(5):395-409.

4. Ru Zhao Y, Xie X, de Koning HJ, Mali WP, Vliegenthart R, Oudkerk M. NELSON lung cancer screening study. *Cancer Imaging.* 2011;11. Spec No A(1A): S79-84.

5. de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. *N Engl J Med.* 2020;382(6):503-513.

6. Park HY, Kang D, Shin SH, et al. Chronic obstructive pulmonary disease and lung cancer incidence in never smokers: a cohort study. *Thorax.* 2020;75(6):506-509.

7. Wiener RS, Schwartz LM, Woloshin S, Welch HG. Population-based risk for complications after transthoracic needle lung biopsy of a pulmonary nodule: an analysis of discharge records. *Ann Intern Med.* 2011;155(3):137-144.

8. Kozower BD, Sheng S, O'Brien SM, et al. STS database risk models: predictors of mortality and major morbidity for lung cancer resection. *Ann Thorac Surg.* 2010;90(3):875-881; discussion 881-883.

9. Patz EF Jr, Pinsky P, Gatsonis C, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med.* 2014;174(2):269-274.

10. Rocco G, Pennazza G, Santonico M, et al. Breathprinting and early diagnosis of lung cancer. *J Thorac Oncol.* 2018;13(7): 883-894.

11. Farraia MV, Cavaleiro Rufo J, Paciência I, Mendes F, Delgado L, Moreira A. The electronic nose technology in clinical diagnosis: a systematic review. *Porto Biomed J.* 2019;4(4):e42.

12. Bos LD, Sterk PJ, Fowler SJ. Breathomics in the setting of asthma and chronic obstructive pulmonary disease. *J Allergy Clin Immunol.* 2016;138(4):970-976.

13. Lamote K, Van den Heuvel MM. Where the nose is going to help the eye: sniffing lung cancer. *Lung Cancer.* 2021;154: 195-196.

14. Haick H, Hashoul D. Lung cancer breath tests. *Expert Rev Respir Med.* 2019;13(7): 597-599.

15. Dragonieri S, Annema JT, Schot R, et al. An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD. *Lung Cancer.* 2009;64(2): 166-170.

16. Krauss E, Haberer J, Barreto G, Degen M, Seeger W, Guenther A. Recognition of breathprints of lung cancer and chronic obstructive pulmonary disease using the Aeonose® electronic nose. *J Breath Res.* 2020;14(4):046004.

17. Poli D, Carbognani P, Corradi M, et al. Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study. *Respir Res.* 2005;6(1):71.

18. Krilaviciute A, Heiss JA, Leja M, Kupcinskas J, Haick H, Brenner H. Detection of cancer through exhaled breath: a systematic review. *Oncotarget.* 2015;6(36):38643-38657.

19. Tsou P-H, Lin Z-L, Pan Y-C, et al. Exploring volatile organic compounds in breath for high-accuracy prediction of lung cancer. *Cancers.* 2021;13(6):1431.

20. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open.* 2016;6(11): e012799.

21. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for

individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med.* 2015;13(1):1.

22. de Vries R, Dagelet YWF, Spoor P, et al. Clinical and inflammatory phenotyping by breathomics in chronic airway diseases irrespective of the diagnostic label. *Eur Respir J.* 2018;51(1):1701817.

23. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. GOLD executive summary. *Am J Respir Crit Care Med.* 2017;195(5):557-582.

24. Postmus PE, Kerr KM, Oudkerk M, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2017;28(suppl 4):iv1-iv21.

25. Früh M, De Ruysscher D, Popat S, Crinò L, Peters S, Felip E. Small-cell lung cancer (SCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2013;24(suppl 6):vi99-vi105.

26. de Vries R, Muller M, van der Noort V, et al. Prediction of response to anti-PD-1 therapy in patients with non-small-cell lung cancer by electronic nose analysis of exhaled breath. *Ann Oncol.* 2019;30(10):1660-1666.

27. de Vries R, Brinkman P, van der Schee MP, et al. Integration of electronic nose technology with spirometry: validation of a new approach for exhaled breath analysis. *J Breath Res.* 2015;9(4):046001.

28. Reda AA, Kotz D, Kocks JW, Wesseling G, van Schayck CP. Reliability and validity of the clinical COPD questionnaire and chronic respiratory questionnaire. *Respir Med.* 2010;104(11):1675-1682.

29. Moor CC, Oppenheimer JC, Nakshbandi G, et al. Exhaled breath analysis by use of eNose technology: a novel diagnostic tool for interstitial lung disease. *Eur Respir J.* 2020;57(1):2002042.

30. Yeomans KA, Golder PA. The Guttman-Kaiser criterion as a predictor of the number of common factors. *Journal of the Royal Statistical Society Series D (The Statistician).* 1982;31(3):221-229.

31. Bikov A, Lázár Z, Horvath I. Established methodological issues in electronic nose research: how far are we from using these instruments in clinical settings of breath analysis? *J Breath Res.* 2015;9(3):034001.

32. Gromski PS, Correa E, Vaughan AA, Wedge DC, Turner ML, Goodacre R. A comparison of different chemometrics approaches for the robust classification of electronic nose data. *Anal Bioanal Chem.* 2014;406(29):7581-7590.

33. Broadhurst DI, Kell DB. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics.* 2006;2(4):171-196.

34. Bouwmeester R, Martens L, Degroeve S. Comprehensive and empirical evaluation of machine learning algorithms for small molecule LC retention time prediction. *Anal Chem.* 2019;91(5):3694-3703.

35. Cuperlovic-Culf M. Machine learning methods for analysis of metabolic data and metabolic pathway modeling. *Metabolites.* 2018;8(1):4.

36. Determan CE Jr. Optimal algorithm for metabolomics classification and feature selection varies by dataset. *Int J Biol.* 2015;7(1):100.

37. Chen X, Muhammad KG, Madeeha C, et al. Calculated indices of volatile organic compounds (VOCs) in exhalation for lung cancer screening and early detection. *Lung Cancer.* 2021;154:197-205.

38. Kort S, Brusse-Keizer M, Schouwink H, et al. Diagnosing non-small cell lung cancer by exhaled-breath profiling using an electronic nose: a multicenter validation study. *Chest.* 2023;163(3):697-706.

39. Ulanowska A, Kowalkowski T, Trawińska E, Buszewski B. The application of statistical methods using VOCs to identify patients with lung cancer. *J Breath Res.* 2011;5(4):046008.

40. Chen X, Cao M, Hao Y, 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;2005:5873-5876.

41. Saalberg Y, Wolff M. VOC breath biomarkers in lung cancer. *Clin Chim Acta.* 2016;459:5-9.

42. Leopold JH, Bos LD, Sterk PJ, et al. Comparison of classification methods in breath analysis by electronic nose. *J Breath Res.* 2015;9(4):046002.

43. Vineis P, Airoldi L, Veglia F, et al. Environmental tobacco smoke and risk of respiratory cancer and chronic obstructive pulmonary disease in former smokers and never smokers in the EPIC prospective study. *BMJ.* 2005;330(7486):277.

44. Skillrud DM, Offord KP, Miller RD. Higher risk of lung cancer in chronic obstructive pulmonary disease. A prospective, matched, controlled study. *Ann Intern Med.* 1986;105(4):503-507.

45. Kischkel S, Miekisch W, Sawacki A, et al. Breath biomarkers for lung cancer detection and assessment of smoking related effects—confounding variables, influence of normalization and statistical algorithms. *Clin Chim Acta.* 2010;411(21-22):1637-1644.

46. Herath P, Wimalasekera SW, Amarasekara TD, Fernando MS, Turale S. Adverse effects of cigarette smoking on exhaled breath carbon monoxide, blood carboxyhemoglobin, and hematological parameters amongst Sri Lankan adult tobacco smokers: a descriptive study. *Popul Med.* 2021;3:1-10.

47. Principe S, van Bragt J, Longo C, et al. The influence of smoking status on exhaled breath profiles in asthma and COPD patients. *Molecules.* 2021;26(5):1357.

48. Benzaquen J, Boutros J, Marquette C, Delingette H, Hofman P. Lung cancer screening, towards a multidimensional approach: why and how? *Cancers (Basel).* 2019;11(2):212.

49. Peikert T, Duan F, Rajagopalan S, et al. Novel high-resolution computed tomography-based radiomic classifier for screen-identified pulmonary nodules in the National Lung Screening Trial. *PloS One.* 2018;13(5):e0196910.

50. Tanoue LT, Tanner NT, Gould MK, Silvestri GA. Lung cancer screening. *Am J Respir Crit Care Med.* 2015;191(1):19-33.

51. Nejman D, Livyatan I, Fuks G, et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science.* 2020;368(6494):973-980.

52. Kerpel-Fronius A, Monostori Z, Kovacs G, et al. Nationwide lung cancer screening with low-dose computed tomography: implementation and first results of the HUNCHEST screening program. *Eur Radiol.* 2022;32(7):4457-4467.

53. Young RP, Hopkins RJ. Lung cancer risk prediction to select smokers for screening CT—letter. *Cancer Prev Res (Phila).* 2012;5(4):697-698; author reply, 699.