Breathomics in pulmonary disease
Fens, N.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
General discussion
1 CONCEPT OF BREATHOMICS
Metabolomics is the scientific study of chemical fingerprints that are the result of specific cellular processes [1]. Every breath that we exhale contains hundreds of different molecules, forming a complex pattern. This fingerprint of the breath, or breathprint, can provide information about the state of health or disease of a subject. The concept of breathomics can be considered to be as old as medicine itself. Already in ancient Greek and traditional Chinese medicine it was recognized that smelling the exhaled breath of patients could aid in making a diagnosis. However, with the emergence of blood and urine tests in the past century, the diagnostic dogma shifted from measuring a whole spectrum of biomarkers to measuring single biomarkers based on pathophysiological reasoning.

Although the focus of diagnostics has shifted towards single biomarkers, recent advances in technology made omics studies in breath a (re-)emerging field in medical science, as described in chapter 2. Qualitative analysis of metabolomic profiles in urine, blood and breath gained interest since the 1970’s [2], and since the end of the 20th century, a revolution in the understanding of the constituents of exhaled breath has taken place by the development of sensitive gas chromatography and mass spectrometry (GC-MS) instruments and electronic noses (eNoses).

2 BACKGROUND OF THE THESIS
In the studies described in this thesis we have focused on the validation of exhaled breath molecular profiling as a diagnostic test in pulmonary diseases, especially in the obstructive airway diseases asthma and COPD. Besides this, we have investigated what comprises the signal that is captured in exhaled breath in asthma and COPD and how exhaled breath testing can improve subphenotyping a heterogeneous disorder such as COPD. Additionally, we addressed some methodological issues concerning breath testing using electronic nose (eNose) technology as well as gas chromatography mass spectrometry (GC-MS) in order to apply these techniques in large multi-center studies. Finally, we looked beyond airway diseases to assess the diagnostic value of exhaled breath testing in acute pulmonary embolism.

In this chapter, all studies will be brought together and the reported findings will be discussed. Further methodological issues concerning both the technique and the validation of exhaled breath testing in medicine will be discussed, as well as aims that will need to be considered in future studies.

3 CONCLUSIONS AND IMPLICATIONS OF THE STUDIES
On evaluation of the data presented in this thesis, the following conclusions and implications can be drawn.
3.1 Breathomics in medicine

- Since the discovery of electronic noses and their application in exhaled breath molecular profiling, great advances have been made with respect to discrimination of diseases by comparing overall breathprints provided by the electronic noses (chapter 2).
  ♦ Implication: both technological (sensor design, materials, breath sampling) and medical (knowledge on specific compounds) advances have been made in the field of exhaled breath testing. Bringing these fields together may lead to tailor-made electronic noses for specific disease states.

3.2 Diagnostic validation in obstructive airway diseases

- As the first step following the STARD guidelines in the validation of electronic noses in the diagnosis of classical asthma and COPD, excellent internal validity of 96% accuracy was reached (chapter 3).
- In the second step in the validation of electronic noses, focused on external validity in a newly recruited group of asthma and COPD patients with classical disease, a good accuracy of 83% was reached (chapter 4).
- As the third step in the validation of electronic noses, a newly recruited group of difficult-to-diagnose asthma patients with fixed airflow obstruction and COPD patients was tested. A good accuracy of 88% was reached (chapter 4).
  ♦ Implication: Following the above determination of diagnostic accuracy, the next steps should include the prospective evaluation of the test performance in large series of patients with suspected disease, i.e. patients with an ‘intention-to-diagnose’. Such a study is currently running. Besides this, the clinical consequences of introducing the test into clinical practice should be evaluated in randomized controlled trials.

3.3 The exhaled signal

- One of the hallmark features of asthma, variability of airways obstruction and caliber, showed no direct influence on the exhaled breathprint as measured by electronic nose (chapter 5).
- On the other hand in COPD, inflammatory markers in sputum are highly associated with both electronic nose exhaled breathprints as well as a list of individual exhaled volatile organic compounds (VOCs) (chapter 6).
  ♦ Implication: Although asthma and COPD show large differences in the exhaled breath profile (see also chapters 3-5) as well as in type and activity of airway inflammation, the above suggests that breath profiles in asthma are also associated to airway inflammation. However, this has to be confirmed in other studies that are currently undertaken.

3.4 Subphenotyping COPD

- Exhaled molecular profiles have additional value in subphenotyping a heterogeneous disease such as COPD into clusters that may help to improve patient care and provide directions for future research (chapter 7).
Implication: As exhaled breathprints are associated to airway inflammation (chapter 6) and contribute to the identification of subphenotypes of COPD (chapter 7), this may lead to the identification of subgroups of COPD patients most likely to benefit from treatment and to design clinical trials towards these phenotypes.

3.5 Methodological validation

• For widespread application of breath testing by electronic nose technology, some methodological issues have to be overcome and clarified. We showed that electronic nose signal stability over time is high, implicating that centralized analysis of exhaled breath samples in multi-center studies is feasible (chapter 8).

Implication: Large scale studies designed to investigate the value of exhaled breath testing can now be undertaken with breath testing at a single centralized center. This overcomes the issue of inter-device reproducibility, and at the same allows for simultaneous analysis of breath samples using different types of sensor technologies including GC-MS. In the future, this may lead to the selection or development of specific sensors for specific diseases and towards a reliable translation of one electronic nose into another, making centralized analysis redundant again.

3.6 Beyond airway disease: Validation in pulmonary embolism

• In a proof-of-principle study, we showed that exhaled breath testing using an electronic nose may increase the exclusion of patients at high risk for pulmonary embolism (PE) without additional imaging (CT) testing, especially when combined with a commonly used clinical decision rule (chapter 9).

Implication: The presence of acute pulmonary embolism is reflected in the exhaled molecular profile, but is masked when other diseases with a specific molecular profile are also present in the same patient. To overcome this problem, a large training set of PE patients without comorbidity should be constructed to avoid masking of the signal. This ‘classical’ PE signal should then be tested in a group of PE-suspected patients with comorbidities. Emphasis should be put on exclusion of the disease, in order to use the breath test in a ‘screen-and-test’ procedure. All subjects, in which the disease could not be excluded with high certainty, should undergo additional imaging testing as in current clinical practice.

4 DIAGNOSTIC VALIDATION IN OBSTRUCTIVE AIRWAY DISEASES

4.1 Diagnostic validation in medicine

Diagnostic tests form the basis of medicine, from which clinical management strategies are directed. As such, the quality of a diagnostic test determines for a large part the quality of care that follows the test result [3]. In order to optimize clinical care, we should start at optimizing diagnostics by overcoming methodological flaws and difficulties in diagnostic research
allowing the introduction of reliable evidence-based diagnostics. However, there are no formal requirements for adoption of new diagnostic tests into routine clinical care, in contrast to new treatments or drugs [3]. Besides this, most research has focused on the development of new treatments and on aetiology of diseases, leaving the connecting step of diagnostics underexposed.

Because the methodology of diagnostic research lacks uniformity, several groups have made recommendations for providing study designs, specific options, requirements and possible pitfalls [3-7]. Although each of the groups is tackling this from a slightly different angle, the general message is shared. Studies for diagnostic accuracy should be designed following a common basic structure, starting at the comparison of extremes (diseased vs healthy), to a broader spectrum of the disease, to patients that exhibit clinical varieties or comorbidities, and ending at prospective evaluation of large series of subjects with an intention-to-diagnose.

4.2 Internal validation in obstructive airway diseases

The first step in assessing the diagnostic accuracy of a new test, or a new application of an existing test, is to determine the internal validity. We performed such an internal validity study for eNose breath testing in chapter 3 to discriminate between patients with ‘classic’ asthma and COPD, and smoking and non-smoking healthy controls. Not only did we confirm other studies showing a distinct exhaled molecular profile in asthma [8-10] and COPD [11-14] as compared to healthy controls, we also showed that the two obstructive inflammatory airway diseases each have a specific exhaled molecular profile, enabling discrimination between the two diseases.

A few methodological issues should be kept in mind when interpreting data from diagnostic accuracy studies. As we showed in chapter 3, the accuracy after internal validation for the discrimination between asthma and COPD by eNose was very high, reaching 96%. However, when COPD was compared to (smoking) healthy controls, a significant but low accuracy of 66% was reached. When asthma was compared to healthy controls, an accuracy of 95% was reached. This indicates that asthma has a distinct exhaled signal, whereas for COPD and healthy controls a large dispersion in exhaled profile is found. This makes the current technique suitable for identification of asthma, but not of COPD. On the other hand, we showed in chapters 6 and 7 that specific inflammatory mediators in COPD are highly associated with exhaled breathprints, and that subphenotypes of COPD with associated breathprints could be identified.

The second issue that needs to be considered when interpreting data from internal validity studies is the possible introduction of bias when constructing the diagnostic model. This may have impact on further studies of external validity in which the same model is being tested. We took several measures to overcome and prevent bias, the most important being the use of internationally standardized and accepted guidelines for characterization of subjects [15,16]. In fact, not all types of bias should necessarily be prevented, such as spectrum and selection
bias. This refers to the study population consisting of patients in a different clinical spectrum than the population in which the test is to be applied [3]. Indeed, the validation of breath molecular profiling using an electronic nose or other high-dimensional techniques such as GC-MS requires a two-step approach. First, a diagnostic model is constructed. This model is then tested in the same (internal validity) and newly recruited (external validity) groups of patients. The groups on which the diagnostic model is based, the training set, should be 1) as large as possible and 2) consist of patients with ‘classic’ disease without interfering signals from other diseases or medication that could possibly mask the disease signal. In later stages of the (external) validation, the test should be evaluated in a broader spectrum of the disease and in patients that do exhibit possible masking signals.

4.3 External validation in obstructive airway diseases
The second step in the assessment of diagnostic accuracy of the electronic nose for asthma and COPD was to determine the external validity, as we reported in chapter 4. To this purpose, a new (validation) set of patients with asthma and COPD was recruited in a different hospital. Patient groups were similar to the groups in the training set with respect to demographic and functional parameters to ensure a true external validation. The diagnostic algorithm for the electronic nose that was constructed in the internal validity study was tested in this newly recruited group of patients, showing a good discrimination between the two diseases with an accuracy of 83% and associated sensitivity of 91% and specificity of 90%. This indicates that the discriminatory potential for asthma and COPD of exhaled breath profiling by an eNose is valid and ‘transferable’ to a different clinical setting. Transferability is the ability of study results to transfer to situations with similar parameters, populations and characteristics, in this case a Dutch teaching hospital with the inclusion of similar patient groups [17].

External validity was further tested by the inclusion of a difficult-to-diagnose group of patients consisting of asthmatics with fixed airways obstruction, functionally resembling COPD. In clinical practice this group poses a day-to-day diagnostic challenge, as gold-standard functional assessment does not discriminate between the diseases. However, diagnosing these patients correctly will improve their disease outcome by being able to start the right therapy. In chapter 4 we described the findings of the external validation in the group of ‘fixed’ asthma. By using the diagnostic model that was based on classic reversible asthma and COPD, the electronic nose consequently classified patients with fixed asthma as having asthma and not as having COPD. By the inclusion of clinically challenging groups of asthma and COPD patients with overlapping spirometry results, we showed that the results of the breath test can be generalized towards a different and clinically relevant population.

4.4 Appraisal
Studies of validation of diagnostic accuracy are essential in the introduction of a new diagnostic test into medical practice. The common basic structure of such evidence-based evaluation
studies requires that the test under investigation is compared to the results of a reference test, or ‘gold standard’ test [5]. However, ‘gold standards’ rarely provide full certainty, or 100% accuracy. When a new test is compared to this ‘gold standard’ test, by definition it will be less accurate, or at best, equal to the ‘gold standard’. New tests superior to the prevailing ‘gold standard’ test may be developed, but would be ignored because they do not agree with the standards [3]. Results of validation studies should be read with this in mind, and the ‘gold standard’ test should receive the same critical appraisal as the test under investigation. In parallel, other ways of assessing the accuracy of new tests should be considered, such as the combination of several reference tests, the response to treatment and prolonged follow-up times.

The trade-off from 96% accuracy for internal validity to 83% for external validity in our studies most likely reflects the heterogeneity of both diseases. For asthma, heterogeneity mainly lies in the presence of atopy, age of onset, predominant inflammatory cell type, presence of associated comorbidities such as nasal polyposis and type of medication. For COPD, subphenotypes of chronic bronchitis, emphysema and a milder mixed type can be identified (see chapter 7). To overcome influences of heterogeneity on the discriminating signal for the diagnostic model, efforts should be put in creating large training sets that include the different disease subphenotypes.

5 THE EXHALED SIGNAL
After showing the ability of exhaled breath profiling by electronic noses to discriminate between asthma and COPD in chapters 3 and 4, the question of what constitutes the exhaled signal in both diseases was raised. Because electronic noses assess the complete mixture of volatile organic compounds (VOCs) in breath, no direction for the type of pathophysiological pathway could be indicated as the primary differentiating signal. Besides this, possible interfering signals also needed to be identified in order to correct for these possibly confounding factors [18]. Hence, several dimensions of the diseases have been investigated in relation to the exhaled signals in chapters 5 and 6.

5.1 Disease-associated factors
Disease-associated factors are likely contributors to the exhaled signal. The most striking differences between asthma and COPD are the dynamics of airway caliber, and the type and activity of airway inflammation [19-23].

5.1.1 Airway caliber
Airway narrowing is one of the hallmark features of asthma [15,24]. It has been reported that airway narrowing leads to a decrease in exhaled nitric oxide (NO) in asthma [25,26], but another study could not confirm this [27]. Furthermore, it has been shown that an acute increase in airway caliber by salbutamol elevates NO [28], but this finding was also not consistent [26,29].
It was not known whether other exhaled components, such as VOCs, are also influenced by airway caliber in asthma. The study in chapter 5 describes the effect on electronic nose breathprints of acute changes in airway caliber during methacholine provocation in asthmatics. Our study was the first to show that an acute change in airway caliber was not associated with changes in molecular breath profile. The implication of this finding is that corrections for the degree of airway obstruction are not required for eNose breathprints. Secondly, this emphasizes that the differences between asthma and COPD in exhaled molecular profiles as measured by electronic noses cannot be explained by a difference in airway caliber. As such, spirometry and eNose can be seen as complementary measurements in the assessment of obstructive airway diseases. This is confirmed in chapter 4, in which we showed a difference in breathprints from patients with fixed asthma and patients with COPD, despite a similar degree of airway obstruction. A more likely explanation for the unique breathprints should be sought in disease activity and/or underlying inflammation. This finding also paves the way for implementation of exhaled breath measurements in the monitoring of asthma and other diseases with variable airways obstruction by repeated eNose assessment.

5.1.2 Airway inflammation

Airway inflammation plays an important role in asthma as well as in COPD [19,20,23,30]. Although the airway inflammation in asthma and COPD is complex, both diseases are characterized by eosinophilic and neutrophilic infiltration of the airway walls, and these inflammatory cells can also be found in sputum [31-34]. In COPD inflammation predominantly involves the small airways and the lung parenchyma, and is often of the neutrophilic type. In contrast, in asthma the larger airways are most affected and the inflammation is typically of the eosinophilic type [23,30]. Even in fixed asthma, a disease state functionally resembling COPD, airway inflammation was found to be different from COPD [19-21], providing a potential target for diagnostics. As we showed in chapter 4, breathprints of fixed asthma patients were consistently classified as 'classic' reversible asthma and not as COPD. This further increased the probability that exhaled breath molecular profiles are mainly determined by the type, activity and perhaps location of airway inflammation rather than airway caliber per se.

In chapter 6, the first study assessing the relationship between exhaled molecular profiles and markers of airway inflammation is reported. We showed indeed that in mild to moderate COPD, the exhaled molecular profile was associated with eosinophilic and neutrophilic inflammation. Interestingly, molecular profiles measured by GC-MS were not only related to the presence of inflammatory cells, but also to the activation of these cells as measured by eosinophilic cationic protein (ECP) and myeloperoxidase (MPO) in sputum. eNose technology was more suitable for the determination of activation of inflammation (ECP and MPO), especially in mild COPD. These results indicate that molecular patterns in exhaled breath reflect airway inflammation in COPD, which is already present in relatively early stages of the disease. It is likely that this is also the case for airway inflammation in asthma, although this remains
to be investigated. This study has a clinical implication. Several studies have shown that selection of COPD patients based on inflammatory type leads to a better treatment response [35-37]. Exhaled breath profiling may be used as a non-invasive way to assess inflammatory type and thereby treatment response. To this purpose, diagnostic management studies should be conducted.

5.2 Treatment-associated factors

Differences in exhaled molecular profile may not be caused solely by disease-associated differences. When interpreting results from diagnostic studies, either single marker tests or high-dimensional tests such as breath profiling, possible confounding factors should not be disregarded. The exhaled profile may be influenced by factors such as medication use. The prevailing treatment of both asthma and COPD consists of inhaled corticosteroids and inhaled short- and long-acting bronchodilators. These medications may exert influence on exhaled profile in two ways. First, they may impact the molecular composition of breath by their local and systemic effect, modifying inflammatory and metabolic pathways. Second, the mode of administration by inhaled aerosols may directly influence the exhaled air sample by containing (parts) of the drug. In chapter 5, we showed that similar alterations in eNose exhaled breathprints were found after challenges with methacholine and isotonic saline. This indicates that the change in breathprint might have been caused by the nebulisation procedure itself, implicating that patients should be instructed not to use their inhalator minutes before exhaled breath measurements. In the studies described in chapters 3 and 4, patients were asked to refrain from taking inhalation medication 3 hours prior to the test. Diagnostic accuracies for asthma and COPD in these studies were not influenced by the use of inhaled corticosteroids (ICS) or long-acting bronchodilators. Therefore, we may conclude that a 3 hour period is sufficient to avoid influence of aerosols on the breathprint, and that asthma and COPD breathprints are not influenced by the chronic use of ICS or long-acting bronchodilators.

5.3 Patient-associated factors

Breathprints may also be affected by patients-associated factors, such as age, comorbidities or current smoking.

5.3.1 Age

It has been shown that with increasing age in healthy controls, the composition of the exhaled air with respect to alkanes and monomethylated alkanes changes [38]. These findings are consistent with the finding of increased oxidative stress in ageing, as well as a decrease in the cytochrome p450 clearance [39,40]. On the other hand, the discrimination between healthy controls, asthmatics and COPD patients by an electronic nose was not influenced by a difference in age, as described in chapters 3 and 4 [8,41,42]. Older asthmatics could be separated from COPD patients with a similar accuracy as young asthmatics [8,42]. Furthermore, eNose
breathprints from old and young asthmatics were indistinguishable [8,42], and old and young controls also could not be discriminated [8]. Taken these findings together, this suggests that although the individual components of exhaled breath change with increasing age, this does not affect the overall profile as measured by electronic nose.

5.3.2 Comorbidity
This thesis describes the value of exhaled breath profiling in asthma and COPD. Taken the fact that this seems feasible with high accuracy, one can easily reason that other diseases might also be detectable in exhaled air. If so, the presence of comorbidities may have influence on the exhaled breathprint, even masking the signal of asthma, COPD, or any other signal of interest. Indeed, several studies have been conducted that show the presence of specific fingerprints of different diseases in exhaled breath. To name a few: diabetes [43-47], many forms of cancer [13,48-53], renal failure [54-57], heart disease [58], bacterial infections of the upper [59,60] and lower [61-63] airways, and tuberculosis [64-67]. This possible interference should be kept in mind when designing studies for determination of diagnostic accuracy for exhaled breath testing. To minimize influence of comorbidity on the diagnostic model and thereby the validity, studies may be designed in two ways. The first option is to construct a training set on which the diagnostic model is based, containing only ‘pure’ i.e. non-comorbid subjects. We showed this option in patients with acute pulmonary embolism, described in chapter 9. Indeed, accuracy for discriminating acute PE from non-PE in suspected patients dramatically increased after separation of the training sets into non-comorbid and comorbid subjects. The second option is to construct a training set of patients with the disease under investigation with mixed comorbidities. When this training set is considerably large, the signal of the disease under investigation will be the common factor in all breathprints, allowing detection. This will however also worsen bias caused by confounding. Factors that happen to correlate with the class discrimination, such as risk factors associated with the disease of interest, will also be discriminatory [68].

5.3.3 Smoking
Cigarette smoke contains toxic and carcinogenic compounds, such as 2,5-dimethylfuran, acetonitrile, benzene, toluene and styrene [69]. When interpreting exhaled breath measurements in current smokers, one should bear in mind that breathprints may be influenced in two ways. First, volatile organic compounds in an exhaled breath sample may not be derived from the body but rather from the cigarette itself. This direct influence on the breathprint can be avoided by asking subjects to refrain from smoking for a certain period before an exhaled breath sample is taken. In our studies, a two-hour period (chapters 3 and 4) appeared to minimize the influence of smoking in COPD patients. Second, it has been shown that cigarette smoke increases oxidative stress in the human body by increasing the amount of free radicals [70,71]. This may lead to the production of oxidative stress-related molecules directly and indirectly
by promoting airway inflammation by activation of eosinophils and neutrophils [72] as we showed in chapter 6 for COPD. In COPD patients, exhaled breathprints showed no difference between current and ex-smokers, as shown in chapters 3, 4 and 6. However, in healthy controls, differences between current and non-smokers are present, as we showed in chapter 3. Therefore, patients in diagnostic studies concerning exhaled breath profiling, should always be matched with their healthy smoking and/or non-smoking counterparts to be able to eliminate the influence of smoking on the test results.

6 SUBPHENOTYPING COPD
COPD is a very heterogeneous disease, and attempts to describe the disease by single disease parameters such as lung function and CT lung density measurements alone have not led to better management of the disease. Therefore, a new taxonomy has been proposed to describe COPD [73], based on patterns of the disease with respect to the different domains: functional, clinical, patient-related, radiological, inflammatory etcetera. The unbiased statistical approach for identification of meaningful subphenotypes of COPD that is described in chapter 7 revealed three distinct clusters of the disease in a community-derived population that differed with respect to severity of airflow limitation, radiologic lung density, presence of comorbidities such as cardiovascular diseases and diabetes, symptoms of dyspnea and sputum production and, interestingly, exhaled breath molecular profile. Cluster analyses aiming at identifying subphenotypes of COPD have been performed previously [74-79]. The clusters that were identified in chapter 7 confirmed and extended the subphenotypes of COPD that were found in those previous studies in a clinical setting, thereby serving as an external validation in early or mild cases of COPD in a primary care setting. Additionally, the clusters that were identified showed distinct exhaled breath molecular profiles, indicating that this technique may be useful to assess such subphenotypes by a fast and non-invasive breath test. Combining the results of the studies presented in chapters 6 and 7 suggests that subphenotypes of COPD that can be identified using exhaled breath molecular profiling may be based on the type and activation of airway inflammation.

7 METHODOLOGICAL VALIDATION
7.1 Repeatability
An essential premise for diagnostic tests is the repeatability. Repeatability can be divided into four forms.
1. Instrumental repeatability: when the same sample of exhaled breath is measured several times, there should be no difference between breathprints. In chapter 8 we showed that instrumental repeatability for both electronic noses and GC-MS is high. Others have shown similar results, for GC-MS [80,81] and ion mobility spectrometry [82].
2. Intra-subject repeatability, intra-day: breath samples taken consecutively should not be distinguishable. Indeed, in chapters 3 and 8 we showed that there was no difference between the first and second electronic nose and the first and second GC-MS measurements taken 5-10 minutes apart in healthy controls, asthmatics and COPD patients. This confirms findings in other studies [80-82].

3. Intra-subject repeatability, between-days: breath samples taken days to months apart from the same subjects should preferably only change as a result of change in disease state, and should not be influenced by sensor drift of the device and subject-related variables such as diet, hormonal cycles, environmental influences and other non-disease related processes. In chapter 3 it was shown that in healthy smoking and non-smoking subjects, breath samples taken 1-48 days apart were indistinguishable by electronic nose. In chapter 5 the same stability of exhaled breathprints was shown for asthmatics. Others have come to similar results using GC-MS analysis [14,80-82].

4. Classification (accuracy) repeatability between groups: this type of repeatability is essentially the same as internal and external validation, as described in chapters 3 and 4. It includes both the long-term stability of the discriminatory signal between groups and the long-term device repeatability.

7.2 Breath collection

The quality of the breath sample is highly dependent of the method of sampling. A list of potential confounders can be thought of in the light of exhaled breath testing, including the VOCs of the researcher taking the sample, environmental VOCs, the material of the breath collection equipment and/or the disinfectants used. To avoid influence of these exogenous VOCs, we used an inspiratory VOC-filter in all of our studies. Besides this, baseline samples of the inspired air were taken and used to correct for any environmental influences.

The relative humidity of exhaled breath is higher as compared to ambient air samples. As humidity may influence the concentration of certain volatiles [83], and because sensors of the polymer electronic noses that were used in most of our studies are extremely sensitive to humidity [84], we controlled the humidity of the samples by using silica filters at the expiratory port of the breath capturing device. A relative humidity of 30-40% proved optimal for this type of electronic nose, as well as in the GC-MS analyses that were performed in the studies described in this thesis.

Breath can be sampled either directly on-line, by blowing into the device, or indirectly off-line by capturing breath in a bag or container. We employed indirect breath sampling by means of Tedlar® (polyvinylfluoride) and Nalophan® (polyethyleneterephthalate; PET) bags in all studies. Advantages of the indirect method include the possibility of repeated measurements from the exact same air sample and the possibility to control the flow, volume, humidity and concentration of the air that is being sampled. Disadvantages are the delay from the time of sampling to the time of analysis, and the possible influence on the composition of the
breath sample by diffusion of certain components into or out of the bag [85]. Direct on-line assessment of samples requires little sample preparation, is user-friendly and quick. However, not all electronic nose devices are suitable for direct measurements due to limited ranges for humidity, flow and volume and required pre-concentration steps.

7.3 Breath storage
Although all studies described in this thesis have employed the immediate measurement of breath samples rather than store them, storage is an important item to be addressed in order to apply this technique on a large scale in multi-center trials. Until the restriction of eNose compatibility or translation between devices has been solved, studies will rely on centralized measurements of breath samples. In chapter 8, we described the stability of the exhaled signal for the discrimination between asthma and healthy controls over a 2 week storage period on Tenax® desorption tubes as to mimic transportation times between centers. Offline gas sampling after (prolonged) storage is already often applied in other than medical fields, such as environmental research and indoor air quality assessment [85]. For individual components, storage may have an effect on stability by several factors including diffusion through the bag in which breath is initially captured, bag surface-induced reactions, reactions with other compounds in the gas mixture and diffusion into the bag of exogenous compounds. Previous studies have focused on single compounds, and have shown a wide variability in stability when stored on desorption tubes, varying from 4 weeks for chemically active compounds [86] to up to a year for inactive compounds [87]. Combining the results of these studies and our study described in chapter 8, focusing on the complex VOC mixture of breath, it seems feasible to store breath samples on Tenax® desorption tubes for at least four weeks.

7.4 Data analysis
The sensor array of an electronic nose provides a pattern of responses when exposed to a gas mixture. Data acquisition is the first step for data analysis, in which all data is converted to an electrical signal pattern that can be processed by pattern recognition techniques [88]. The output is a pattern vector that is passed into the second step, feature extraction. Principal component reduction is the most commonly used type of feature extraction resulting in dimensionality reduction. New variables, called principal components, are created that describe the variance within the dataset and are not correlated to each other. These principal components are fed into classification algorithms to examine the differences and similarities between patterns leading to the smallest classification error [88]. Linear discriminant analysis is the pattern recognition technique that is commonly used in electronic nose data analysis, and is also used in the studies described in this thesis. Grossly, classification techniques can be divided into two types: unsupervised and supervised analyses. Cluster analysis methods such as hierarchical and k-means cluster analysis are examples of unsupervised analyses, making no a priori assumption for classification. Discriminant analysis is an example of supervised
classification, in which a variable determines class membership. Supervised methods can be subdivided into methods that assess predictive values of group membership, such as regression techniques, and methods that assess the distribution of and interaction between variables resulting in accuracy, sensitivity and specificity. Examples of conventional supervised classification algorithms include principal component regression, discriminant factor analysis, analysis of variance between groups (ANOVA), partial least squares, support vector machine, logistic regression and artificial intelligence techniques including neural networks and fuzzy logic [80,88]. The variety of analysis methods for eNose data emphasizes that there is much to gain – or to lose – by the choice of classification algorithm [88]. Discriminant analysis for instance is appropriate for normally distributed data such as the data in the studies in this thesis, whereas support vector machine makes no assumption about data distribution and may better fit a dataset with non-normally distributed data.

Exhaled breath testing using eNoses is especially suitable for exclusion of disease, with optimization of sensitivity. A trade-off is made against a relatively high percentage of false-positives, which will have to undergo further diagnostic testing. The number of false-positives should be kept as low as possible, by choosing the most appropriate algorithm and breathprint cut-off value. In addition, one of the main causes of false discoveries in general (i.e. finding a discrimination between classes when there is in fact no difference) is inadequate sample size [68]. This error typically occurs in omics studies, when the number of variables largely exceeds the number of patients. Therefore, studies should be conducted with sufficient sample sizes. Besides this, the discriminatory potential of the model should be checked for its accuracy in a subset of the data that was not used in the generation of the model; the validation set. This can be done in three ways [68]. Internal validation/cross-validation is usually performed using the leave-one-out (jackknife) method. This means that for $N$ times ($N =$ number of patients), the predictive model is constructed on all data except for one point and a prediction is made for that point. Similarly, this can be performed in a 2:1 or 1:1 ratio for training and validation sets from the same population. Second, bootstrapping can be performed. Using this technique, a random sample of $n$ patients can be extracted from $N$ sampled patients, where each patient can be selected several times. Thereby, a large number of ‘virtual’ datasets can be created and validated. Third, an external validation procedure can be performed. This involves the inclusion of a completely new and independent validation set, from another population, as we described in chapter 4. The latter technique closely approaches application in clinical practice and is seen as the gold standard of validation. In this thesis, all steps described above have been undertaken and are described in chapters 3 and 4.

8 BEYOND AIRWAY DISEASE: VALIDATION IN PULMONARY EMBOLISM

Given the good accuracy for discriminating between airway diseases such as asthma and COPD in our own studies, and the abundance of studies targeting diseases such as lung cancer
[13,49,89-91] and tuberculosis [64-67], we expanded the focus to pulmonary embolism (PE). Exhaled breath testing is particularly suitable as a first-line test in a screen-and-test procedure. Aiming at a high sensitivity and negative predictive value to exclude disease may eventually lead to a reduction in the amount of patients suspected of acute pulmonary embolism that have to undergo CT imaging. Reducing the amount of CT scans that has to be performed to exclude acute PE will lower the risk of radiation associated side-effects. This holds especially for the increase in breast cancer in young women associated with the use of spiral CT-scanning [92,93].

In chapter 9, we described a proof-of-principle study aiming at the discrimination of CT-confirmed absence and presence of acute pulmonary embolism in patients with suspected disease. Predefined subgroups were made with respect to comorbidities known to have their own specific breathprint, including diabetes [43-47], any form of cancer [13,48-53], renal failure [54-57] and heart disease [58]. In the group with comorbidities, PE could not be distinguished from non-PE. On the other hand, in the group without comorbidities, there was a clear signal for PE in exhaled breath. When combined with the Wells score, a commonly used clinical decision rule [94], sensitivity even further increased. The high sensitivity in the non-comorbid group indicates that exhaled breath testing may already be of value to exclude PE in the group that is most likely to benefit from a reduction in CT-scans, namely young women [93]. Sensitivity in the comorbid group may be increased by constructing a large training set consisting of ‘pure’ PE patients providing a solid reference standard of PE breathprints. The next step would be to test this model in an external validation study with prospective enrollment of patients suspected of PE, and eventually, to design a diagnostic management study [5,68].

9 DIRECTIONS FOR FUTURE RESEARCH

9.1 Validation

Unlike newly developed drugs and therapies, there is no formal standard for acceptance of diagnostic tests into routine care [3]. Until health authorities adopt such formal standards, efforts should be put into careful validation of diagnostics following the current guidelines [5]. The results described in this thesis are the first steps towards introduction of exhaled breath molecular profiling for obstructive airway diseases in medical practice. The necessary next steps would be:

1. To assess the accuracy for patients with suspected disease, i.e. with an intention-to-diagnose. This study is currently running for asthma and COPD.
2. To assess the influence of different comorbidities on exhaled breathprints. This may be divided into comorbidities that are related to the disease under investigation, such as nasal polyposis and gastro-esophageal reflux disease for asthma, and cardiovascular disease and diabetes in COPD, and comorbidities that are unrelated to the disease.
2. To assess the value of exhaled breath profiling in monitoring of airway disease and early detection of exacerbations. This study is currently running for asthma. When this shows to be feasible, a clinical management study should be performed to assess whether outcomes are improved by early detection of exacerbations and thereby early start of treatment.

3. To assess the possibility of prediction of treatment response. In asthma and COPD, not all patients respond to first or second line treatment. Side-effects and unnecessary treatment may be avoided when it is possible to predict the response to treatment beforehand.

Cutoff values of the test for diseased versus healthy or between disease states may be adjusted depending on the goal of the test (Table 1). In a screening setting, a high sensitivity is preferred to exclude disease. In other words, a high negative predictive value is recommended and a trade-off is made with respect to the amount of false-positives that will have to undergo further diagnostic testing. On the other hand, a high specificity and positive predictive value is warranted in confirmation of diseases, mostly in a clinical setting.

9.2 Methodology
Although exhaled breath molecular profiling has been applied in the medical field for more than three decades, there are still some methodological issues to be solved or improved before this technique can be implemented on a large scale in clinical practice.

A particular issue that needs to be addressed is the translation between different electronic nose devices. Naturally, devices operating with different sensor techniques provide different kinds of exhaled profiles. However, electronic noses with identical sensor technologies also provide different profiles. As a consequence, each electronic nose device should be trained and validated separately. To overcome this problem, efforts should be made in the statistical translation of one device into another [88,95,96], or to the design of sensors that provide identical profiles [97].

A second issue that needs further investigation is the identification of specific sets of volatile organic compounds (VOCs) in different diseases by gas chromatography and mass spectrometry (GC-MS). This will not only provide information about which pathophysiological pathways are involved, possibly leading to new targets for therapy, but it will also allow technicians to construct sensors specifically sensitive to these compounds. Eventually this may lead to the development of tailored electronic noses for different applications.

Finally, the method of breath collection needs optimization and standardization. On the one hand, breath collection should be as ‘clean’ as possible, avoiding influence of environmental and breathing setup derived VOCs. On the other hand, efforts should be put into designing a device that is easy to use for both patients and physicians, such as the breath ethanol test being used by the police. The golden mean should be found to optimize device performance and clinical usefulness.
### Table 1 Diagnostic tests: concepts.

<table>
<thead>
<tr>
<th>Disease prevalence</th>
<th>Goal</th>
<th>Negative predictive value (NPV)</th>
<th>Positive predictive value (PPV)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>False-positives</th>
<th>False-negatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula</td>
<td>Cases/1000</td>
<td>-</td>
<td>( \frac{\Sigma \text{True negative}}{\Sigma \text{Index test negative}} )</td>
<td>( \frac{\Sigma \text{True positive}}{\Sigma \text{Index test positive}} )</td>
<td>( \frac{\Sigma \text{True positive}}{\Sigma \text{Gold standard positive}} )</td>
<td>( \frac{\Sigma \text{True negative}}{\Sigma \text{Gold standard negative}} )</td>
<td>( \text{Cases index test positive, gold standard negative} )</td>
</tr>
<tr>
<td>Screening setting</td>
<td>Low</td>
<td>Exclude disease</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Often present</td>
</tr>
<tr>
<td>Clinical setting</td>
<td>Medium</td>
<td>Detect disease</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Present</td>
</tr>
<tr>
<td>Top clinical setting</td>
<td>High</td>
<td>Confirm disease</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Usually absent</td>
</tr>
</tbody>
</table>
REFERENCES


