

Scientific Report for the study “*breath-gas profiles of healthy volunteers*”
performed at the “*Arbeitskreis Vorsorge- und Sozialmedizin*” (AKS) in
Bregenz, Austria

Study duration: 3 months (Jul 15 - Oct 15, 2007)

Günther Diem, Hans Concini, Karl Unterkofler and Anton Amann

11. Dez 2007

Abstract.

Background: A pilot study has been carried out to define typical characteristics of the trace gas compounds in exhaled breath of non-smokers and smokers to determine normal concentration values in healthy volunteers. Such normal values will be helpful for interpretation of breath analysis data from patients who suffer from various diseases, such as lung cancer. As volunteers we recruited healthy person undergoing a health check-up in Vorarlberg (Austria). The volunteer recruitment was organized by the “Arbeitskreis Vorsorge- und Sozialmedizin” (AKS) in Bregenz, Austria.

Methods: Exhaled breath was analyzed using Proton Transfer Reaction-Mass Spectrometry (PTR-MS) for 219 volunteers (50 smokers, 166 non-smokers, 3 ex-smokers).

Results: The median concentrations for formaldehyde were ~4 ppb, for methanol ~190 ppb, for acetonitril ~5 ppb (non-smokers) and ~40 ppb (smokers), for acetone ~500 ppb, for isoprene ~200 ppb, for benzene ~0.7 ppb (non-smokers) and ~ 1.8 ppb (smokers) and for toluene ~5 ppb. Volatile organic compounds corresponding to product ions at twelve mass-to-charge ratios (m/z 28, 42, 67, 71, 79, 81, 83, 93, 97, 109, 123, 137) in the PTR-MS spectra differentiated between smokers and non-smokers. The Youden-index (= maximum of sensitivity+specificity-1, YI) as a measure for differentiation between smokers and non-smokers was YI= 0.30 for ions at the m/z values 28 (tentatively identified as HCN), YI= 0.68 for $m/z=$ 42 (tentatively identified as acetonitrile) and YI= 0.60 for $m/z=$ 79 (tentatively identified as benzene). No statistically significant difference between smokers and non-smokers was observed for the product ions at $m/z=$ 31 and 33 (compounds tentatively identified as formaldehyde and methanol).

Conclusions: When interpreting the exhaled breath of lung cancer or COPD patients, who often smoke, compounds appearing at the above-mentioned 12 mass-to-charge ratios should be considered with appropriate care to avoid misdiagnosis. Validation studies in larger numbers of patients with more precise delineation of their smoking behavior and using additional analytical techniques such as GC/MS and SIFT-MS should be carried out.

Keywords: breath analysis, exhaled air profile, smoking, volatile organic compounds (VOCs), proton transfer reaction mass spectrometry (PTR-MS).

1. Introduction

The biochemical and molecular biological diagnostic methods used in medicine have developed very rapidly in recent decades, the main focus being on blood and urine analyses. Clearly, these methods require that a sample of the body fluid be taken, which is quite invasive, especially for blood sampling, and uncomfortable for the donor. Nevertheless, these analyses are an essential, irreplaceable tool in clinical diagnosis. Blood analysis is usually concerned with the large molecular weight non-volatile compounds such as proteins and ions that are present and not with the low molecular weight volatile species that are mostly lost when a blood sample is taken. However, in recent years it has emerged that these volatile compounds will be present in exhaled breath, some now known to be present at trace levels in the parts-per-million (ppm) and parts-per-billion (ppb) levels, and that these can be valuable indicators of metabolic status and can distinguish between the healthy and diseased state if their levels can be measured to an acceptable accuracy [1]. So the major challenge is to be able to identify and quantify these volatile compounds to sufficient accuracy to be useful in diagnosis. This is now becoming possible, due to remarkable developments in gas analysis techniques and in sampling methodology that have occurred during the last decade. Routine breath analysis would be a valuable addition to the armoury of the clinician, especially since breath sampling is a non-invasive technique, totally painless and agreeable to patients and can be achieved in real time with the results immediately available to the clinician. However, at this time, breath analytical techniques are not so well developed as those for urine and blood analyses and thus are not yet widely utilized in clinical practice, but this new addition to medical diagnostic techniques is gaining momentum as the potential of breath analysis is being realized and more scientists and clinicians are becoming involved in this area of science and medicine. To date, the United States Federal Drug Administration (FDA) has approved only the following compounds for breath testing:

- ethanol (law enforcement)
- hydrogen (carbohydrate metabolism)
- nitric oxide (asthma)
- carbon monoxide (neonate jaundice)
- $^{13}\text{CO}_2$ (*H. pylori* infection)
- branched hydrocarbons (heart transplant rejection).

This list is still quite limited even though it has been established in pilot studies that some other breath metabolites, when elevated above normal, are indicative of disease, for example, elevated breath ammonia occurs in kidney and liver dysfunction [2] and breath acetone is elevated in diabetes [3]. Also, there is clear evidence that breath isoprene is related to cholesterol biosynthesis [4] and there are strong indications that the endogenously produced ethanol that is seen in exhaled breath is associated with gut bacterial overgrowth [5]. These pilot studies will be referred to again later. A reason for the few FDA approvals is mostly because of the stringent requirements laid down by the FDA: sampling procedures are critically investigated and measurement reproducibility is considered as paramount to avoid misleading artefacts.

Breath analysis is very attractive, because of its non-invasive nature and because it can easily be realized for sick patients, including children and elderly persons. Multiple breath samples can be collected into bags or onto traps, and now on-line breath sampling can be achieved with the advent of new experimental techniques. Of these new techniques, selected ion flow tube mass spectrometry (SIFT-MS), proton-transfer reaction mass spectrometry (PTR-MS), laser spectrometry and ion mobility spectroscopy (IMS) are particularly promising for real time breath analyses. Also, the well established gas chromatography mass spectrometry (GC-MS) now in conjunction with solid phase micro extraction (SPME), although slower in providing analysis, has a valuable role in breath analysis.

We expect that the non-invasive technique of exhaled breath analysis will play a considerable role for diagnosis and therapeutic monitoring in the future. Our present work is a contribution to respective clinical applications of exhaled breath analysis.

2. Methods

Sample collection

A cohort of 219 volunteers was recruited; all individuals gave informed consent for participation in the study. Demographic data are presented in Table I. The volunteers completed a questionnaire describing their current smoking status (active smokers, non-smokers) and the time elapsed since their last smoke. The classification as smoker/non-smoker/ex-smoker is based on the self-declaration of the volunteers. The amount of smoking (in packyears) has not been determined. Ex-smokers have only been considered for Fig 4, but are not used for comparisons between smokers and non-smokers. Samples of mixed alveolar exhaled breath (including dead space air) were collected in 3-litre-volume Tedlar Bags with parallel collection of ambient air (also in Tedlar bags). The samples were collected at different daytime independent of the time of meals and were processed within 24 hours at most. Before measurement, the bags were heated to 40°C for at least 15 min. For all our samples we measured CO₂-content, sorting our samples with low CO₂-concentration. The study was approved by the local ethics committee.

PTR-MS instrument used

A high-sensitivity proton transfer reaction mass spectrometer (PTR-MS, 3 turbopumps) with Teflon rings (instead of Viton rings) was used for our measurements. The count rate of primary ions (H₃O⁺) was around 10⁷ counts per second. Dwell time was 0.5 sec for each mass-to-charge ratio measured ($m/z=21$ - $m/z=230$). Typical compounds used for determination of transmission coefficients were acetonitrile, acetaldehyde, acetone, DMS, 2-butanone, benzene, toluene, p-xylene, benzaldehyde, chlorobenzene, 1,2 dichlorobenzene, 1,2,4 trichlorobenzene. These compounds do not show fragmentation (of their respective protonated form). Concentrations of these compounds were chosen in a range leading to ~10% reduction of primary counts, with subsequent observation of recovery of primary ion counts (measuring at $m/z=21$ and the specific mass-to-charge ratio of the respective non-fragmenting compound). The length of the drift tube of our PTR-MS is 9.3 cm, with an applied voltage of 600 V. The usual pressure in the drift tube was ~2.3 mbar (with slight variations). In accordance with the instructions of the manufacturer (Ionicon GesmbH, Innsbruck), we computed concentrations with using only H₃O⁺ as primary ion (not considering the first water cluster H₂O.H₃O⁺).

Mass-spectrometric analysis

Proton transfer reaction-mass spectrometry allows on-line monitoring of VOCs with volume mixing ratios as low as a few parts per trillion (pptv) [6, 7]. Chemical ionization, based on proton-transfer reactions with H_3O^+ as the primary reactant ion, is a versatile method for identification and quantification of the mixtures of organic molecules. In our study, each sample, including samples of ambient air, was measured three times with mass-to-charge ratios (m/z) ranging from 21 to 230. The median concentrations of these three measurements were used for further statistical analysis. Concentrations of compounds related to some m/z have been calculated based:

- either on a “standard” rate constant for protonation of $k = 2 \cdot 10^{-9} \text{ cm}^3 \text{ sec}^{-1}$ (for compounds which are not identified, the concentration thus being uncalibrated)
- or on specific thermal equilibrium protonation rate constants for the compounds methanol ($k = 2.7 \cdot 10^{-9} \text{ cm}^3 \text{ sec}^{-1}$), acetonitrile ($k = 5.1 \cdot 10^{-9} \text{ cm}^3 \text{ sec}^{-1}$), isoprene ($k = 2 \cdot 10^{-9} \text{ cm}^3 \text{ sec}^{-1}$) and acetone ($k = 3.9 \cdot 10^{-9} \text{ cm}^3 \text{ sec}^{-1}$). For isoprene, which apart from appearing at m/z 69 fragments to m/z 39 (~10%) and m/z 41 (~40%), we used the concentration computed for m/z 69 multiplied by a calibration factor of 2.24.

Identification of compounds is notoriously difficult with PTR-MS. Judging from our GCMS-investigations, we know that methanol, acetone and isoprene are present in almost everybody's exhaled breath and that acetonitrile arises in increased concentrations in the breath of smokers. At the respective mass-to-charge ratios (m/z 33, m/z 42, m/z 59, m/z 69) other compounds may be present, even though in low concentrations. Incidentally, protonated isoprene does not only show up at m/z 69, but partly fragments in PTR-MS to m/z 39 (~10% of protonated isoprene) and m/z 41 (~40% of protonated isoprene). For formaldehyde and hydrogen cyanide, we cannot presently rely on GCMS-measurements. The compounds dimethylsulfoxide, toluene, dimethylfuran and dimethylpyrazole in Table II are not more than an “educated guess”.

The concentrations relating to product ions at m/z 31 (tentatively identified as protonated formaldehyde) have been corrected for isotope effects from m/z 30 (= NO^+ which contributes $^{15}\text{NO} + \text{N}^{17}\text{O} = 0.37\% + 0.04\% = 0.41\%$ to m/z 31). On mass-to-charge ratio $m/z = 31$ one may also observe fragments from reaction products of ethanol and O_2^+ or of methanol and O_2^+ . Accurate absolute values for formaldehyde concentrations can only be achieved with PTR-MS by appropriate calibration measurements. The ions at m/z 43 (tentatively identified as originating from isopropanol) may partly originate from compounds other

than isopropanol. It can be both $C_3H_7^+$ (as from propanol) or CH_3CO^+ as sometimes occurs from the reactions of aldehydes, ketones and carboxylic acids [8, 9]. Ions at m/z 31 might also be fragments from reaction products of ethanol or methanol with O_2^+ .

The age effect of exhaled breath samples is negligible (apart from water, which quickly diffuses through the walls of Tedlar bags). Acetonitrile seems to diffuse quickest through bag walls, with an exponential decay constant $\tau \sim 31h^*$.

Statistical analysis

Concentrations of compounds are expected to be log-normally distributed, since the contributing physiological factors act multiplicatively and not additively. If the concentrations of compounds are log-normally distributed, the logarithms of the concentrations are normally distributed. This was tested with Lilliefors Test (with level of significance at 5%). Histograms of distributions of concentrations are therefore shown using a logarithmic concentration scale.

Since the data are expected to be log-normally distributed, the concentrations are expressed by giving medians of concentrations and geometric standard deviation (GSD), instead of mean and standard deviation (which would be appropriate parameters for normally distributed concentrations). Repeated-measures analysis of variance (ANOVA where Lilliefors Test confirmed log-normal distribution, Kruskal-Wallis otherwise) was used to compare the *logarithmic* concentrations of the different groups (smokers vs. non-smokers) [10, 11]. Statistical results were considered to be significant if $p < 0.01$. Receiver-operator-characteristics (ROC) curves [12-14] were applied to determine the thresholds for the concentrations of compounds that yielded the highest combined accuracy for distinguishing patients with the high and low concentration of definite substances. Sensitivity, specificity, as well as positive and negative predictive values were determined for these thresholds. The Youden-index was determined, which is defined to be the maximum of (sensitivity+specificity-1). We may illustrate Youden-index with some examples:

- if sensitivity= 0.8 and specificity= 0.9, Youden-index= $0.8+0.9-1= 0.7$,
- if sensitivity= 0.5 and specificity= 0.9, Youden-index= $0.5+0.9-1= 0.4$.

^{*}) Herbig, J, personal communication

We consider sensitivity, specificity and ROC-curves much more instructive than p-values: ROC-curves do not depend too much on the numbers n_1 and n_2 of volunteers in the two groups considered, whereas p-values are very sensitive to n_1 and n_2 .

Selection of data

In certain situations, the inhaled air shows a higher concentration of some compounds than the exhaled air. In such situations the corresponding concentrations of the compound in exhaled air may *not* reflect the blood concentrations of this compound (if blood concentrations are involved at all, which is not the case for a compound like, e.g., nitric oxide, which is produced in the lungs and the sinuses [15-17]). A similar caveat holds if the concentration of a compound in inhaled air is just below the concentration in exhaled air. We therefore not only considered the raw concentrations of compounds in exhaled breath, but also applied a *filter* to these raw concentrations as described in the following.

Filtering data: A value for the expiratory concentration is considered if and only if

$$(\text{inspiratory concentration})_i \leq 0.5 * (\text{expiratory concentration})_i . \quad (1)$$

Hence, the filter discards all those expiratory concentrations which are less than double the respective inspiratory concentration. For the compounds, where the concentration in exhaled air is expected to be higher (i.e., endogenous compounds from the human body), this filter works well with the factor 0.5 - if this factor is increased, more samples are added, if this factor is decreased less samples are taken into account.

For compounds, where the influence depends less on the human body (e.g. but on cigarette smoke), there have to be *exceptions* to this filter condition (1) for very low expiratory concentrations; if we compare, say non-smokers with smokers, the expiratory concentrations of some compounds in non-smokers are often so small that the indoor air concentrations (inhaled) and the expiratory concentrations are in the same range. If these expiratory concentrations are filtered out, almost all data are “lost”. Therefore, we do *not* filter out these expiratory concentrations, but have to concede that these expiratory concentrations are only *upper bounds* for the “real” expiratory concentrations (which would appear if the indoor air would be absolutely clean and free of any contamination).

To formulate the *exceptions* to our filter condition (1) in a precise quantitative way, we consider a logistic regression (setting non-smoker = 0, smoker = 1; see Figure 1) and choose the *marginal concentration* as being that particular concentration for which the logistic regression curve takes a value of 0.3 (= 30%). All expiratory concentrations below the marginal concentration are taken into account (both for smokers and non-smokers).

If this value (here 0.3) is too low, one risks that most samples of non-smokers are filtered, especially the ones with higher concentration. This results in a false decrease of the overall concentration of non-smokers and the statistics would show a larger difference than in reality exists. If this factor is too high, samples with low concentration of smokers, which were filtered out by the filter-rule (1) are taken into account and the overall concentration of smokers would decrease. Moreover, more samples of non-smokers with higher concentrations would pass the filter and would enlarge the overall-concentration of non-smokers. Therefore the statistic would show a smaller difference than in reality exists.

Nevertheless, it should be noted that these values (below 30% of the logistic regression) do not necessarily represent exhaled breath concentrations of some systemic compound in the blood of non-smokers, but possibly indoor air concentrations, only (of compounds which are just inhaled and exhaled).

We consider raw concentrations *and* filtered data. The filtering is a kind of cross-check, hinting at problems with high indoor air concentrations. For compounds with roughly equal concentrations in smokers and non-smokers (e.g., formaldehyde or methanol), the second part of the filtering process (taking into account the expiratory concentrations below the marginal concentration) is not effective, and therefore the filtered concentrations may be unacceptably high.

We *never* use differences (expired concentration – inspired concentration) and consequently never use “negative concentrations”. Whenever a VOC behaves like carbon dioxide, differences do not make sense: the concentration of carbon dioxide in exhaled air is ~4%, independent of the CO₂ concentration in inhaled air (0%, 1% or 2% in indoor air). The differences (expired concentration – inspired concentration) in concentration of CO₂ would nevertheless be very different (namely 4%, 3% and 2%) without any physiological reason for this in the body.

Receiver-operator characteristics (ROC-curves)

To differentiate between smokers and non-smokers, a threshold concentration c_0 can be chosen, non-smokers being expected to show lower concentration than c_0 and smokers being expected to show higher concentrations than c_0 . Such a threshold concentration c_0 gives rise to a corresponding sensitivity and specificity (for detection of smokers). Sensitivity is defined as the number of true positives [i.e., smoker & \geq threshold] divided by the number of all smokers. Specificity is defined as the number of true negatives [i.e., non-smoker & $<$ threshold] divided by the number of all non-smokers. If many different candidates for threshold concentrations c_0 are chosen, the corresponding sensitivities may be plotted versus the corresponding (1-specificity): this is called an ROC-curve [12, 13, 18]. The Youden-Index is the maximum of (sensitivity+specificity-1). If the sensitivity and the specificity are at 70%, the Youden-

Index is 0.4. If the sensitivity and the specificity are at 90%, the Youden-Index is 0.8.

3. Results

The median concentrations for formaldehyde were ~ 4 ppb, for methanol ~ 190 ppb, for acetonitril ~ 5 ppb (non-smokers) and ~ 40 ppb (smokers), for acetone ~ 500 ppb, for isoprene ~ 200 ppb, for benzene ~ 0.7 ppb (non-smokers) and ~ 1.8 ppb (smokers) and for toluene ~ 5 ppb.

Ions at twelve mass-to-charge ratios (m/z 28, 42, 67, 71, 79, 81, 83, 93, 97, 109, 123, 137) were selected out of the mass spectrometric profile (m/z 21 - 230) for the exhaled breath of smokers versus non smokers using discriminant analysis (Table II).

Figure 2 shows the derived concentrations of compounds tentatively identified as benzene (m/z 79), acetonitrile (m/z 42), formaldehyde (m/z 31) and methanol (m/z 33) presented as histograms on a ppb log scale separately for the groups of smokers and controls (non-smokers). It can be seen that the distributions are essentially log-normal (as are those for several common breath metabolites studied using SIFT-MS [19-22]) and the concentrations of hydrogen cyanide and acetonitrile are significantly higher in the breath of smokers in comparison with non-smokers.

Threshold concentrations that yielded highest combined sensitivity and specificity were determined using ROC curves to distinguish smokers from non smokers (Table III). The Youden-index (= maximum of sensitivity+specificity-1, YI) as a measure for differentiation between smokers and non-smokers was YI= 0.30 for ions at the m/z values 28 (tentatively identified as HCN), YI= 0.68 for $m/z= 42$ (tentatively identified as acetonitrile) and YI= 0.60 for $m/z= 79$ (tentatively identified as benzene). An example of a ROC curve for the m/z 42 ion is shown in Figure 3. For the ions at m/z 31 (tentatively identified as formaldehyde) and m/z 33 (tentatively identified as methanol) we did *not* observe differences in concentrations between smokers and non-smokers.

The correlation coefficient R for the concentrations of acetonitrile and benzene is $R= 0.71$, for the concentrations of acetonitrile and hydrogen cyanide $R= 0.53$, and for the concentrations of benzene and hydrogen cyanide $R= 0.55$ (see Fig 4). Due to these correlations, the joint analysis of the concentrations of two different compounds does not give rise to a substantial increase in differentiation between smokers and non-smokers.

Our results indicate (see Table IV) that there are no other m/z which show higher concentrations for smokers in comparison with non-smoking healthy volunteers (apart from m/z which are isotopes of the

m/z's mentioned above, and not taking into consideration water clusters and m/z for the compounds released by Tedlar bags).

4. Discussion

The main result of the present study is the identification of distinctive characteristics of smokers' exhaled air (breath) profiles and the delineation of reference concentrations for the volatile biomarkers of smoking using PTR-MS. The screening of human exhaled breath for VOCs characteristic of certain diseases is gaining increasing attention in the recent literature [23, 24]. Yet many pathological conditions that may be diagnosed by breath analysis (e.g. lung cancer) commonly coexist with a variety of morbidities and/or are related to substance abuse, e.g., tobacco smoking, drug, alcohol, etc. Therefore, the results of diagnostic breath testing may be distorted by volatiles having an exogenous origin.

In addition to 12 mass-to-charge ratios with higher concentrations for smokers as compared with the concentrations in non-smoking volunteers we found 7 mass-to-charge ratios (m/z 35, m/z 46, m/z 49, m/z 59, m/z 64, m/z 68, m/z 100), where the smokers show *lower* concentrations than non-smokers (with isotopic effect at m/z 60, see Table IV). We do not discuss these cases here. By introducing Table IV, and by excluding water clusters, primary ions and isotopic effects we try to be more precise than Moser et al. [25] who just stated that “significant differences in exhaled breath composition could be found between smokers and non-smokers in 32 out of 179 masses”.

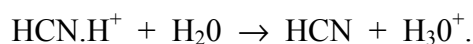
PTR-MS is now an established tool for the rapid determination of exhaled breath profiles of volatile gases either in real-time or using breath samples collected into bags or onto traps [26-32]. The results of a gas chromatography mass spectrometric (GC/MS) study of the profiles of exhaled breath in a healthy population have been reported [33] and several selected ion flow tube mass spectrometer (SIFT-MS) studies of the distributions of the common breath metabolites have been carried out [19-22], including a study of acetonitrile in the exhaled breath and urine headspace of smokers [34]. However, none of the known investigations has provided a comprehensive overlook of smoking-related VOCs in human breath, being mainly focused on the quantification of a single or a few chemicals of smoking origin [6, 34-44]. The present pilot investigation is the attempt to circumscribe the specific characteristics of exhaled air profiles in smokers that can be determined using PTR-MS.

Ionic species at 12 m/z values, selected by discriminant analysis, and hence the corresponding compounds in the breath of smokers, can be tentatively attributed to the substances given in Table II (where the attributions to dimethylsulfoxide, toluene, dimethylfuran and dimethylpyrazole are not more than an “educated guess”). The occurrence of benzene, acetonitrile and 2,5-dimethylfuran in the exhaled breath of smokers is well established [37, 41, 42, 44-49].

This is in concordance with the present results, which also show median concentrations of these compounds in smokers' breath within the same range. The present study as well shows that the concentrations of acetonitrile and benzene are correlated, and that this is also the case for the combination of benzene with hydrogen cyanide and for the combination of acetonitrile with hydrogen cyanide (see Figure 4). Due to this correlation, the combined use of two different marker compounds does not necessarily increase the quality of differentiation between smokers and non-smokers.

Such volatiles as hydrogen cyanide, acetonitrile and benzene (tentatively attributed to the m/z 28, 42 and 79) are well known toxic components of the cigarette smoke [50-53]. Hence, their presence in the exhaled air of smokers is not surprising. We should mention that the concentrations of the compounds indicated by the ions at m/z values 31 and 33 (tentatively identified as formaldehyde and methanol, respectively) are not significantly different in the exhaled breath of this cohort of smokers and non-smokers, despite the fact that formaldehyde and methanol have also been found in the mainstream cigarette smoke [52, 53].

Finally, some limitations of the present study should be discussed. Identification of compounds measured by PTR-MS is always tentative. In particular, overlap of different protonated compounds having the same m/z values may occur. For example, protonated 1,3-butadiene, which is expected to appear at $m/z=55$, has been reported as one of the markers of smoking behavior at the level of $360 \mu\text{g}/\text{m}^3$, corresponding to a few ppb, but this compound cannot be detected using PTR-MS, since the water cluster ion $(\text{H}_2\text{O})_2 \text{H}_3\text{O}^+$ also appears at m/z 55 in the PTR-MS spectrum. Also, the quantification of compounds with proton affinities close to that of water (such as hydrogen cyanide and formaldehyde) gives rise to concentrations which are lower than the actual ones: for these compounds the protonated form (e.g., $\text{HCN}\cdot\text{H}^+$) partly loses its proton to water again [54] in moist samples:



Recently, it has been shown by SIFT-MS experiments ^{**)} that HCN is present in the breath of healthy persons at a median level of about 10 ppb, which is much greater than the median value indicated by the present PTR-MS measurements. Incidentally, HCN along with acetonitrile and benzene is known to be present in inhaled cigarette smoke [55], which is the most likely reason for its higher levels in exhaled breath of smokers as compared to non-smokers.

^{**)} Spanel, P: personal communication

This pilot study has identified twelve volatile organic compounds, VOCs, which are at significantly higher concentrations in the exhaled breath of smokers than non-smokers. Of these compounds, acetonitrile is confirmed as the clearest indicator, as previously shown by other studies [6, 38]. Our results of this compound in exhaled breath of non-smokers is higher than in other studies [6, 34, 38]. This may be an indicator of passive smoking, a subject of great topical interest.

Although our twelve selected VOCs in breath occur following cigarette smoking and decrease with the time after the last smoke, their presence still must be interpreted with caution, since some may also have their origins in adverse clinical conditions such as lung cancer or COPD.

Thus, our findings should be regarded as tentative, and validation studies with the analysis of alveolar air samples, taking into consideration the amount of packyears, respiratory and heart rates and level of blood pressure, including control groups of healthy probands and COPD patients need to be carried out, ideally employing additional analytical techniques such as SIFT-MS and GC/MS, which allow precise (not tentative) identification of the detected compounds.

References

- [1] Manolis, A., *The diagnostic potential of breath analysis*. Clin Chem, 1983. **29**(1): 5-15.
- [2] Davies, S., Spanel, P., and Smith, D., *Quantitative analysis of ammonia on the breath of patients in end-stage renal failure*. Kidney Int, 1997. **52**(1): 223-8.
- [3] Tassopoulos, C.N., Barnett, D., and Fraser, T.R., *Breath-acetone and blood-sugar measurements in diabetes*. Lancet, 1969. **1**(7609): 1282-6.
- [4] Stone, B.G., Besse, T.J., Duane, W.C., Evans, C.D., and DeMaster, E.G., *Effect of regulating cholesterol biosynthesis on breath isoprene excretion in men*. Lipids, 1993. **28**(8): 705-8.
- [5] Sajjad, A., Mottershead, M., Syn, W.K., Jones, R., Smith, S., and Nwokolo, C.U., *Ciprofloxacin suppresses bacterial overgrowth, increases fasting insulin but does not correct low acylated ghrelin concentration in non-alcoholic steatohepatitis*. Aliment Pharmacol Ther, 2005. **22**(4): 291-9.
- [6] Hansel, A., Jordan, A., Holzinger, R., Prazeller, P., Vogel, W., and Lindinger, W., *Proton-Transfer Reaction Mass-Spectrometry - Online Trace Gas-Analysis at the Ppb Level*. International Journal of Mass Spectrometry, 1995. **150**: 609-619.
- [7] Lindinger, W., Hansel, A., and Jordan, A., *On-line monitoring of volatile organic compounds at pptv levels by means of proton-transfer-reaction mass spectrometry (PTR-MS) - Medical applications, food control and environmental research*. International Journal of Mass Spectrometry, 1998. **173**(3): 191-241.
- [8] Smith, D. and Spanel, P., *Selected ion flow tube mass spectrometry (SIFT-MS) for on-line trace gas analysis*. Mass Spectrom Rev, 2005. **24**(5): 661-700.
- [9] Smith, D. and Spanel, P., *Selected Ion Flow Tube Mass Spectrometry, SIFT-MS, for On-line Trace Gas Analysis of Breath*, in *Breath Analysis for Clinical Diagnosis and Therapeutic Monitoring*, Amann, A. and Smith, D., Editors. 2005, World Scientific: Singapore.
- [10] Kleinbaum, D., Kupper, L., Muller, A., and Nizam, K., *Applied Regression Analysis and Other Multivariable Methods*. 1998, Pacific Grove (CA): Brooks/Cole Publishing Company.
- [11] Wassermann, L., *All of Statistics. A Concise Course in Statistical Inference*. 2004, New York: Springer.
- [12] Rao, G., *What is an ROC curve?* J Fam Pract, 2003. **52**(9): 695.
- [13] Faraggi, D. and Reiser, B., *Estimation of the area under the ROC curve*. Stat Med, 2002. **21**(20): 3093-106.
- [14] Walsh, S.J., *Goodness-of-fit issues in ROC curve estimation*. Med Decis Making, 1999. **19**(2): 193-201.
- [15] Dweik, R., *Nitric oxide in exhaled breath: a window on lung physiology and pulmonary disease*, in *Breath Analysis for Clinical Diagnosis and Therapeutic Monitoring*, Amann, A. and Smith, D., Editors. 2005, World Scientific: Singapore.
- [16] Gustafsson, L., *Exhaled nitric oxide: how and why we know it is important*, in *Breath Analysis for Clinical Diagnosis and Therapeutic Monitoring*, Amann, A. and Smith, D., Editors. 2005, World Scientific: Singapore.
- [17] Lundberg, J., *Nasal nitric oxide measurements as a diagnostic tool: ready for clinical use?*, in *Breath Analysis for Clinical Diagnosis and Therapeutic Monitoring*, Amann, A. and Smith, D., Editors. 2005, World Scientific: Singapore.
- [18] Fluss, R., Faraggi, D., and Reiser, B., *Estimation of the Youden Index and its associated cutoff point*. Biom J, 2005. **47**(4): 458-72.

- [19] Turner, C., Spanel, P., and Smith, D., *A longitudinal study of methanol in the exhaled breath of 30 healthy volunteers using selected ion flow tube mass spectrometry, SIFT-MS*. *Physiol Meas*, 2006. **27**(7): 637-48.
- [20] Turner, C., Spanel, P., and Smith, D., *A longitudinal study of ammonia, acetone and propanol in the exhaled breath of 30 subjects using selected ion flow tube mass spectrometry, SIFT-MS*. *Physiol Meas*, 2006. **27**(4): 321-37.
- [21] Turner, C., Spanel, P., and Smith, D., *A longitudinal study of breath isoprene in healthy volunteers using selected ion flow tube mass spectrometry (SIFT-MS)*. *Physiol Meas*, 2006. **27**(1): 13-22.
- [22] Turner, C., Spanel, P., and Smith, D., *A longitudinal study of ethanol and acetaldehyde in the exhaled breath of healthy volunteers using selected-ion flow-tube mass spectrometry*. *Rapid Commun Mass Spectrom*, 2006. **20**(1): 61-8.
- [23] Amann, A. and Smith, D., eds. *Breath Analysis for Clinical Diagnosis and Therapeutic Monitoring*. 2005, World Scientific: Singapore.
- [24] Amann, A., Spanel, P., and Smith, D., *Breath analysis: the approach towards clinical applications*. *Mini reviews in Medicinal Chemistry*, 2007. **7**115 - 129.
- [25] Moser, B., Bodrogi, F., Eibl, G., Lechner, M., Rieder, J., and Lirk, P., *Mass spectrometric profile of exhaled breath--field study by PTR-MS*. *Respir Physiol Neurobiol*, 2005. **145**(2-3): 295-300.
- [26] Wisthaler, A., Tamas, G., Wyon, D.P., Strom-Tejsten, P., Space, D., Beauchamp, J., Hansel, A., Mark, T.D., and Weschler, C.J., *Products of ozone-initiated chemistry in a simulated aircraft environment*. *Environ Sci Technol*, 2005. **39**(13): 4823-32.
- [27] Wisthaler, A., Strom-Tejsten, P., Fang, L., Arnaud, T.J., Hansel, A., Mark, T.D., and Wyon, D.P., *PTR-MS assessment of photocatalytic and sorption-based purification of recirculated cabin air during simulated 7-h flights with high passenger density*. *Environ Sci Technol*, 2007. **41**(1): 229-34.
- [28] D'Anna, B., Wisthaler, A., Andreasen, O., Hansel, A., Hjorth, J., Jensen, N.R., Nielsen, C.J., Stenstrom, Y., and Viidanoja, J., *Atmospheric chemistry of C3-C6 cycloalkanecarbaldehydes*. *J Phys Chem A Mol Spectrosc Kinet Environ Gen Theory*, 2005. **109**(23): 5104-18.
- [29] Amann, A., Telser, S., Hofer, L., Schmid, A., and Hinterhuber, H., *Exhaled breath as a biochemical probe during sleep*, in *Breath Analysis for Clinical Diagnosis and Therapeutic Monitoring*, Amann, A. and Smith, D., Editors. 2005, World Scientific: Singapore. p. 305 - 316.
- [30] Janovsky, U., Scholl-Bürgi, S., Karall, D., Beauchamp, J., Hansel, A., Poupart, G., Schmid, A., and Amann, A., *Breath gas analysis in patients suffering from propionic acidaemia*, in *Breath Analysis for Clinical Diagnosis and Therapeutic Monitoring*, Amann, A. and Smith, D., Editors. 2005, World Scientific: Singapore. p. 401 - 407.
- [31] Ledochowski, M., Amann, A., and Fuchs, D., *Breath Gas Analysis in Patients with Malabsorption Syndromes*, in *Breath Gas Analysis for Medical Diagnostics*, Amann, A. and Smith, D., Editors. 2005, World Scientific: Singapore.
- [32] Wehinger, A., Schmid, A., Mechtcheriakov, S., Ledochowski, M., Grabmer, C., Gastl, G., and Amann, A., *Lung cancer detection by proton transfer reaction mass spectrometric analysis of human breath gas*. *Int J Mass Spec*, 2007. **265**: 49 - 59.
- [33] Phillips, M., Herrera, J., Krishnan, S., Zain, M., Greenberg, J., and Cataneo, R.N., *Variation in volatile organic compounds in the breath of normal humans*. *J Chromatogr B Biomed Sci Appl*, 1999. **729**(1-2): 75-88.
- [34] Abbott, S., Elder, J., Spanel, P., and Smith, D., *Quantification of acetonitrile in exhaled breath and urinary headspace using selected ion flow tube mass spectrometry*. *Int J Mass Spectrom*, 2003. **228**: 655 - 665.

- [35] Cox, B.D. and Whichelow, M.J., *Carbon monoxide levels in the breath of smokers and nonsmokers: effect of domestic heating systems*. J Epidemiol Community Health, 1985. **39**(1): 75-8.
- [36] Euler, D.E., Dave, S.J., and Guo, H., *Effect of cigarette smoking on pentane excretion in alveolar breath*. Clin Chem, 1996. **42**(2): 303-8.
- [37] Gordon, S.M., Wallace, L.A., Brinkman, M.C., Callahan, P.J., and Kenny, D.V., *Volatile organic compounds as breath biomarkers for active and passive smoking*. Environ Health Perspect, 2002. **110**(7): 689-98.
- [38] Jordan, A., Hansel, A., Holzinger, R., and Lindinger, W., *Acetonitrile and Benzene in the Breath of Smokers and Nonsmokers Investigated by Proton-Transfer Reaction Mass-Spectrometry (Ptr-Ms)*. International Journal of Mass Spectrometry and Ion Processes, 1995. **148**(1-2): L1-L3.
- [39] Low, E.C., Ong, M.C., and Tan, M., *Breath carbon monoxide as an indication of smoking habit in the military setting*. Singapore Med J, 2004. **45**(12): 578-82.
- [40] McLaughlin, S.D., Scott, B.K., and Peterson, C.M., *The effect of cigarette smoking on breath and whole blood-associated acetaldehyde*. Alcohol, 1990. **7**(4): 285-7.
- [41] Perbellini, L., Princivale, A., Cerpelloni, M., Pasini, F., and Brugnone, F., *Comparison of breath, blood and urine concentrations in the biomonitoring of environmental exposure to 1,3-butadiene, 2,5-dimethylfuran, and benzene*. Int Arch Occup Environ Health, 2003. **76**(6): 461-6.
- [42] Prazeller P, K.T., Jordan A, Holzinger R, Hansel A, Lindinger W, *Quantification of passive smoking using Proton Transfer Reaction-Mass Spectrometry*. Int J Mass Spectrom, 1998. **179**(L1-L4).
- [43] Senthilmohan, S.T., McEwan, M.J., Wilson, P.F., Milligan, D.B., and Freeman, C.G., *Real time analysis of breath volatiles using SIFT-MS in cigarette smoking*. Redox Rep, 2001. **6**(3): 185-7.
- [44] Wallace, L., Pellizzari, E., Hartwell, T.D., Perritt, R., and Ziegenfus, R., *Exposures to benzene and other volatile compounds from active and passive smoking*. Arch Environ Health, 1987. **42**(5): 272-9.
- [45] Wan-Kuen Jo, K.-W.P., *Utilization of Breath Analysis for Exposure Estimates of Benzene Associated with Active Smoking* Environmental Research 2000. **83**(2): 180-187.
- [46] Pellizzari, E.D., Wallace, L.A., and Gordon, S.M., *Elimination kinetics of volatile organics in humans using breath measurements*. J Expo Anal Environ Epidemiol, 1992. **2**(3): 341-55.
- [47] Brunnemann, K.D., Kagan, M.R., Cox, J.E., and Hoffmann, D., *Determination of benzene, toluene and 1,3-butadiene in cigarette smoke by GC-MDS*. Exp Pathol, 1989. **37**(1-4): 108-13.
- [48] Brugnone, F., Perbellini, L., Maranelli, G., Romeo, L., Alexopoulos, C., and Gobbi, M., *[Effects of cigarette smoking on blood and alveolar air levels of benzene]*. Med Lav, 1990. **81**(2): 101-6.
- [49] Wester, R.C., Maibach, H.I., Gruenke, L.D., and Craig, J.C., *Benzene levels in ambient air and breath of smokers and nonsmokers in urban and pristine environments*. J Toxicol Environ Health, 1986. **18**(4): 567-73.
- [50] Campbell, J.K., Rhoades, J.W., and Gross, A.L., *Acetonitrile as a constituent of cigarette smoke*. Nature, 1963. **198**: 991-2.
- [51] Kensler, C.J. and Battista, S.P., *Components of Cigarette Smoke with Ciliary-Depressant Activity. Their Selective Removal by Filters Containing Activated Charcoal Granules*. N Engl J Med, 1963. **269**: 1161-6.
- [52] Hoffmann, D., Djordjevic, M.V., and Hoffmann, I., *The changing cigarette*. Prev Med, 1997. **26**(4): 427-34.
- [53] Rodgman, A., *Toxic chemicals in cigarette mainstream smoke - Hazard and hoopla*. Contributions to Tobacco Research, 2003. **20**(8): 481 - 545.

- [54] Karl, T., Jobson, T., Kuster, W.C., Williams, E., Stutz, J., Shetter, R., Hall, S.R., Goldan, P., Fehsenfeld, F., and Lindinger, W., *Use of proton-transfer-reaction mass spectrometry to characterize volatile organic compound sources at the La Porte super site during the Texas Air Quality Study 2000*. Journal of Geophysical Research-Atmospheres, 2003. **108**(D16): -.
- [55] Bigourd, D., Cuisset, A., Hindle, F., Matton, S., Fertein, E., Bocquet, R., and Mouret, G., *Detection and quantification of multiple molecular species in mainstream cigarette smoke by continuous-wave terahertz spectroscopy*. Optics Letters, 2006. **31**(15): 2356-2358.

Table I Characteristics of the subject groups

		Smokers		Non-smokers		Ex-smokers		Total	
Male	Age	Mean±SD		Mean±SD		Mean±SD		Mean±SD	
		43.55±11.30		48.58±15.36		62±0		47.37±14.52	
	n	n	%	n	%	n	%	n	%
		31	26.3	86	72.9	1	0.8	118	100
Women	Age	Mean±SD		Mean±SD		Mean±SD		Mean±SD	
		44.79±12.73		50.72±14.57		34.00±19.80		49.26±14.52	
	n	n	%	n	%	n	%	n	%
		19	18.8	80	79.2	2	2.0	101	100
All	Age	Mean±SD		Mean±SD		Mean±SD		Mean±SD	
		44.02±11.75		49.61±14.98		43.33±21.39		48.24±14.52	
	n	n	%	n	%	n	%	N	%
		50	22.8	166	75.8	3	1.4	219	100

Table II

Concentrations of compounds in the breath of smokers versus non-smokers in parts-per-billion, ppb, indicated according to marker ions at the mass-to-charge ratios, m/z , (see text). We used the standard rate constant for protonation $k = 2 \cdot 10^{-9} \text{ cm}^3 \text{ sec}^{-1}$, apart for some tentatively identified compounds where the rate constants are known: methanol (m/z 33, $k = 2.7 \cdot 10^{-9} \text{ cm}^3 \text{ sec}^{-1}$), acetonitrile (m/z 42, $k = 5.1 \cdot 10^{-9} \text{ cm}^3 \text{ sec}^{-1}$). The concentrations of the m/z values 31 and 33 (tentatively formaldehyde and methanol) are *not* significantly different in the exhaled breath of this cohort of smokers and non-smokers. Classification as smoker/non-smoker/ex-smoker was based on self-declaration of volunteers. GSD: geometric standard deviation. The concentrations of hydrogen cyanide and formaldehyde are underestimated by PTR-MS measurements (due to their low proton affinity).

Table II (upper panel): using filtered data (*ANOVA, ** Kruskal-Wallis)

m/z	Tentative identification of VOCs	Smokers				Non-smokers				p-Value: significant difference in exhaled values
		inhaled concentration* ppb		exhaled concentration* ppb		Inhaled concentration* ppb		exhaled concentration* ppb		
		Median	GSD	Median	GSD	Median	GSD	Median	GSD	
m/z 28	hydrogen cyanide	0.85	1.60	1.00	1.47	0.87	1.60	0.84	1.26	< 0.001 *
m/z 31	formaldehyde	11.80	1.78	3.78	1.56	10.58	1.85	3.68	1.42	n.s. *
m/z 33	methanol	68.98	2.20	173.04	1.77	95.80	2.53	199.68	1.72	n.s. *
m/z 42	acetonitrile	13.65	2.73	42.20	2.99	14.17	3.26	5.10	2.10	< 1.1 x10 ⁻¹² **
m/z 67		0.68	1.40	3.46	1.72	0.71	1.44	2.38	1.63	< 7.5 x10 ⁻⁰⁷ *
m/z 71		3.35	1.27	5.78	1.51	3.65	1.41	4.36	1.43	<1.5 x10 ⁻⁰⁵ *
m/z 79	benzene	0.83	1.75	1.82	2.22	0.97	1.84	0.66	1.46	< 2.5x10 ⁻⁰⁹ *
	dimethylsulfoxide									
m/z 81		3.94	1.56	8.10	2.71	4.16	1.74	4.95	2.24	<7.5x10 ⁻⁰⁵ **
m/z 83		4.98	1.54	7.43	3.11	6.00	1.60	5.35	2.21	<8.1x10 ⁻⁰⁵ **
m/z 93	toluene	3.69	1.67	5.94	1.55	4.51	1.73	4.43	1.52	< 0.001 **
m/z 97	dimethylfuran, dimethylpyrazole	1.68	1.59	2.72	2.16	1.84	1.72	1.20	1.67	< 1.5x10 ⁻¹² **
m/z 109		2.33	1.36	2.89	1.50	2.56	1.35	2.70	1.38	<0.022 **
m/z 123		0.45	1.32	0.62	1.71	0.49	1.45	0.42	1.27	<1.1x10 ⁻¹² **
m/z 137		1.15	1.49	2.11	2.78	1.27	1.64	1.54	2.26	<0.0015

Table II (lower panel): without filtering the data (*ANOVA, ** Kruskal-Wallis)

m/z	Tentative identification of VOCs	Smokers				Non-smokers				p-Value: significant difference in exhaled values
		inhaled concentration*		exhaled concentration*		Inhaled concentration*		exhaled concentration*		
		ppb		ppb		ppb		ppb		
		Median	GSD	Median	GSD	Median	GSD	Median	GSD	
m/z 28	hydrogen cyanide	0.85	1.50	1.10	1.47	0.87	1.60	0.86	1.31	< 1.4 x10 ⁻⁰⁵ **
m/z 31	formaldehyde	11.80	1.78	3.97	1.70	10.58	1.85	3.75	1.51	n.s. *
m/z 33	methanol	68.98	2.20	173.04	1.77	96.13	2.54	199.68	1.72	n.s. *
m/z 42	acetonitrile	13.65	2.73	39.46	2.55	14.17	3.26	5.18	2.17	< 1.0 x10 ⁻¹⁵ **
m/z 67		0.68	1.40	3.46	1.72	0.71	1.45	2.38	1.63	< 7.5 x10 ⁻⁰⁷ *
m/z 71		3.35	1.27	5.97	1.50	3.65	1.41	4.41	1.46	<5.4 x10 ⁻⁰⁵ *
m/z 79	benzene	0.83	1.75	1.79	2.04	0.97	1.84	0.70	1.53	< 1.0 x10 ⁻¹⁵ *
	dimethylsulfoxide									
m/z 81		3.94	1.56	8.10	2.71	4.18	1.76	4.95	2.24	<7.5x10 ⁻⁰⁵ **
m/z 83		4.98	1.54	7.43	3.11	6.00	1.60	5.35	2.21	<8.1x10 ⁻⁰⁵ **
m/z 93	toluene	3.69	1.67	6.47	1.57	4.51	1.73	4.48	1.55	< 0.0003 **
m/z 97	dimethylfuran, dimethylpyrazole	1.68	1.59	2.86	2.11	1.84	1.72	1.24	1.75	< 5.3x10 ⁻¹² **
m/z 109		2.33	1.36	3.32	1.47	2.58	1.35	2.79	1.44	<0.011 *
m/z 123		0.45	1.32	0.72	1.60	0.49	1.45	0.43	1.32	<5.3x10 ⁻¹⁴ **
m/z 137		1.15	1.49	2.11	2.78	1.27	1.64	1.54	2.26	<0.0015

Table III Classification value, Sensitivity, specificity for maximal Youden-Index [18] for the discriminating ions at the m/z values relating to breath compounds of smoking origin *for filtered data*. Classification as smoker/non-smoker/ex-smoker was based on self-declaration of volunteers. For all possible values, the value to classify between the groups is taken for which the Youden-Index is at its maximum.

m/z	Classification value [ppb]	Sensitivity %	Specificity %	max. Youden-Index
28	1.03	84.2	46.2	0.30
42	13.69	87.0	80.6	0.68
67	2.42	51.5	78.0	0.30
71	5.93	83.1	46.9	0.30
79	1.13	92.5	67.6	0.60
81	13.93	90.3	24.0	0.14
83	6.91	67.3	56.0	0.23
93	5.58	77.8	56.5	0.34
97	1.33	62.0	89.4	0.51
109	3.77	95.4	29.5	0.25
123	0.50	83.2	76.2	0.59
137	6.44	94.5	18.0	0.13

Table IV: All mass-to-charge ratios between 28 and 230 were checked for differences in concentration between smokers and non-smokers. Table IV gives additional information on the reasons why certain mass-to-charge ratios were not considered as showing different concentrations for a particular volatile compound between smokers and non-smokers. Typically, mass-to-charge ratios for water clusters and N,N-dimethyl-acetamide and phenol (released from Tedlar bags and arising at $m/z=88$ and $m/z=95$) were not considered. Also mass-to-charge ratios which are expected to be only isotope effects (e.g. $m/z=70$ can be expected to be an isotope effect from isoprene $m/z=69$) were not considered. Finally, only 12 mass-to-charge ratios show an effect of smoking on the respective concentration. This might contrast with the result of Moser et al. [25] that “Significant differences in exhaled breath composition could be found between smokers and non-smokers in 32 out of 179 masses”. Colorcode: *green*= significantly higher in smokers as compared to non-smokers; *magenta*: significantly lower in smokers than in non-smokers; *yellow*= could be considered, but $p>0.01$ is possible resp. very low concentrations.

m/z	after filtering	Non-smoker	smoker	before filtering	Non-smoker	smoker
28	significant	0.84	1.00	significant	0.86	1.10
29	not significant;	27.17	27.30	not significant;	27.26	27.55
30	NO ⁺ from the ion source; all samples filtered	581.83	572.25	NO ⁺ from the ion source	581.83	572.25
31	not significant	3.68	3.78	not significant	3.75	3.97
32	O ₂ ⁺ from the ion source	929.60	947.23	O ₂ ⁺ from the ion source	929.60	947.23
33	not significant; concentration smoker < concentration non-smoker	199.68	173.04	not significant; concentration smoker < concentration non-smoker	199.68	173.04
34	not significant; concentration smoker < concentration non-smoker	34.52	33.90	not significant; concentration smoker < concentration non-smoker	34.52	33.90
35	significant; (0.01 < p < 0.05) concentration smoker < concentration non-smoker	1.02	0.85	significant; (0.01 < p < 0.05) concentration smoker < concentration non-smoker	1.02	0.85
36	not significant; difference in inhaled air; concentration	2.80	2.52	not significant; difference in inhaled air; concentration	2.80	2.52

	smoker < concentration non-smoker			smoker < concentration non-smoker		
37	water cluster	2816.12	2874.43	water cluster	2816.12	2874.43
38	isotope of water cluster	4.30	3.87	isotope of water cluster	4.30	3.87
39	not significant	67.26	87.52	not significant	71.79	88.32
40	not significant	1.63	2.41	not significant	1.71	2.42
41	not significant; concentration smoker < concentration non-smoker	285.14	379.80	not significant; concentration smoker < concentration non-smoker	316.70	379.80
42	significant	5.10	42.20	significant	5.18	39.46
43	not significant	257.69	262.18	not significant	258.53	262.18
44	not significant	14.11	14.15	not significant	14.17	14.15
45	protonated carbon dioxide (CO ₂ H ⁺) disturbs measurements on m/z 45	64.14	69.67	protonated carbon dioxide (CO ₂ H ⁺) disturbs measurements on m/z 45	67.16	74.44
46	significant (0.01 < p < 0.05); concentration smoker < concentration non-smoker	44.95	30.50	significant (0.01 < p < 0.05); concentration smoker < concentration non-smoker	44.95	30.50
47	not significant	27.98	50.39	not significant	29.54	53.32
48	not significant	0.86	1.28	not significant;	0.91	1.37
49	significant concentration smoker < concentration non-smoker	1.45	0.90	significant; concentration smoker < concentration non-smoker	1.45	0.90
50	not significant; concentration < 1 ppb	0.26	0.25	not significant; concentration < 1 ppb	0.26	0.25
51	not significant; concentration smoker < concentration non-smoker	1.52	1.29	not significant; concentration smoker < concentration non-smoker	1.52	1.29
52	isotope effect of m/z 51; concentration	0.08	0.08	isotope effect of m/z 51; concentration	0.08	0.08

	<<1 ppb			<<1 ppb		
53	significant; concentration < 1 ppb	0.38	0.52	significant; concentration < 1 ppb	0.38	0.55
54	significant; concentration < 1 ppb	0.21	0.26	significant; concentration < 1 ppb	0.22	0.30
55	water cluster	10.33	12.16	water cluster	10.33	12.16
56	isotope of m/z 55	0.59	0.86	isotope of m/z 55	0.59	0.87
57	Tedlar-bag related m/z	17.41	18.95	Tedlar-bag related m/z	17.56	19.31
58	not significant; concentration smoker < concentration non-smoker;	1.48	1.35	not significant; concentration smoker < concentration non-smoker;	1.48	1.36
59	significant (0.01 < p < 0.05); concentration smoker < concentration non-smoker	523.16	435.71	significant (0.01 < p < 0.05); concentration smoker < concentration non-smoker	523.16	435.71
60	significant (0.01 < p < 0.05); concentration smoker < concentration non-smoker	38.60	31.91	significant (0.01 < p < 0.05); concentration smoker < concentration non-smoker	38.60	31.91
61	not significant; concentration smoker < concentration non-smoker;	113.49	105.67	not significant; concentration smoker < concentration non-smoker;	113.49	105.67
62	not significant; concentration smoker < concentration non-smoker	3.45	3.25	not significant; concentration smoker < concentration non-smoker	3.45	3.25
63	not significant; concentration smoker < concentration non-smoker	20.86	19.65	not significant; concentration smoker < concentration non-smoker	20.86	19.65
64	significant; (0.01 < p < 0.05);	0.42	0.35	significant; (0.01 < p < 0.05);	0.42	0.35

	concentration smoker < concentration non-smoker; concentration < 1 ppb;			concentration smoker < concentration non-smoker; concentration < 1 ppb;		
65	not significant; concentration smoker < concentration non-smoker; concentration < 1 ppb	0.59	0.52	not significant; concentration smoker < concentration non-smoker; concentration < 1 ppb	0.59	0.52
66	significant; concentration << 1 ppb;	0.05	0.08	significant; concentration << 1 ppb;	0.05	0.08
67	significant	2.38	3.46	significant	2.38	3.46
68	significant; (0.01 < p < 0.05); concentration smoker < concentration non-smoker;	1.26	1.42	significant; (0.01 < p < 0.05); concentration smoker < concentration non-smoker;	1.26	1.42
69	significant; (0.01 < p < 0.05)	201.05	235.57	significant; (0.01 < p < 0.05)	201.05	235.57
70	isotope of m/z 69	2.91	3.29	isotope of m/z 69	2.91	3.29
71	significant	4.36	5.78	significant	4.41	5.97
72	not significant; concentration smoker < concentration non-smoker	1.95	1.83	not significant; concentration smoker < concentration non-smoker	1.95	1.83
73	effected by water cluster	7.86	10.35	effected by water cluster	7.86	10.35
74	not significant; concentration smoker < concentration non-smoker	16.45	15.75	not significant; concentration smoker < concentration non-smoker	16.45	15.75
75	not significant	13.38	13.58	not significant	13.38	13.58
76	not significant; concentration smoker < concentration non-smoker;	1.03	0.99	not significant; concentration smoker < concentration non-smoker;	1.03	0.99

77	not significant; concentration smoker < concentration non-smoker; concentration < 1 ppb	0.66	0.60	not significant; concentration smoker < concentration non-smoker; concentration < 1 ppb	0.66	0.60
78	not significant; concentration < 1 ppb	0.16	0.18	<i>significant;</i> <i>(0.01 < p < 0.05)</i> concentration < 1 ppb	0.17	0.22
79	<i>significant</i>	0.66	1.82	<i>significant</i>	0.70	1.79
80	isotope of m/z 79 concentration < 1 ppb	0.29	0.74	isotope of m/z 79 concentration < 1 ppb	0.29	0.77
81	<i>significant</i>	4.95	8.10	<i>significant</i>	4.95	8.10
82	<i>significant;</i> <i>concentration <</i> <i>1 ppb;</i>	0.34	0.79	<i>significant;</i> <i>concentration <</i> <i>1 ppb;</i>	0.34	0.79
83	<i>significant</i>	5.35	7.43	<i>significant</i>	5.35	7.43
84	isotope of m/z 83, concentration <1ppb	0.47	0.77	isotope of m/z 83, concentration <1ppb	0.47	0.77
85	<i>significant;</i> <i>(0.01 < p < 0.05)</i>	1.58	1.73	<i>significant;</i> <i>(0.01 < p < 0.05)</i>	1.64	1.83
86	not significant; concentration <1 ppb	0.45	0.47	not significant; concentration <1 ppb;	0.45	0.47
87	not significant	2.40	2.49	not significant	2.40	2.49
88	not significant; concentration smoker < concentration non-smoker	13.34	10.39	not significant; concentration smoker < concentration non-smoker	13.34	10.39
89	possibly influenced by isotope of m/z 88 (N,N- dimethyl acetamide released by Tedlar bags)	5.84	7.08	possibly influenced by isotope of m/z 88 (N,N- dimethyl acetamide released by Tedlar bags)	5.84	7.38
90	not significant; concentration	2.11	1.99	not significant; concentration	2.11	1.99

	smoker < concentration non-smoker			smoker < concentration non-smoker		
91	not significant; concentration smoker < concentration non-smoker	7.83	8.27	not significant; concentration smoker < concentration non-smoker	7.83	8.27
92	not significant; concentration < 1 ppb	0.64	0.67	not significant; concentration < 1 ppb	0.65	0.73
93	significant	4.43	5.94	significant	4.48	6.47
94	not significant	2.55	2.51	not significant	2.55	2.51
95	not significant	18.60	19.16	not significant	18.60	19.16
96	not significant	1.43	1.79	not significant	1.43	1.79
97	significant	1.20	2.72	significant	1.24	2.86
98	isotope of m/z 97	0.22	0.33	isotope of m/z 97	0.22	0.35
99	not significant; concentration smoker < concentration non-smoker	2.88	2.75	not significant; concentration smoker < concentration non-smoker	2.96	2.81
100	significant; (0.01 < p < 0.05); concentration < 1 ppb; concentration smoker < concentration non-smoker	0.86	0.72	significant; (0.01 < p < 0.05); concentration < 1 ppb; concentration smoker < concentration non-smoker	0.86	0.72
101	not significant; concentration < 1 ppb	0.85	0.86	not significant; concentration < 1 ppb	0.85	0.86
102	not significant; concentration < 1 ppb;	1.10	1.08	not significant; concentration < 1 ppb;	1.10	1.08
103	not significant	1.64	1.81	not significant	1.64	1.81
104	not significant; concentration < 1 ppb	0.41	0.39	not significant; concentration < 1 ppb	0.41	0.39
105	not significant; concentration < 1 ppb;	0.62	0.62	not significant; concentration < 1 ppb;	0.62	0.62
106	not significant	1.28	1.45	not significant	1.28	1.45
107	not significant	3.70	3.27	not significant	3.70	3.27

	concentration smoker < concentration non-smoker			concentration smoker < concentration non-smoker		
108	not significant	5.72	6.03	not significant	5.72	6.03
109	significant	2.70	2.89	significant; (0.01 < p < 0.05)	2.79	3.32
110	significant; difference in inhaled air (non-smoker > smoker); concentration < 1 ppb	0.27	0.30	significant; difference in inhaled air (non-smoker > smoker); concentration < 1 ppb	0.28	0.33
111	significant; concentration < 1 ppb	0.69	0.86	significant; concentration < 1 ppb	0.70	0.90
112	not significant; concentration < 1 ppb difference in inhaled air (non-smoker > smoker);	0.22	0.22	not significant; concentration < 1 ppb difference in inhaled air (non-smoker > smoker);	0.23	0.23
113	not significant; concentration < 1 ppb;	0.52	0.54	not significant; concentration < 1 ppb;	0.53	0.56
114	not significant; concentration < 1 ppb; concentration smoker < concentration non-smoker	0.23	0.20	not significant; concentration < 1 ppb; concentration smoker < concentration non-smoker	0.23	0.20
115	not significant; concentration < 1 ppb; concentration smoker < concentration non-smoker	0.51	0.48	not significant; concentration < 1 ppb; concentration smoker < concentration non-smoker	0.51	0.48
116	not significant; concentration < 1 ppb concentration smoker <	0.39	0.37	not significant; concentration < 1 ppb concentration smoker <	0.39	0.37

	concentration non-smoker			concentration non-smoker		
117	not significant	1.84	1.94	not significant	1.84	1.94
118	not significant; concentration smoker < concentration non-smoker; concentration <1 ppb	0.46	0.39	not significant; concentration smoker < concentration non-smoker; concentration <1 ppb	0.46	0.39
119	not significant	1.56	1.62	not significant	1.56	1.62
120	not significant	1.12	1.20	not significant	1.12	1.20
121	not significant; difference in inhaled air (non-smoker >smoker);	1.20	1.24	not significant; difference in inhaled air (non-smoker >smoker);	1.20	1.24
122	not significant; difference in inhaled air (non-smoker >smoker); concentration <1 ppb	0.25	0.27	not significant; difference in inhaled air (non-smoker >smoker); concentration <1 ppb	0.25	0.27
123	significant (but relatively low concentrations < 1ppb)	0.42	0.62	significant (but relatively low concentrations < 1ppb)	0.43	0.72
124	not significant; concentration <1 ppb	0.85	0.92	not significant; concentration <1 ppb	0.85	0.92
125	not significant	1.23	1.23	not significant	1.23	1.23
126	not significant; concentration smoker < concentration non-smoker	2.70	2.59	not significant; concentration smoker < concentration non-smoker	2.70	2.59
127	not significant; concentration <1 ppb	0.67	0.68	not significant; concentration <1 ppb	0.67	0.68
128	not significant; concentration smoker < concentration non-smoker concentration	0.49	0.44	not significant; concentration smoker < concentration non-smoker; concentration	0.49	0.44

	<1 ppb			<1 ppb;		
129	not significant; concentration <1 ppb	0.37	0.37	not significant; concentration <1 ppb	0.37	0.37
130	not significant; concentration <1 ppb	0.11	0.10	not significant; concentration <1 ppb	0.11	0.10
131	not significant; concentration <1 ppb	0.31	0.33	not significant; concentration <1 ppb	0.31	0.33
132	not significant; concentration <1 ppb	0.11	0.09	not significant; concentration <1 ppb	0.11	0.09
133	not significant; concentration smoker < concentration non-smoker	1.08	0.91	not significant; concentration smoker < concentration non-smoker	1.08	0.91
134	not significant; concentration <1 ppb;	0.16	0.16	not significant; concentration <1 ppb;	0.16	0.16
135	not significant; concentration <1 ppb; concentration smoker < concentration non-smoker	0.68	0.67	not significant; concentration <1 ppb; concentration smoker < concentration non-smoker	0.68	0.67
136	not significant	1.93	1.89	not significant	1.93	1.89
137	significant	1.54	2.11	significant	1.54	2.11
138	significant concentration <1 ppb	0.30	0.37	significant concentration <1 ppb	0.30	0.37
139	not significant; difference in inhaled air (non-smoker >smoker); concentration <1 ppb	0.35	0.39	not significant; difference in inhaled air (non-smoker >smoker); concentration <1 ppb	0.35	0.39
140	significant; difference in inhaled air (non-smoker >smoker); concentration	0.09	0.09	significant; difference in inhaled air (non-smoker >smoker); concentration	0.09	0.09

	<<1 ppb			<<1 ppb		
141 - 230	not significant; concentration <1 ppb;			not significant; concentration <1 ppb		

Figure 1 The example of the logistic regression curve and 30% marginal concentration (*not filtered*) for acetonitrile as calculated from the ion count rate at $m/z = 42$ using $k=5.1 \times 10^{-9} \text{ cm}^3 \text{ s}^{-1}$.

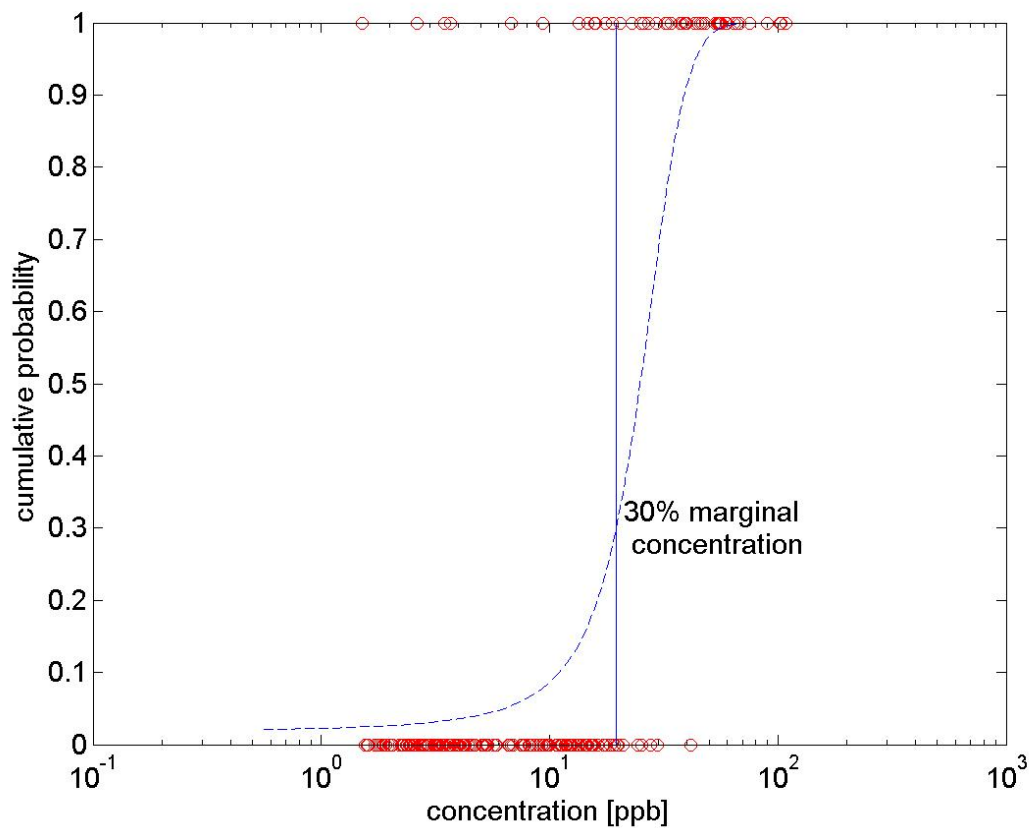
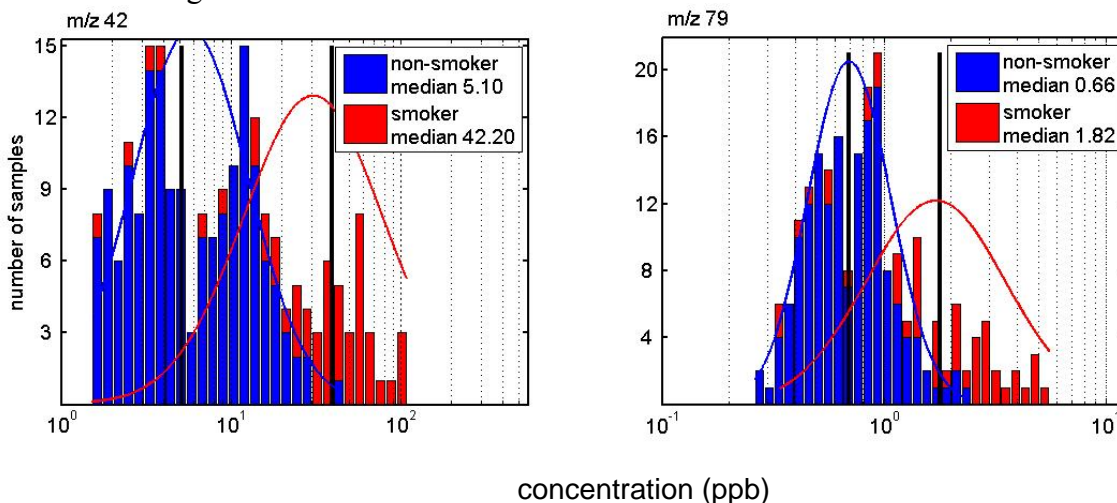


Figure 2

a) Examples

Histograms of *filtered* concentrations of acetonitrile (m/z 42) and benzene (m/z 79) present in the breath gas of smokers and non smokers (significantly higher concentrations ($p < .01$)). The curves show the estimated log-normal distribution and the vertical black lines show the median-values.



b) Counter-examples

Histograms of concentrations (*without filtering*) of formaldehyde (m/z 31) and methanol (m/z 33) present in the breath gas of smokers and non smokers. These do *not* show a significant difference in concentrations. The curves show the estimated normal distribution and the vertical black lines show the median-values. Filtering would not make sense for m/z 31 and m/z 33, since there are no differences in concentration between smokers and non-smokers, and therefore all very low concentrations would be eliminated (see Methods Section).

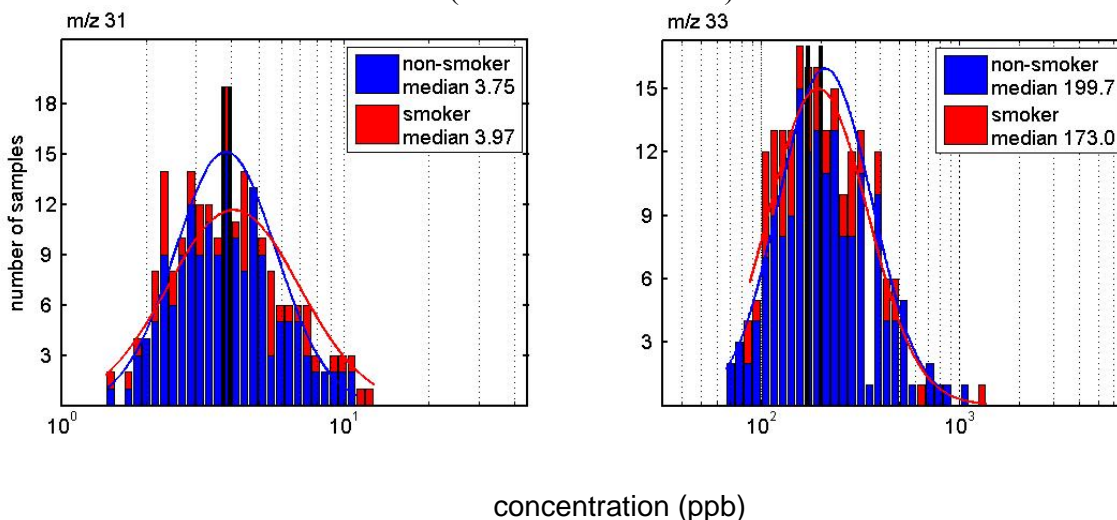


Figure 3 Receiver operating characteristic (ROC) curve for *filtered* concentrations of m/z 42. This plot demonstrates the ROC curve of prediction of the breath test for the m/z 42 in a view of the continuum of sensitivity and specificity (the black point marks point of maximum Youden-index, the red one of maximal accuracy).

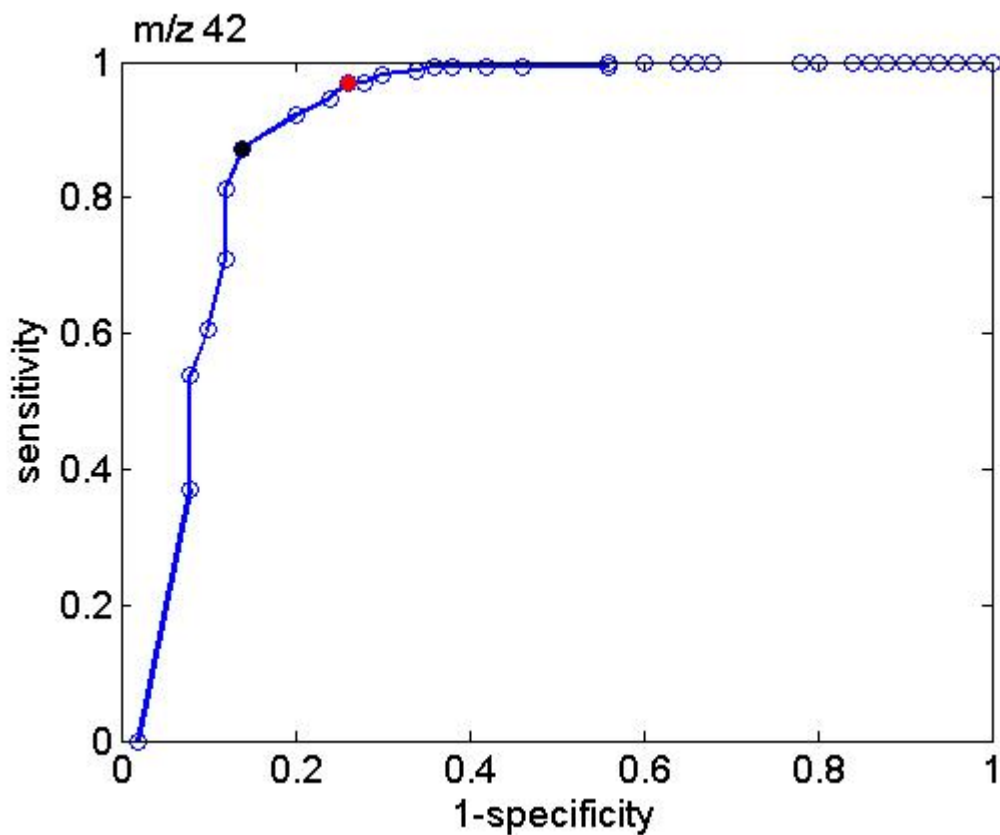
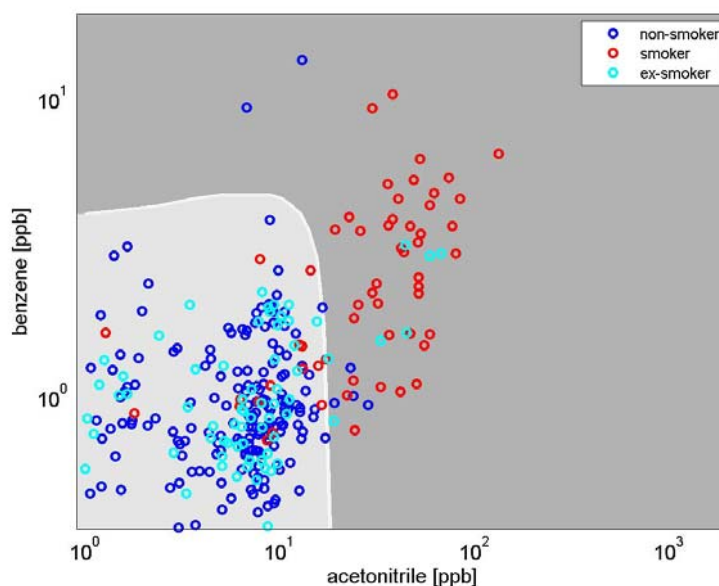


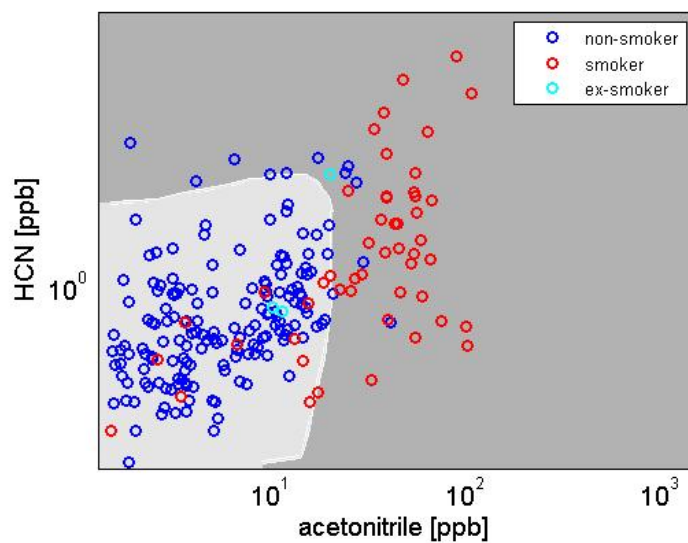
Figure 4 *Filtered* concentrations (parts-per-billion, ppb) of compounds in breath derived from product ions at m/z 42 and m/z 79 (upper panel) identified as acetonitrile and benzene respectively and those derived from product ions at m/z 28 and m/z 42, identified as hydrogen cyanide and acetonitrile (middle panel), and hydrogen cyanided and benzene (lower panel) in smokers, non-smokers and ex-smokers. Between all the three pairs there is a significant correlation. In these pictures we can see a classification based on two substances, the area where smokers are classified is dark grey, the one for non-smoker light grey. The classification was computed with a quadratic discriminant analysis (MATLAB® command classify.m with quadratic boundaries between groups) based on *filtered* data. The sensitivity, specificity and the Youden-Index are shown in the tables besides the plots.

(Fig 4 upper panel)



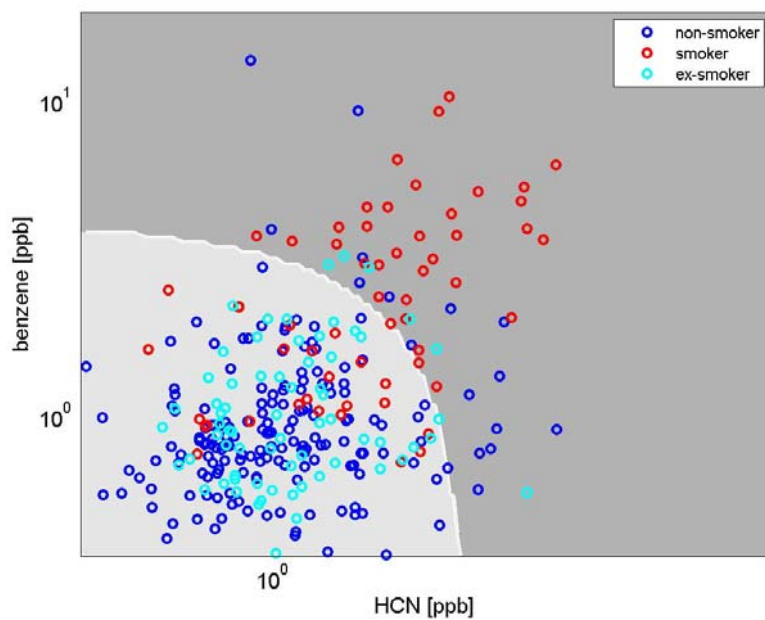
Sensitivity	Specificity	Youden-Index	Correlation coefficient R
%	%		
91.7	78.4	0.70	0.71

(Fig 4 middle panel)



Sensitivity %	Specificity %	Youden- Index	Correlation coefficient R
64.9	78.9	0.44	0.53

(Fig 4 lower panel)



Sensitivity %	Specificity %	Youden-Index	Correlation coefficient R
92.1	68.0	0.60	0.55