Methodological considerations for large-scale breath analysis studies: lessons from the U-BIOPRED severe asthma project.

Waqar M. Ahmed, Paul Brinkman, Hans Weda, Hugo H. Knobel, Yun Xu, Tamara M. Nijser, Royston Goodacre, Nicholas Rattray, Teunis J. Vink, Marco Santonico, Giorgio Pennazza, Paolo Montuschi, Peter J. Sterk, Stephen J. Fowler, on behalf of the U-BIOPRED study group

Author affiliations:
1. School of Biological Sciences, University of Manchester, United Kingdom.
2. Department of Respiratory Medicine, Academic Medical Center (AMC), University of Amsterdam, The Netherlands.
5. School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, United Kingdom.
6. Department of Environmental Health Sciences, Yale School of Public Health, Yale University, New Haven, United States.
7. Unit of Electronics for Sensor Systems, Faculty of Engineering, Campus Bio-Medico University of Rome, Rome, Italy.
8. Department of Pharmacology, Faculty of Medicine, Catholic University of the Sacred Heart, Rome, Italy.
9. Manchester Academic Health Sciences Centre, Manchester University NHS Foundation Trust, United Kingdom.

*Corresponding author
Email: stephen.fowler@manchester.ac.uk

#Membership of the U-BIOPRED study group is listed in the supplementary information
Abstract

Methods for breath sampling and analysis require robust quality assessment to minimise the risk of false discoveries. Planning large-scale multi-site breath metabolite profiling studies also requires careful consideration of systematic and random variation as a result of sampling and analysis techniques. In this study we use breath sample data from the recent U-BIOPRED cohort to evaluate and discuss some important methodological considerations such as batch variation and correction, variation between sites, storage and transportation, as well as inter-instrument analytical differences. Based on this we provide a summary of recommended best practices for new large scale multi-site studies.

Keywords

Multi-site studies, breath analysis, metabolomics, quality assurance, quality control, U-BIOPRED

1. Introduction

The deployment of ‘omics technologies in clinical research is proving invaluable in unravelling complex biological mechanisms and for the discovery of important markers for inflammatory lung disease[1]. However, for studies to produce clinically meaningful results, a large number of samples are typically required to provide statistical power and hence reduce the risk of false discoveries[2]. Robust sampling and analysis methods are equally important, especially where additional sites and instruments will be prone to increased systematic variation.

In metabolomic studies, a standardised and well-controlled sampling and analysis approach must be adopted throughout the whole experimental pipeline [3–6]. The use of metabolomic workflows in breath research is challenging as breath samples typically contain high intra- and inter-sample variation, a number of sampling variables (such as diet and medical history), and the presence of
exogenous artefact or contaminant compounds. As a result, and because of limited long term storage options, it is usually not possible to form pooled quality control samples of breath to evaluate analytical variation[7]. In addition, breath sample analysis can result in a high number of features, often exceeding the number of samples. This dominant effect, which may be specific to a single site or instrument, can influence downstream data analysis[8]. It is therefore important to consider additional quality assessment procedures in breath research where large samples sizes are used. Studies with multiple sites can introduce new variation from unavoidable inconsistent sampling techniques, sample storage and transportation, and lack of control samples. Methodological considerations have previously been reported for the collection and analysis of breath samples, and international task forces have recommended breath sampling and analysis standardisation[9–13]. However, specific guidance for large scale multi-site studies remains limited.

The aim of this investigation is to address methodological considerations for large scale multi-site studies that may otherwise be overlooked throughout collection and analysis of breath samples. To show examples of methodological considerations, we will use data from the U-BIOPRED (Unbiased BIOmarkers in PREDiction of respiratory disease outcomes) severe asthma cohort study, and consider the following objectives:

- To identify previously-reported disease-specific and artefact VOCs;
- To use the extracted VOCs to assess variation between sites and patients;
- To assess the effect of variation from storing breath samples in sorbent-packed tubes;
- To evaluate sampling and analytical inter-batch variation within multivariate data and subsequently compare batch correction methods;
- Compare multivariate datasets from different instruments using Procrustes analysis.
2. Materials and Methods

2.1. Study information

The U-BIOPRED study is a large scale European multi-site study with the objective to reveal novel phenotypes and therapeutic targets in severe asthma[14,15]. Asthma subtype was categorised as either mild/moderate or severe, as defined by the Innovative Medicines Initiative (IMI)[16]. Several ‘omics technologies were employed in a systems biology approach using samples including blood, urine, sputum supernatants, and breath from adult and paediatric (school-aged and pre-school children) patient cohorts. Extensive clinical data and patient-reported outcomes were also collected. In this work we will focus on data from the breath volatilomics part of the study.

2.2. Breath sampling and analysis

Samples were collected and analysed between 2012 and 2015. After 5 min breathing room air through a VOC filter (A2, North Safety, Middelburg, Netherlands), participants were asked to breathe a single vital capacity into a ten litre Tedlar® bag (SKC Inc, Eighty Four, PA, United States) via a three way valve[17]. Within 10 min, the mixed expiratory breath sample was sequentially purged onto two sorbent tubes containing Tenax GR (stainless steel, 6 mm x 177.8 mm, Gerstel, Mülheim an der Ruhr, Germany) by using a custom-built peristaltic pump (medical grade tubing, bore 4.8 mm, wall thickness 2.4 mm, Williamson Pumps, Brighton, UK) at a flow rate of 250 mL/min. A diagram of this layout is shown in the supplementary information (Figure S1). After local sampling, the tubes were tightened with brass caps, sealed in an air-tight packaging, and transported by air or land to the Academic Medical Centre, Amsterdam, for central analysis and distribution. Sorbent tubes were stored in refrigerators (2 to 8 °C) prior to analysis.

The first sample was used for eNose analysis. VOCs were released from the sorbent tube using a thermal desorption oven (TDS 3, Gerstel, Mülheim an der Ruhr, Germany), after which the sample was transferred into a Tedlar bag with nitrogen as carrier gas. Subsequent analysis was carried out by a composite eNose platform. The sample handing method has been described previously[18].
eNose platform consisted of four eNoses from four different brands, using distinct sensor technologies: 1) Cyranose C320 using carbon black-polymer sensors[19] (two instruments were included in this study), 2) Tor Vergata eNose using quartz crystal microbalances (QMB) covered with metalloporphyrins[20], 3) Common Invent eNose using metal oxide semiconductor sensors[21], and 4) Owlstone Lonestar based on field asymmetric ion mobility spectrometry[22]. A combined total of 158 sensor response signals were used in the analysis (Common Invent = 8, Cyranose = 32, Lonestar = 110, Tor Vegata = 8). The Lonestar was setup to scan between 0 and 100 % dispersion field (magnitude of electric field) in 51 steps and a compensation voltage between -6 and +6 Volt in 512 steps.

The second sorbent tube was sent to Philips Research in Eindhoven, The Netherlands, in order to be analysed centrally by thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS), using a similar method as previously described[23]. In brief, samples underwent automated TD (Gerstel, Mülheim an der Ruhr, Germany) using helium as the carrier gas. The sample was then purged onto a packed liner, heated to 300 °C for 3 min, and subsequently transferred to a Tenax TA cold trap (kept at -150 °C) for 2 min. The cold trap was then heated to 280°C at 20°C/s, after which the sample was injected into the GC column (VF1-ms, 30 m × 0.25 mm, 1 µm, 100% dimethylpolysiloxane, Varian Chrompack, Middelburg, The Netherlands) using a splitless injection method with helium as a carrier gas at a flow rate of 1.2 mL/min. The GC oven (7890 N GC, Agilent, Santa Clara, CA, USA) was set to a ramped temperature programme with the following parameters: 40 °C (5 min hold), ramp to 300 °C at 10 °C/min (5 min hold). Compounds were then transferred to a MS with a Time-of-Flight mass analyser (LECO Pegasus 4D, LECO, Sint Joseph, MI, USA) in electron ionization positive mode (70eV), with a mass spectral acquisition range between 29 to 450 Da.

2.3. Data pre-processing and treatment

Raw GC-MS files were converted to netCDF files, and uploaded to an open source knowledge management platform (tranSMART) for quality checks and to ensure data transparency between...
project partners[24]. Both pre-processing and statistical analyses were performed in the R software environment (R Development Core Team 2017, version 3.4.2) using relevant R packages (XCMS, BatchCorrMetabolomics, and Vegan) described in detail previously [25–27]. GC-MS samples were pre-processed to create an ion fragment data matrix for further analysis and extraction of known VOCs.

Several putative asthma-related VOCs (and other commonly reported VOCs such as C₅ to C₁₅ alkanes, acetone, isoprene, toluene, and internal standards) [28–33] were targeted for further assessment, based on predicted retention indices and mass fragments (extracted from Pubchem and NIST online chemical databases). Retention indices were predicted using known retention times of alkanes[34]. These alkanes were assigned a retention index number using their carbon number (e.g. 800 for C₈, or 900 for C₉) and subsequently used as anchors for cubic spline interpolation of unknown VOCs, for example, if a VOC had a retention time between octane and nonane, its retention index would fall between 800 and 900. In addition, a retention index threshold of ± 20 was applied.

The identity of compounds that met these search criteria were then confirmed by pre-processing chromatograms using AMDIS (version 2.72) and subsequently NIST14 library search. This meant the identifications conformed to MSI level 2[35]. It is important to note that the Tedlar bag method used in this study may not be adopted in other multi-site studies where other sampling devices may be preferred[36]. Both known and unknown VOCs were therefore evaluated as shown in the fault tree analysis in Figure 1, except for calculating LODs and linear ranges for identified compounds.
Figure 1. A fault tree analysis of breath sample VOC origin and artefact determination
Data from e-nose instruments were auto scaled (mean = 0, standard deviation = 1) and the ComBat algorithm was used to correct for systematic drift. MS data were normalised by the internal standard toluene-d8, and features were range scaled. Missing values (i.e. zero values) within the GC-MS data matrix were imputed using random forest proximity[37,38] for multivariate analyses and samples with more than 70 % missing values removed.

2.4 Univariate and multivariate analysis

To assess the significance of variation of VOCs between groups, non-parametric Kruskal-Wallis test was used, and if statistically significant, a post-hoc test was performed consisting of a Mann–Whitney U test between two groups with Bonferroni correction. With regard to assessing how breath sample VOCs respond to storage duration, any samples that would be influenced by patient groups or site variation were removed from analysis. Therefore, breath samples from adults with severe asthma from the Amsterdam site were used. These samples had a maximum storage time of 39 days and were randomised within each storage batch. Linear regression was used to determine if known exogenous and endogenous VOCs were correlated with duration of sample storage. To evaluate the effectiveness of common batch correction methods, a distance matrix calculated using Bhattacharyya distance measure was used in principal component analysis (PCA), retaining the first three PCs based on scree plot analysis, where the resulting inter-batch metric was the mean Bhattacharyya distance (0 equal to no batch effect). The metric was adopted from and described in more detail by Wehrens et al.[27].

To measure and compare dissimilarity between instruments (i.e. pairwise analysis of GC-MS versus e-nose, and e-nose versus e-nose), a Procrustes test with principal coordinates analysis (PCoA) was performed. In summary, paired samples were aligned for each feature matrix (i.e. per instrument), and subsequently converted to a distance matrix using the Jaccard measure, as described in previous studies[39,40]. PCoA was then performed on the distance matrices retaining the first 3 PCs, which explained the most variance based on scree plot analysis. As it is important to correct for the
occurrence of any negative eigenvalues when using the Jaccard measure (non-Euclidean) with PCoA (Euclidean)[39], we used Lingoes’ correction to convert negative eigenvalues to non-negative where a constant value double the value of negative eigenvalues, and added to all eigenvalues resulting in a corrected non-negative eigenvector[41].

Pairwise congruence between two instruments for the same patient sample was assessed using Procrustes analysis[42,43]. Briefly, an input matrix is geometrically transformed (i.e. rotating and stretching/shrinking) to find the optimal superimposition on a target matrix, and such that the sum of squared distances ($m^2$) is minimised and normalised (where 1 = data are completely different, and 0 = data are exactly the same). Subsequently the correlation $r$ between the two data matrices after Procrustes transformation is derived from $m^2$ ($r = \sqrt{1 - m^2}$). In order to assess the statistical significance of the similarity measured by the single Procrustes superimposition (non-symmetric), a Procrustean test with 1000 permutations is performed and an empirical $p$-value was derived by counting the number of cases where $r$ of permuted data had been higher than the one calculated from the original data and divided by the number of permutations performed (i.e. 1000). Detailed explanations and considerations of the Procrustean test applied to MS data have been published [42,43].

3. Results

3.1. Sample and metadata description

A total of 298 breath samples were collected in the study, of which 164 were first visit (baseline) samples and used in this work, each sample comprising breath from a single patient, thereby reducing any confounding effect from longitudinal samples. Baseline samples included adults with mild/moderate ($n = 11$), and severe ($n = 42$) asthma; school aged children with mild/moderate ($n = 15$), and severe ($n = 37$) asthma; and pre-school aged children with mild/moderate ($n = 21$), and severe ($n = 32$) wheeze. Clinical characteristics and definitions of asthma severity from within the U-BIOPRED consortium are described elsewhere[14–16]. The remaining samples were from healthy
subjects \((n = 6)\). Breath samples were collected at the following sites: Amsterdam \((n = 78)\), Copenhagen \((n = 19)\), London, \((n = 47)\), Manchester \((n = 12)\), and Southampton \((n = 8)\). With the exception of Copenhagen and Southampton sites, which comprised breath samples solely from children and adults (respectively), all sites provided a mixture of age and asthma subtype groups.

3.2. Determination of VOCs from breath samples

To assess the variation of VOCs using sample groups (i.e. by age, site, asthma severity), we first identified known breath VOCs (acetone, isoprene, and toluene), C5-15 alkanes, and putative asthma VOCs compiled from selected studies using TD-GC-MS profiling techniques. A final target list of 53 asthma VOCs was used to search the U-BIOPRED ion fragment matrix. From this search, 42 VOCs were extracted, of which 16 had identification confirmed (see Table 1 and Figure S2). Confirmed VOCs were classed as aldehydes \((n = 2)\), ketone \((n = 1)\), sulphide \((n = 1)\), furan \((n = 2)\), and hydrocarbons \((n = 10)\), of which seven were branched hydrocarbons.
Table 1. A list of VOCs putatively identified (MSI level 2) as biomarkers of asthma and extracted from U-BIOPRED breath sample GC-MS data. Also shown are their average molecular weight, base peak mass fragment, the predicted retention index, the KEGG description of that VOC, and the reference from which they were identified.

<table>
<thead>
<tr>
<th>VOC</th>
<th>Molecular Weight (g/mol)</th>
<th>Extracted MS fragments</th>
<th>Predicted RI</th>
<th>KEGG description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>carbon disulfide</td>
<td>76.13</td>
<td>44, 76</td>
<td>537</td>
<td>Non-specific</td>
<td>[30]</td>
</tr>
<tr>
<td>2-butane</td>
<td>72.11</td>
<td>43, 57, 72</td>
<td>577</td>
<td>Non-specific</td>
<td>[32]</td>
</tr>
<tr>
<td>2-methylpentane</td>
<td>86.18</td>
<td>57, 86</td>
<td>584</td>
<td>Lipid peroxidation</td>
<td>[29]</td>
</tr>
<tr>
<td>2-methylfuran</td>
<td>82.10</td>
<td>39, 53</td>
<td>588</td>
<td>Carcinogen</td>
<td>[28]</td>
</tr>
<tr>
<td>3-methylfuran</td>
<td>82.10</td>
<td>39, 53, 82</td>
<td>608</td>
<td>Carcinogen</td>
<td>[28]</td>
</tr>
<tr>
<td>benzene</td>
<td>78.11</td>
<td>51, 78</td>
<td>660</td>
<td>Xenobiotic</td>
<td>[31,33]</td>
</tr>
<tr>
<td>2-methylhexane</td>
<td>100.20</td>
<td>85, 100</td>
<td>677</td>
<td>Lipid peroxidation</td>
<td>[29]</td>
</tr>
<tr>
<td>octane</td>
<td>114.23</td>
<td>43, 57, 114</td>
<td>800</td>
<td>Lipid peroxidation</td>
<td>[29]</td>
</tr>
<tr>
<td>p-xylene</td>
<td>106.17</td>
<td>91, 106</td>
<td>872</td>
<td>Non-specific</td>
<td>[30,31]</td>
</tr>
<tr>
<td>cumene</td>
<td>120.20</td>
<td>105, 120</td>
<td>928</td>
<td>Carcinogen</td>
<td>[30]</td>
</tr>
<tr>
<td>octanal</td>
<td>128.22</td>
<td>128</td>
<td>978</td>
<td>Lipid peroxidation</td>
<td>[28]</td>
</tr>
<tr>
<td>2-methyloctane</td>
<td>156.31</td>
<td>43, 57, 128, 156</td>
<td>1061</td>
<td>Lipid peroxidation</td>
<td>[32]</td>
</tr>
<tr>
<td>nonanal</td>
<td>142.24</td>
<td>57</td>
<td>1084</td>
<td>Lipid peroxidation</td>
<td>[28]</td>
</tr>
<tr>
<td>dodecane</td>
<td>170.34</td>
<td>170</td>
<td>1200</td>
<td>Lipid peroxidation</td>
<td>[33]</td>
</tr>
<tr>
<td>2,6,11-trimethyldodecane</td>
<td>212.42</td>
<td>57, 71, 212</td>
<td>1275</td>
<td>Lipid peroxidation</td>
<td>[32]</td>
</tr>
<tr>
<td>2,6,10-trimethyldodecane</td>
<td>212.42</td>
<td>57, 71</td>
<td>1332</td>
<td>Lipid peroxidation</td>
<td>[29,32]</td>
</tr>
</tbody>
</table>
Known analytical artefacts including polydimethylsiloxanes, phenol, and \( N,N \)-dimethylacetamide (DMAC) were identified. Figure 2 shows a combined mass spectrum highlighting sampling and instrument artefacts using the mean of sample intensities. Reproducible and stable artefact peaks have been used to calibrate an instrument for qualitative analysis, as previously shown[44].

![Mass Spectrum](image)

**Figure 2.** A combined mass spectrum of all fragments within the MS acquisition range (m/z 29-400) and their mean percentage abundance, normalised to the most abundant peak. Highlighted are Tedlar bag sampling artefacts (phenol, and DMAC), and instrument artefacts (polydimethylsiloxane).

### 3.3. Variation from patient factors and between sites

Out of the previously identified VOCs, we found increased levels for dodecane (Figure 3a) for the London site (adults = 4, school = 16, pre-school = 20), in comparison to the Amsterdam (adults = 14, school = 11, pre-school = 10) and Manchester (adults = 8, school = 3, pre-school = 1) sites (Kruskal-Wallis \( p < 0.001 \)). Similarly, pentane (shown in Figure 3b) has shown difference (Kruskal-Wallis \( p = 0.029 \)) in intensity between London and Copenhagen sites (Mann-Whitney U, Bonferroni corrected \( p = 0.027 \)) within the pre-school cohort only.
Figure 3. Relative mass fragment intensities in patient breath samples for (a) dodecane across three sites including Amsterdam (n = 35), London (n = 40), and Manchester (n = 12), and (b) pentane across three sites including Amsterdam (n = 10), Copenhagen (n = 16), and London (n = 20). Significance between these groups are highlighted by asterisks or not significant (NS) after a Kruskal-Wallis test followed by a Mann-Whitney-U test with Bonferroni correction. Also shown for dodecane (a) are patient age groups (where i = adults, j = school children, and k = pre-school children) within each site.

Using the adult cohort only (from all sites), we found octanal to be increased (Mann–Whitney U p = 0.048) for patients with mild-to-moderate asthma (n = 11) when compared to patients with severe asthma (n = 22), as shown in Figure 4.
Figure 4. Relative mass fragment intensity showing increased octanal in patient breath samples for moderate (n = 11) compared to severe (n = 22) asthma subtypes, with a significance of p = 0.049 after a Mann-Whitney-U test.

3.4. Variation from sample storage

After analysis of storage variation, we found no significant variation for breath samples stored refrigerated for up to 39 days (Kruskal-Wallis p = 0.514). Several known VOCs were selected, and their intensity shift analysed against sample storage duration. These included several alkanes including hexane, octane, decane, dodecane, and tetradecane; common exogenous VOCs benzene and toluene, and common endogenous VOCs acetone and isoprene. Figure 5 illustrates variation of the VOCs described against sample storage duration.
Figure 5. Selected VOCs (normalised and log scaled) for samples stored for up to 39 days (mean of 28 days), annotated with $R^2$ derived from a linear fit. Samples were limited to severe asthma adult patients from the Amsterdam site.
3.5. Assessment of multivariate datasets: Comparison of batch correction methods

To compare correction methods of these batches, we adopted a method described by Wehrens et al. [27]. In the U-BIOPRED study, breath was sampled in uncontrolled batches, where one batch may be associated to one site within a patient visit period. Samples were then analysed in controlled batches, and each sample was spiked with a four-component deuterated internal standard mixture. Mean inter-batch distances (IBD) are shown in Table 2.

**Table 2. Comparison of selected normalisation methods using the mean Bhattacharyya distance within sampling and analytical batches, where low inter-batch distance indicates larger point cloud overlap between batches, which in turn means an overall low batch effect.**

<table>
<thead>
<tr>
<th>Normalisation method</th>
<th>Analytical batch distance</th>
<th>Sample batch distance</th>
<th>Combined average</th>
</tr>
</thead>
<tbody>
<tr>
<td>No normalisation</td>
<td>0.75</td>
<td>0.69</td>
<td>0.72</td>
</tr>
<tr>
<td>toluene-$d_8$</td>
<td>0.43</td>
<td>0.54</td>
<td>0.49</td>
</tr>
<tr>
<td>acetone-$d_6$</td>
<td>2.52</td>
<td>2.91</td>
<td>2.72</td>
</tr>
<tr>
<td>Sum</td>
<td>0.79</td>
<td>0.63</td>
<td>0.71</td>
</tr>
<tr>
<td>Mean</td>
<td>0.79</td>
<td>0.63</td>
<td>0.71</td>
</tr>
<tr>
<td>Median</td>
<td>0.78</td>
<td>0.75</td>
<td>0.77</td>
</tr>
<tr>
<td>Sum of squares</td>
<td>0.87</td>
<td>0.90</td>
<td>0.89</td>
</tr>
</tbody>
</table>

When compared to untransformed data (combined IBD = 0.72), we show that using an internal standard (toluene-$d_8$) improved batch correction (0.49) more than using scaling factors (between 0.71 and 0.89). In our case, scaling factors such as normalisation by sample sum or sample mean (0.71) performed better than normalisation by sample median (0.77) or sample variation (0.89). The combined IBD for acetone-$d_6$ was 2.72. Figure 6 illustrates the sampling batch distance for data
normalised by acetone-$d_6$ (Figure 6a), and toluene-$d_8$ (Figure 6b), where less distance between
eclipses is visible for the latter meaning less variation between batches.

*Figure 6. Visual representations of GC-MS data normalised by internal standards a) acetone-$d_6$ and
b) toluene-$d_8$, where batches are represented with the same colour and shape. The mean inter-batch
Bhattacharyya distance is shown for each scores plot.*

3.6. Assessment of multivariate datasets: Comparison between instruments

After central analysis by TD-GC-MS and an e-nose platform, Procrustes analysis was performed.
Table 3 shows a pairwise matrix of correlations between instruments (Procrustes error $m^2$ and 95%
confidence intervals shown in Table S3).
Table 3. Pairwise matrix of similarity correlations derived from the Procrustes test, between GC-MS and e-nose instruments.

<table>
<thead>
<tr>
<th></th>
<th>GC-ToF-MS</th>
<th>Lonestar</th>
<th>Cyanose</th>
<th>Tor Vergata</th>
<th>Comon Invent</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC-ToF-MS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lonestar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R 0.252</td>
<td></td>
<td>R 0.112</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(p = 0.132)</td>
<td></td>
<td>(p = 0.878)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyanose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R 0.112</td>
<td>R 0.183</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(p = 0.878)</td>
<td>(p = 0.204)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tor Vergata</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R 0.211</td>
<td>R 0.189</td>
<td>R 0.745</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(p = 0.255)</td>
<td>(p = 0.397)</td>
<td>(p = 0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comon Invent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R 0.173</td>
<td>R 0.355</td>
<td>R 0.392</td>
<td>R 0.216</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(p = 0.525)</td>
<td>(p = 0.003)</td>
<td>(p = 0.003)</td>
<td>(p = 0.184)</td>
<td></td>
</tr>
</tbody>
</table>

Significant values (α = 0.05) are highlighted in bold.

No significant similarities were found between GC-MS and e-nose data. For e-nose comparisons, Cyranose and Tor Vergata instruments showed correlation (R = 0.745, p = 0.001). Less strong correlations were found between Lonestar and Comon Invent (R = 0.355, p = 0.003), and between Cyranose and Comon Invent (R = 0.392, p = 0.003). To illustrate this Procrustes transformation, Figure 7 shows an example of a Procrustes superimposition plot for dissimilarity – between GC-MS and Lonestar instruments, and similarity – between Cyranose and Tor Vergata instruments.
Figure 7. Procrustes superimposition plots using distance matrices of aligned samples (paired numbers represent patient identification) where a) Lonestar data (red) transformed onto GC-MS data (blue), and b) Cyranose e-nose data (red) transformed onto Tor Vergata e-nose data.

4. Discussion

4.1. Summary of findings

In this study, we have assessed breath sample data from the recent U-BIOPRED severe asthma cohort study, and have shown possible sources of variation which must be considered when planning a large scale or multi-site study.

To achieve this, we used a targeted approach, where VOCs have been described in literature and linked to asthma, and used the KEGG database to search for their metabolic origins. It is important to stress that KEGG is curated by experts and links several chemical, biological, and pathway databases however, information of breath metabolites is limited as there are few databases that provide this information. With regard to VOCs found within U-BIOPRED breath data, many hydrocarbons may have an exogenous source and not be linked to internal metabolism, as are many VOCs found in breath samples. A KEGG search for benzene and p-xylene revealed these VOCs as
xenobiotic. Furan based cyclic compounds such as 2- and 3-methylfuran and cumene were described in KEGG as carcinogenic. These previously identified compounds may not be directly related to asthma but may constitute epiphenomena such as differing exposures or dietary influence. It is also possible that furans may be derived from aldehydes (in similarity to the Paal-Knorr furan synthesis reaction) where multiple carbonyl functional groups are reduced to form an ether functional group. Carbon disulfide, p-xylene, and 2-butanone are non-specific as they have multiple origins other than host inflammation, such as products of microbial metabolism[45], and the human airway contains a complex community of bacteria whose metabolic outputs will also be measured in breath sampling.

Data from both KEGG and the selected studies suggest that alkanes (both saturated and unsaturated) are associated with lipid peroxidation, a hallmark of host inflammation, where volatile alkanes would be breakdown products of fatty acids (FAs). In addition, volatile aldehydes or ketones are also linked to lipid peroxidation, as carbonyl group breakdown products of FAs. Patients with severe asthma had a high daily dose of corticosteroids, and their lower abundance of octanal when compared to mild/moderate (Figure 4) may be linked to suppression of lipid peroxidation. Although octanal may arise from other latent factors such as diet or age differences, this suppression is also evident in a previous study investigating breath VOCs from patients with chronic obstructive pulmonary disease (COPD) – which shares some clinical characteristics with severe asthma (such as airflow obstruction, frequent exacerbations, and high doses of inhaled steroids) – where a lower number of heavier VOCs were identified in comparison to non-COPD controls[46].

Similar volatile metabolic products, such as 2,6,11,15-tetramethyl-hexadecane and nonanal, have also been found in breath VOC profiles for patients with lower respiratory tract infections [47]. This may indicate the increased production of long chain unsaturated FAs due to a change in membrane fluidity, or defective efferocytosis in asthmatic patients [48]. Additionally, unsaturated FAs have been found not to induce cytokine release in vitro when compared to saturated FAs [49]. Shorter methylated alkanes, such as 2-methylhexane, may arise as breakdown products. It was not possible
to NIST-match several methylated alkanes because many share similar mass fragments and retention time, and therefore requires targeted analysis for accurate identification.

Sampling devices may introduce artefact VOCs depending on the type of material used, for example gas sampling bag contaminants such as phenol and N,N-dimethylacetamide [36]. Both the latter compounds were found in breath samples in this study, as breath was collected using Tedlar bags, and their associated signals data were removed to prevent a confounding effect during statistical analyses. Latent variables and confounders must be considered in multi-site studies, and it is important to assess any variation due to patient factors, asthma subtype, and sampling sites. Previous studies have shown variation in breath profiles due to geographical location[50,51], and these findings are supported by our analysis, where we found significant differences in levels of pentane and dodecane between Copenhagen and London sites. It is likely that the difference may be due to exogenous environmental alkanes before sampling breath, but this potential contamination is especially relevant as these VOCs may also be linked to lipid peroxidation. Variation between sites can also occur from differences in site-specific sampling differences or patient demographics such as age[52], or gender[53], as demonstrated using metabolomics on human serum[54]. In the example shown for dodecane, levels for pre-school and school aged children were increased compared to adults, and therefore age may be main contributing variable rather than a method-related effect, as the majority of school-aged patients used the same method as adult patients.

We found no VOCs strongly correlate with sample storage duration for up to 39 days, although we show that benzene and toluene (the known exogenous compounds used for this test) show the strongest correlation from the VOCs selected and decrease over time in breath samples stored less than 40 days. However this may be attributed to sampling, storage, and analytical conditions, where the presence of intrinsic and unidentifiable artefacts may have influenced other VOCs within storage. It is also important to state that the acetone and isoprene are not adequately trapped on
Tenax GR sorbent material. The analysis carried out was a basic measurement, and we suggest using quality assurance samples which include known breath compounds (stored and analysed alongside breath samples) to measure variation by sample storage. Previous studies have investigated the effect of sample storage on VOCs. Using e-nose devices to measure variation, van der Schee et al. found no variation where breath samples were stored for up to two weeks[18]. Kang et al. further investigated breath sample storage for a longer duration analysed using TD-GC-MS. They recommended storage duration may be extended up to 1.5 months[55], however this was specific to samples frozen at -80˚C and therefore not comparable to other studies where samples were refrigerated or stored at room temperature. In metabolomic experiments, -80˚C is recommended as metabolites do not react with each other, however this is dependent on the type of metabolites within a sample[4], and the time taken after freeze thaw and before thermal desorption.

In metabolomic studies, batch variation is intrinsic to sampling and analysis, especially for longitudinal studies. Data were batch corrected using normalisation to the internal standard. In our test we found the worst performing method was normalisation by an internal standard (acetone-\textit{d}_6, Figure 6a) unsuitable for the sorbent material used in our analysis, as it characteristics mean it is outside the ideal Tenax GR capture range. This indicates significant “over-normalisation” of data, thereby producing a false batch effect. It is important to monitor any batch variation effect to prevent misclassification. A common batch correction method is to normalise samples by an internal standard or scaling factors[27,56]. Other methods such as ComBat or retention time alignment may also help to correct known batch variation[57].

Breath samples may be analysed centrally or on-site (i.e. near patient analysis), therefore it is useful to assess intra-sample similarity between different instruments[58,59]. The expectation is instruments should show similar results for the same sample, however the instrument mechanism may overshadow any sample comparability. Analysis by GC-MS produces high resolution ion fragment data. In contrast, e-nose data is limited to the response to several sensors. This is shown in
Figure 7a, where GC-MS form additional clusters of data compared to the Lonestar instrument. Regarding e-nose comparisons, although e-nose sensing mechanisms are different, we have shown good superimposition (and therefore similarity) between Cyranose and Tor Vergata, where the reactive layer of both sensors may have similar sensing properties despite their different technology (in asthmatic patients). We have shown that the sensing mechanism may overshadow biological differences, however the response signal may also account for environmental influences such as humidity and temperatures, especially for those sensors with a polymer-based reactive layer which may also show co-linearity in the sensor drift. This is especially important for multi-site studies as clinical sites may have different parameters and limits for controlling temperature and humidity levels. Furthermore, using e-nose sensors in multi-site studies may negatively affect reproducibility for the same sensor or between different sensors. Additional controlled experiments are required to elucidate the origin of sensor responses.

4.2. Recommendations for future research

Based on our data, we provide several recommendations for breath volatile analysis for large-scale and multi-site studies from study initiation, sample collection, and sample analysis:

- To minimise variation between all sampling and analysis sites, quality assurance procedures should be incorporated into a study design. Such measures may include instrument calibration or maintenance schedule, quality monitoring, data auditing at predefined intervals with open access to metadata, and regular staff training events.

- Sampling devices (e.g. phenol and DMAC from Tedlar bags) or instrument connections (e.g. system leaks, loose connections) can be prone to contamination and therefore must be regularly serviced and/or cleaned.

- Strict storage and transportation methods should be in place where samples are stored consistently across sites without long-term storage. In addition, sample storage conditions,
dry-purging of breath samples, and transport parameters need to be defined within the study design.

- Data from multiple instruments should be compared to investigate similarities and differences in the analytical methods used.

Additional recommendations influenced from metabolomics-based literature useful for future studies are as follows:

- Perform regular instrument proficiency testing and use quality control samples to assess variation between instruments, as in previous studies[58,60].

- Consider data pre-treatment methods, for example, log transformation is not suitable for high RSD or missing values, and Pareto-scaling can be sensitive to high sample variance[61], a common occurrence in breath analysis[8].

- Consider using several models for multivariate analysis, especially for e-nose data analysis, as shown by Gromski et al. and Leopold et al. [62,63].

- Breath samples analysed with GC-MS are known to contain high numbers of metabolites that are collinear in nature, and any multivariate models must be optimised and validated prior to reporting[2,64,65].

- Consider using multiple databases, as highlighted by Vinaixa et al.[66] and the use of external standards, as recommended by the Metabolomic Standards Initiative [35,67], to verify the identification of a VOC.

- For e-nose analysis consider decentralised usage of exchangeable sensor systems, together with a central cloud-based analysis[68].

To summarise the main methodological considerations for breath sampling and analysis with focus on large-scale and multi-site studies, a cause-and-effect diagram was produced, shown in Figure 8. A checklist is also provided in the supplementary information (Table S4) for new multi-site and/or longitudinal studies.
Figure 8. Cause-and-effect diagram highlighting the main methodological considerations for large-scale and multi-site studies, where MSI is Metabolomics Standards Initiative.
5. Conclusions

Exhaled breath analysis is a rapidly developing field aiming for diagnostic use within the clinical environment. In the near future we envisage personalised treatment of people with asthma based on molecular phenotypes and detected by point-of-care sensors. However, the requirement for large scale multi-centre studies will introduce new sources of variation, and in this work, we have explored several issues that may affect the results and interpretation of such as studies. Using examples from the recent U-BIOPRED severe asthma study, we have shown the importance of assessing variation which may arise between sites, patients, and instruments (which could be exploited to obtain complementary information), with the overall aim to reduce the risk of false discoveries when interpreting results. Finally, we also submit recommendations for researchers to consider if they are performing breath analysis using metabolomics approaches.

Acknowledgements

We would like to thank and acknowledge members of the U-BIOPRED consortium members (see Table S5), industrial partners, and contributors. U-BIOPRED has received funding from the Innovative Medicines Initiative (IMI) Joint Undertaking under grant agreement no. 115010, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) and European Federation of Pharmaceutical Industries and Associations (EFPIA) companies’ in-kind contribution (www.imi.europa.eu). Additionally, we would also like to thank Lieuwe Bos (Academic Medical Centre, The Netherlands), Maxim Wilkinson (University of Manchester, United Kingdom), Tom Wenseleers (University of Leuven, Belgium), and Yuanyue Wang (Royal Philips B.V, The Netherlands) for their assistance with data pre-processing.
References


Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification Anal. Chem. 78 779–87

[26] Oksanen J 2015 Multivariate analysis of ecological communities in R: vegan tutorial R Doc. 43


[36] Beauchamp J, Herbig J, Gutmann R and Hansel A 2008 On the use of Tedlar® bags for breath-gas sampling and analysis J. Breath Res. 2 046001


[38] Stekhoven D J and Bühlmann P 2012 Missforest-Non-parametric missing value imputation for mixed-type data Bioinformatics 28 112–8


[40] AlRabiah H, Xu Y, Rattray N J W, Vaughan A a, Gibreel T, Sayqal A, Upton M, Allwood J W and Goodacre R 2014 Multiple metabolomics of uropathogenic E. coli reveal different information
content in terms of metabolic potential compared to virulence factors. Analyst 139 4193–9


