Challenge 3: Searching for exhaled breath volatile biomarkers: How can we correct for environmental contamination?

Presented by Ran Wang, Carl Whitfield, Waqar Ahmed and Stephen J Fowler, The University of Manchester

Background: Capturing volatile organic compounds (VOCs) in exhaled breath gives direct and entirely non-invasive access to a vast number of biomolecules arising directly from the airways, lung and blood, potentially reflective of both local and systemic metabolism; it is therefore potentially an ideal medium for biomarker discovery. As VOC collection is non-invasive and feasible, even in children, it has been studied in virtually all respiratory diseases including airway disease, respiratory infections, cancer and pulmonary fibrosis. However, despite this, there is lack of standardisation in sampling and analytical methods, hindering its progression to clinical application.

The problem: the level of exhaled VOCs is influenced by the concentration inspired from the environment. Despite VOC-filtered air being used during both environment and breath sample collections, this is not completely effective and we have found that the volatiles detected in the environmental samples still significantly confound the difference in VOCs detected in exhaled breaths between patient groups. This highlights the importance of needing to control for environmental contamination during the biomarker discovery in exhaled breaths.

The challenge: To date there is lack of data published around environmental contamination during exhaled breaths sampling and the best way of adjusting for this. There is also no consensus amongst breathomics researchers and industry partners regarding this important issue.

Research question(s):

1) Can a biophysical model of VOC exchange in the lung use biochemical properties (e.g. solubility, blood:air partition coefficient, metabolic production/consumption) to predict the relationship between background and exhaled concentrations?
2) What statistical relationships are observed in the data? Do they match the biophysical predictions? Are they different for asthmatic subjects?
3) Can the model be extended to account for other confounding factors (e.g. flow rates, breathing patterns, bronchoconstriction, or ventilation heterogeneity).

Data available: We have two independent datasets with samples collected using the same methods. The RADicA study dataset contains breaths and background samples from >300 adults and children with symptoms suggested of asthma and >150 healthy volunteers. We also have breaths and background samples collected in 25 patients with asthma during a randomised controlled trial to test circadian variability. This study

contains data of 6-hourly VOC collection (both breaths and background) within the same individual over 24 hours.

Subject to a data-sharing agreement, we will also be able to provide online (i.e. time-series) exhaled VOC measurements for the purposes of model development and validation.

Relevant expertise: This challenge will be particularly relevant to those with expertise in statistics, statistical modelling, data science, biomedical engineering, chemical engineering and computational modelling.