

K. Hopes,^a M. Cauchi,^{*a} C. Walton,^a H. MacQueen,^b W. Wassif^c and C. Turner^b

K. Hopes,^a M. Cauchi,^{*a} C. Walton,^a H. MacQueen,^b W. Wassif^c and C. Turner^b

Experiments into the relationship between diet and health have been an area of high interest for a long time. In this study, we investigate the application of multivariate data analysis to differentiate between rat populations fed on two different diets: normal rat diet (control) and Western affluent diet (WAD). Two sets of data were acquired and analysed: one from a biochemical clinical analyser, taking measurements of blood-based biochemical markers; the other from the analysis of the volatile organic compounds (VOCs) emitted from faecal samples from the same animals using selected ion flow tube mass spectrometry (SIFT-MS). Five classes were considered: weanlings, 12 month controls, 12 month WADs, 18 month controls, and 18 month WADs. Data from the biochemical analyser, weanlings and 18 month WAD fed rats showed significant differences from the other measurement classes. This was shown in both the exploratory analysis and through multivariate classification. Classification of control diet *versus* WAD diets suggested there are differences between classes with 92% accuracy for the 12 month classes and 91% for the 18 month classes. Cholesterol markers, especially as low density lipoprotein-cholesterol (LDL), were the main factor in influencing WAD samples. The data from the SIFT-MS analysis also produced very good classification accuracies. Classification of control diet *versus* WAD diets using the H_3O^+ precursor ion data suggested there are differences between classes with 71% accuracy for the 12 month classes and 100% for the 18 month classes. These findings confirm that total cholesterol and LDL-cholesterol are elevated in the 18 month WAD-fed rats. We therefore suggest that the analysis of VOCs from faecal samples in conjunction with multivariate data analysis may be a useful alternative to blood analysis for the detection of parameters of health.

Received 27th January 2015,
Accepted 25th February 2015

DOI: 10.1039/c5an00182j

www.rsc.org/analyst

There are recommended daily amounts of nutrients for the human body to ensure maintenance and repair of key systems and functions.^{1,2} For optimum performance, macronutrients (in the form of carbohydrates, proteins and fats) should contribute to the diet in an appropriate balance, and micronutrients should be present in sufficient – but not excessive – quantities. The quantities of certain nutrients taken in by an individual significantly affect what happens internally and can vary the probability of developing certain illnesses/diseases.

Many nutritional imbalances have been linked specifically to disease in human and rat models.³ High plasma sodium and low plasma potassium are linked to hypertension,⁴ and

there is a wealth of information linking diet with cardiovascular function.⁵ Liver and kidney function are compromised by a high fat, high carbohydrate diet,⁶ and the immune system is significantly affected by sub-optimal diets.⁷

A study concluded that a reduction in salt intake lowers blood pressure in individuals of any gender or ethnicity.⁸ Deficiencies in glucose uptake and amino acids have a deleterious effect on the immune system by negatively affecting T-cell function and reducing immune-cell activation, respectively.⁹ These factors therefore potentially compromise health. Other studies have shown that cardiovascular disease can be caused by an excessive increase in serum cholesterol levels and sugar consumption.¹⁰

Evidence of dietary effects on health can be seen on a large scale. Whole countries or regions are known to have diets that are more beneficial to an individual's health than others. The Okinawan diet generally consists of low levels of saturated fat, high antioxidant intake, and low glycaemic load. These features are likely to contribute to the decreased risk of cardio-

^a*Cranfield University, College Road, Cranfield MK43 0AL, UK.*

E-mail: m.cauchi@cranfield.ac.uk

^bThe Open University, Walton Hall, Milton Keynes, MK7 6AA, UK

^cBedford Hospital NHS Trust, Kempston Road, Bedford, MK42 9DJ, UK

vascular disease, some cancers, and other chronic diseases seen in the people of south Japan.¹¹ Another study stated that in general, the Okinawan diet resulted in longer lives.¹²

A recent study investigated using a clinical analyser reports the normal serum concentrations and activities of biochemical markers related to nutrition, inflammation and disease.¹³ Such analysers can be programmed to sequentially analyse many biochemical markers of health and disease from small volumes of clinical samples such as whole blood or serum. The authors also investigated how these parameters change with age and diet in the rat. Their findings showed that in rats fed a normal nutritionally balanced rat diet (the 'control' rats), ageing resulted in a general decrease in potassium, iron, and serum albumin concentrations. There was also a decrease in the activities of aspartate aminotransferase and alanine aminotransferase. The control rats showed an increase in total and high density lipoprotein (HDL) cholesterol with age.

Further changes were seen in rats of another group which were fed a high carbohydrate, high fat and low protein diet – a Western affluent diet.⁶ These changes were in serum concentrations of sodium, urea, creatinine and triacylglycerols (TAG), and a change in activity of alkaline phosphatase markers which can be linked to kidney, liver and cardiovascular health.

Volatile organic compounds (VOCs) are becoming more useful as markers of health and many studies have been carried out recently on the link between VOCs and health.¹⁴ For instance, there is known to be a link between acetone and blood glucose concentration in both healthy people and those with diabetes¹⁵ and a variety of other links between VOCs and disease have been noted.¹⁴ Breath analysis was initially used as a non-invasive diagnostic or monitoring tool for health and disease, but the analysis of the VOCs coming from blood, urine, faeces and other clinical samples is now becoming more widely used depending on whether a systemic or localised condition is being investigated.

There are a number of methods available for the analysis of VOCs, however mass spectrometry (with or without chromatography for compound pre-separation) is the most reliable for individual compound identification and quantification. Rapid and quantitative analysis of VOCs can be achieved using selected ion flow tube mass spectrometry (SIFT-MS).^{16,17} SIFT-MS uses a fast flow tube to study the reaction of precursor ions with sample molecules in gas or vapour form. A precursor selection quadrupole and flow tube technology enables the selected precursor ions (H_3O^+ , NO^+ , and O_2^+) to react in turn with the sample molecules to produce product ions through chemical ionisation (CI). These product ions are separated in a downstream quadrupole and are then detected and quantified. A kinetic database is then used to quantify the concentrations of various molecules present in the sample.

The particular precursor ions are chosen because they have slow reaction rates with the components of air, but react quickly with trace gases and vapours that may be used in research. This technology, unlike most CI techniques, is able to use all three reagents rapidly in turn on the same instru-

ment,¹⁸ making compound identification and quantification easier.

Searching for individual biomarkers, either in gas or vapour phase, or liquid phase has had some success, especially for some individual conditions.^{19,20} However, when looking at the majority of health conditions or other more complex systems, it has become apparent that a range of biomarkers is needed to encompass the changes that have occurred. In many cases, the changes in concentrations of these biomarkers are subtle, and a picture of changes can only really be clearly seen when a combination of biomarkers is looked at. Capturing the pattern of these changes in a meaningful way requires the use of multivariate data analysis to simplify the data by reducing the number of variables used in investigating the condition to identify those that are significant. This is because carrying out statistical analysis of individual markers is in most cases likely to only show non-significant changes which are inconclusive; combining datasets to look at overall changes is much more powerful and more likely to yield statistically significant results. Multivariate data analysis has been employed in numerous studies such as the diagnosis of bladder cancer²¹ and gastrointestinal diseases.²²

The present study employs multivariate data analysis to determine the effect that diet has on certain biomarkers over time (as the rat ages) *via* two independent laboratory instrumental techniques: biochemical analysis and SIFT-MS. This will be achieved by evaluating changes in known biomarkers in rats fed on two different diets over a period of 18 months. It is envisaged that the proposed novel methodology may prove to be more useful in the field in addition to being cheaper, faster, less invasive and more accurate. This could lead to a better understanding of the effects of diet on the body over time, and hence to a recognition of an optimal diet for each stage of life.

Experimental

Ethics

The animal work in this study was approved by the Open University Animal Welfare and Ethics Review Board as part of a project originally approved in November 2007, and reviewed annually since then. All work was carried out in accordance with the UK Animals (Scientific Procedures) Act 1986.

Sample collection

The strain of rat employed in the study was the male Sprague-Dawley. These were bred in-house and maintained in an enriched environment, on a 14h light: 10 h dark cycle. Rats were caged in Scantainers (Scanbur Technology, Denmark) at ambient temperature (19–23 °C) and 50 ± 10% humidity. They were given cardboard tubes and aspen wood blocks (LBS Biotechnology, UK). They were weaned at three weeks of age directly onto one of two diets: the experimental Western affluent diet (WAD; Western RD) or the control standard rat



Table 1 Constituents of the control (RM3) and Western affluent (WAD) diets, from manufacturer's data. The diets also contained vitamins and minerals

	RM3	WAD
Fat (% w/w)	4.25	21.40
Protein (% w/w)	22.39	17.50
Fibre (% w/w)	4.21	3.50
Ash (% w/w)	7.56	4.10
Carbohydrate (% w/w)	51.20	50.00
Energy (kcal g ⁻¹)	3.32	4.63

chow diet (RM3). Both diets were supplied by Special Diet Services, Witham, Essex, UK, and were available *ad libitum*. The WAD was high in fat and carbohydrate and low in protein, and the control diet consisted of nutritionally balanced rat feed. Constituents of the diets are compared in Table 1.

The rats were kept in pairs and their health checked regularly. In the event of one member of the pair dying prematurely before the end of the experiment the other rat was maintained on its own until use. In the majority of cases where both rats in a cage survived, they were harvested at the same time, as this is best practice for husbandry. Faecal samples were collected from cages and were therefore the product of two rats.

A total of 35 rats were used in this study. Seven weanlings were used as controls. At each time point (12 months and 18 months) 4 animals eating the control diet and 10 animals eating the experimental diet were harvested. The reason for the smaller number of control animals is that we already have substantial data on such animals and did not wish to repeat this data collection needlessly. Blood and faecal samples were taken from rats from each experimental group straight after weaning, at 12 months or at 18 months. All animals were harvested at the same time of day, 5 hours into the light phase.

The blood samples were taken by cardiac puncture after the rats were deeply anaesthetised. Coagulation occurred at room temperature, then the clots were centrifuged and the supernatants stored at -80 °C until analysis.

Biochemical data – data acquisition and analysis

The study used an automated biochemical clinical analyser – the Cobas Integra 800 analyser (Roche Diagnostics,

Mannheim, Germany) to measure the concentrations or activities of several key biochemical markers, as shown in Table 2. Samples are automatically analysed for all biochemical markers after being loaded into the instrument's autosampler. In particular the following serum electrolytes were measured: urea and creatinine to assess the renal profile; albumin, bilirubin, alkaline phosphatase (AP), aspartate aminotransferase (AST), alanine aminotransferase (ALT) and gamma glutamyl transferase (GGT) to measure liver function; high and low density lipoproteins (HDL and LDL, making up total cholesterol), and triacylglycerols (TAGs) to assess the lipid profile.

The raw data were uploaded into MATLAB R2008a (MathWorks Inc., USA). Before the data could be analysed, it was necessary to handle missing values within the data. Missing values arise most often because of the small volumes of blood harvested from the rats; since the analyser measures parameters sequentially those measured towards the end of the series are most likely to be missing as the sample is depleted by the earlier analyses.

There were 9 missing values in the “LDL-cholesterol” variable; 1 missing value in the Creatinine variable; 1 missing value in “HDL” and 1 missing value in “cholesterol: HDL” (the ratio between total cholesterol and HDL). Handling missing values is very important. Inserting zeroes is dangerous and can bias the data; likewise removal of columns and rows containing missing values within a data set could result in severe loss of useful information. In some instances, one can replace the missing values with the minimum of the value in the column or row with the assumption that the missing values are due to the levels of biomarker being below detectable levels, but this can still lead to bias in the data.

A more intuitive approach to handling missing data is to impute the missing values using an algorithm such as the “statistically inspired modification of partial least squares” (or SIMPLS) which is commonly employed in multivariate calibration.²³ This involves the creation of a mathematical model which is then used to predict the missing values in a given data column.²⁴ This is what was employed in this work.

Data pre-treatment in the form of data scaling was applied. For the biochemical data, auto-scaling was selected due to the different units of concentration or activity in each column.²⁵ Within the MATLAB application software, the PCA (principal

Table 2 Biochemical analytes investigated in this study and their relevance to clinical diagnosis. Letters in brackets are the abbreviations used below

Metabolic profile	Kidney function	Liver function	Lipid profile	Other
Sodium (NA)	Urea (UREA)	Albumin (ALB)	Total cholesterol (CHOL)	Glucose (GLC)
Potassium (K)	Creatinine (CRT)	Bilirubin	HDL-cholesterol (HDL)	Insulin (IN)
Iron (FE)		Alkaline phosphatase (AP)	LDL-cholesterol (LDL)	Testosterone (T)
Transferrin		Aspartate aminotransferase (AST; A)	Triacylglycerol (TAG)	Oestradiol (OST)
Ferritin		Alanine aminotransferase (ALT)	Cholesterol : HDL ratio (cHDL)	Cortisol (COR)
Calcium (CA)				
Adjusted calcium (ACA)		Gamma-glutamyl transferase (GGT)		
Phosphate (PO4)				
Mg (MG)				



components analysis) function²⁶ from PLS Toolbox (v3.5, Eigenvector Research Inc., USA) was used for exploratory analysis.^{24,26} PCA score plots were generated to determine the proximity of samples to one another, and thus indicate whether there could be significant differences between classes. The principal components (PCs) are in effect a new coordinate system in which the samples are projected, and in which a proportion of the total variance is captured within each PC. The PCs are thus an ordered set meaning that PC1 will contain the most variance (*e.g.* containing the most important characteristics which pertain to the most influential variables) then PC2 the next, and so on. Redundant data are captured in the higher-order PCs which tend to contain much lower variances (<1%).

To further establish the extent of the differences between groups within the biochemical data set, multivariate classification with partial least squares discriminant analysis (PLS-DA) in conjunction with leave-one-out cross-validation (LOO-CV) in which the number of latent variables (LVs) was varied from 1 to 10.^{24,27} LVs are “hidden” variables that are inferred from observable variables. All five classes were simultaneously compared then one class was compared against another: 12 month control *versus* 12 month western affluent diet (WAD), and 18 month control *versus* 18 month WAD. PLS-DA loadings were found for the lowest-order latent variables (LV) at the highest classification accuracies. These are overlaid over a PLS-DA scores plot in order to produce biplots, which aid in determining which variables (biochemical markers) influence the distribution of the samples most along the appropriate LV axes.

Finally, permutation testing was performed in which the class values were randomised for each permutation (300 times) and the mean accuracy was attained. If the mean value was found to be substantially lower than the actual accuracy calculated then the latter is deemed to be statistically significant, and thus not due to chance.

SIFT-MS VOC analysis – data acquisition

Each sample consisted of six faecal pellets from the experimental or control rats. The pellets were placed in a sample bag made from 65 mm diameter Nalophan NA tubing (Kalle UK). The bags were sealed and then filled with hydrocarbon-free air and equilibrated in an incubator at 40 °C. One end of each sample bag was connected *via* a Swagelok fitting directly to the SIFT-MS capillary inlet for analysis of the rat faecal headspace, and headspace analysis was carried out on each sample after equilibration. The SIFT-MS is a MkII model manufactured by PDZ Europa, UK.

The sample VOCs react with one of three precursor ions (H_3O^+ , NO^+ or O_2^+) to generate product ions, which are then separated *via* a quadrupole and detected and counted (in counts per second) by a channeltron detector. Thus the data obtained are in the form of counts per second at each mass-to-charge ratio (m/z), representing product ions, from m/z 10 to m/z 140 for 30 seconds using each precursor. Data generated *via* all three precursor ions were thus collected; data from the

H_3O^+ ions were analysed whilst data from the NO^+ and O_2^+ precursor ions were employed to confirm the identities of suggested various ions.

SIFT-MS VOC analysis – multivariate data analysis

There were no missing values within the SIFT-MS data. Data pre-processing involved normalisation of the m/z values against precursor ions (*e.g.* normalised against m/z 19, the H_3O^+ precursor), followed by removal of the precursor and associated ions from the dataset prior to subsequent analysis. Lastly, any m/z values where zero abundance was recorded in all samples were removed.

PCA was applied to the SIFT-MS data. Three outlying samples were identified by visualisation of the distribution of the cases within the PCA score plot (not shown) in which the outlying sample was very distant from its respective cohort; these were one weanling, one 12 month WAD and one 18 month WAD.

Multivariate classification was performed using PLS-DA in conjunction with leave-one-out cross validation (LOO-CV) to all five classes then by classifying one class against another, namely control *versus* WAD for both the 12 month and 18 month classes. PLS-DA biplots were also produced indicating the m/z of the ions that influence the distribution of the cases. Finally, permutation testing was performed as described previously.

Results

Biochemical data – exploratory data analysis

Fig. 1 shows a two-dimensional PCA scores plot of PC2 *versus* PC3 after imputation of the missing values using the SIMPLS algorithm followed by auto-scaling of the data. The circles represent the weanlings (WEAN); inverse triangles are the 12 month control diet rats (12M C); the filled triangles are 12 month old WAD (12M W); the squares are 18 month old



Fig. 1 Data from blood samples analysed by the biochemical analyser. Two dimensional PCA score plot of PC2 *vs.* PC3 showing separation achieved by PCA following replacement of missing values *via* imputation, and auto-scaling. Variances captured in parentheses.





Fig. 2 Data from clinical analyser. PCA loadings for PC1, PC2 and PC3 showing influencing biomarkers ascertained by PCA following replacement of missing values *via* imputation, and auto-scaling. Abbreviations are defined in Table 2. Variances captured in parentheses.

control rats (18M C), and the diamonds are the 18 months WAD rats (18M W). Separation between the groups is visible.

The plot shows a distribution of data mostly along PC2 suggesting increase in age from right to left. There is a lone weanling sample at the lower left quadrant of the PCA score plot; its presence was captured in the PC1 axis to an extent that it dominated the PC2 *versus* PC1 plot (not shown). There is also a lone 18 month WAD sample in the lower right quadrant whose presence is captured in the PC3 axis. These two samples are likely to be outliers. The weanlings and 18 month WAD samples are in distinctive separate groups. The 12 month WAD values overlap the 12 month and 18 month control values. The PCA loadings pertaining to PCs 1, 2 and 3 (Fig. 2) illustrate the contribution of the blood-borne biochemical markers to the distribution of the PCA scores plot.

Fig. 2 shows the largest positive and negative loadings for PC1 to PC3. In PC1, it can be seen that the UREA, alanine aminotransferase (ALT) and adjusted calcium (ACA) dominate PC1 whilst insulin (IN), Glucose (GLC), low-density lipoprotein (LDL) cholesterol, oestradiol (OST) and testosterone (T) dominate PC2. Finally, potassium (K), calcium (CA), albumin (ALB), glucose (GLC) and total cholesterol (CHOL) dominate PC3. Insulin is also opposite to glucose, *i.e.* there is a negative correlation between these two variables.

Biochemical data – multivariate classification

Partial least squares discriminant analysis (PLS-DA) was applied to the biochemical data. Simultaneous classification of the five age/diet classes was performed initially producing an overall classification accuracy of 70% at LV4. The weanlings, 12 month and 18 month WADs were classified well yet the 12 month and 18 month controls were not; this conforms to the overlapping samples observed in the PCA score plot (Fig. 1).

Application of PLS-DA to distinguish between the 12 month controls and 12 month WADs produced an overall accuracy of 92% with a specificity of 100% and a sensitivity of 89%. PLS-DA also distinguished between the 18 month controls and 18 month WADs producing an overall accuracy of 91% with a specificity of 100% and a sensitivity of 86%. The corresponding PLS-DA biplots are shown in Fig. 3A and 3B respectively.

The controls and WAD samples have separated well on opposite sides of the graph for both the 12 month and 18 month time categories. It can be seen that the 12 month



(A) 12 month



(B) 18 month

Fig. 3 PLS-DA biplots of biochemical data for control *versus* WAD for auto-scaled data pre-processed with replacement of missing values *via* imputation. (A) 12 month; (B) 18 month.



control and WAD samples are, respectively, more tightly grouped compared to the 18 month samples.

This biplot shows how much influence the loadings (individual biomarkers) have on the scores (samples).

In Fig. 3A, high-density lipoprotein (HDL) cholesterol, insulin (IN), magnesium (MG), oestradiol (OST), potassium (K) and alanine aminotransferase (ALT) influence the control group, whilst the LDL-cholesterol (LDL), testosterone (T), alkaline phosphatase (AP), aspartate aminotransferase (AST (A)), total cholesterol : HDL-cholesterol (cHDL), triacylglycerol (TAG), sodium (NA), calcium (CA), total cholesterol (CHOL) and cortisol (COR) influence the WAD group. It is also interesting to observe that glucose (GLC) appears to be non-contributory to the distribution of the samples since it is very close to the origin (Fig. 3A).

The OST and AP compounds contribute to LV1 to a greater extent than to LV2 (Fig. 3A); likewise urea (UREA) and testosterone (T) contribute to LV2 to a greater extent than to LV1. Transferrin, gamma-glutamyl transferase (GGT), bilirubin and ferritin (mentioned in Table 1) do not appear in the biplot (Fig. 3) because they each contained the same respective numerical values for each sample/case and were thus automatically omitted from the analysis during the data pre-processing step.

The permutation testing carried out to test for statistical significance for both the 12 month and 18 month time classes showed that both were statistically significant ($p \leq 0.05$) with mean class-randomised accuracies of 64% (Observed accuracy: 92%) and 62% (Observed accuracy: 91%) respectively.

SIFT-MS data – exploratory data analysis

Exploratory data analysis *via* PCA of the SIFT-MS data set was carried out. The PCA scores plot (Fig. 4) suggests that diet may be captured along PC2, because the controls and weanlings (with the exception of a 12 month control) are below the PC2

axis, whilst the WAD rats (with the exception of one 12 month WAD) are above the PC2 axis. The plot also suggests that age is captured in PC1 but it is not entirely clear.

SIFT-MS data – multivariate classification

Multivariate classification using PLS-DA with leave-one-out cross-validation was carried out on the 12 month and 18 month time points. An overall classification accuracy of 71% was attained with a specificity of 67% and a sensitivity of 75% for the 12 month time points at LV3. However, the overall accuracy attained for the 18 month time points was 100% at LV5 (100% specificity and sensitivity) which suggests that as the time progresses the distinction between the two diets increases. Fig. 5 and 6 show the PLS-DA biplots which indicate the m/z ions that contribute to the distinction between the two classes (Control *versus* WAD) for the 12-month and 18-month age groups, respectively.

The biplots in Fig. 5 and 6 indicated two main groups of loadings: one surrounding the control samples; the other surrounding the WAD samples. The m/z ions that appear to be highly influential are 83 and 18.

Fig. 5 shows that the m/z ions that influence the 12 month control samples are 43, 63, 79, 81, and 97; the m/z ions that influence the 12 month WAD samples are 18, 30, 47, 65, and 83. Fig. 6 shows that the m/z ions that influence the 18 month control samples are 17, 45, 59, and 77; the m/z ions that influence the 18 month WAD samples are 30, 36, 43, 47, 64, 65, 79 and 97. Some of these ions may be tentatively identified as acetaldehyde (m/z 45, 63, 81), acetone (m/z 59, 77), 1- or 2-propanol (m/z 43, 79, 97), acetic acid (m/z 79, 97); ammonia (m/z 18, 36), and ethanol (m/z 47, 65, 83), although others will need confirmation through the use of another technique such as gas chromatography mass spectrometry (GC-MS).

Finally, the results from the permutation testing for both the 12 month and 18 month time points were both found to be statistically significant ($p \leq 0.05$) with mean class-randomised accuracies of 64% (Observed accuracy: 71%) and 69% (Observed accuracy: 100%) respectively.

Discussion

The combination of the results from the biochemical assays of serum and the volatile organic compound data set from faecal headspace analysis using SIFT-MS shows clear differences between the blood and faecal samples from the rats fed the WAD diet and those fed the control diet, particularly at 18 months. The weanlings also showed very different results, and were well separated from the others, as seen in the PCA scores plots (Fig. 1 and 4).

Biochemical data

Examining the biochemical data, the PCA scores plots show that the weanlings are clearly grouped together using auto-scaled data. This naturally would be expected as the data comes from rats that should all physically be in the same state

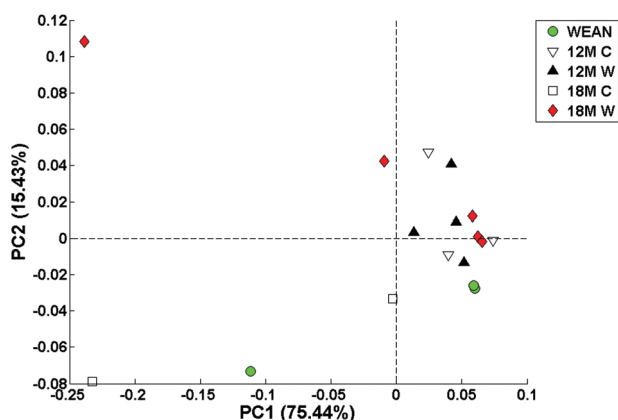
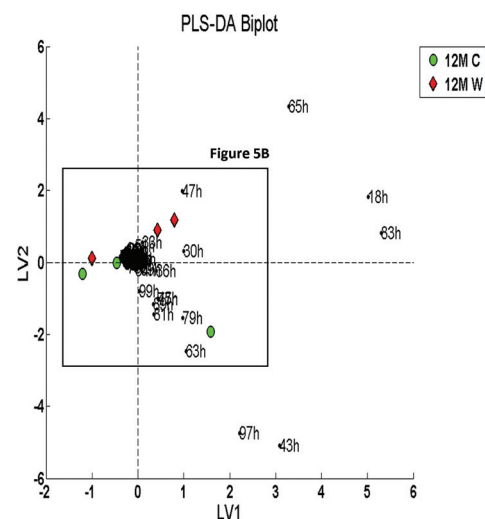
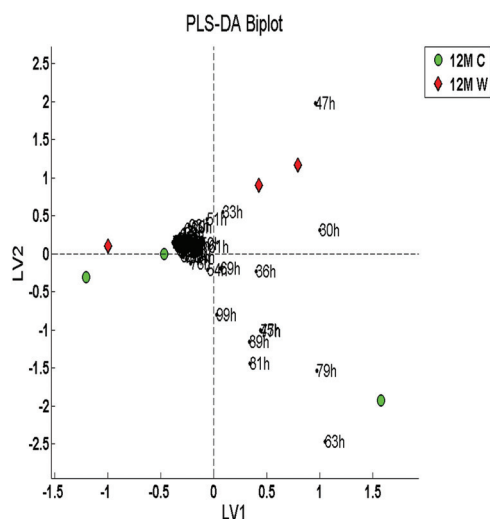


Fig. 4 PCA score plot (PC2 vs. PC1) for SIFT-MS data (following normalisation against the H_3O^+ precursor ions, removal of the precursor ions and their adducts, and removal of m/z ions containing only zeroes in all the samples) following removal of the three outlying samples. Variances captured in parentheses.



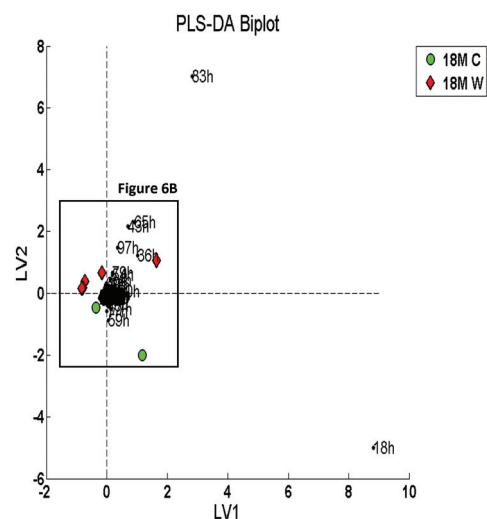


(A) 12 month

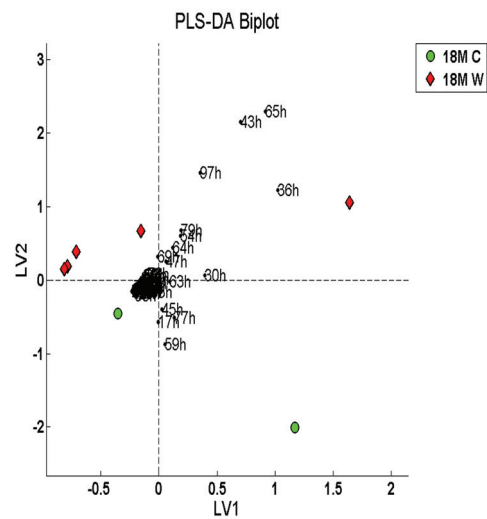


(B) 12 month (Zoomed)

Fig. 5 PLS-DA biplots of SIFT-MS data (following normalisation against the H_3O^+ precursor ions, removal of the precursor ions and their adducts, and removal of m/z ions containing only zeroes in all the samples) for control *versus* WAD: (A) 12 month; (B) 12 month (Zoomed). Data labels refer to m/z values, so for example, 18 h is m/z 18 using the H_3O^+ precursor.



(A) 18 month



(B) 18 month (Zoomed)

Fig. 6 PLS-DA biplots of SIFT-MS data (following normalisation against the H_3O^+ precursor ions, removal of the precursor ions and their adducts, and removal of m/z ions containing only zeroes in all the samples) for control *versus* WAD: (A) 18 month; (B) 18 month (Zoomed). Data labels refer to m/z values, so for example, 18 h is m/z 18 using the H_3O^+ precursor.

and have had the same feed (milk) until they were weaned onto a solid food diet. These results give us confidence that the rats in this study were all starting from a very similar physiological point, which is important for the rest of the experiment. There was, however, one weanling that was a significant distance from the others, and thus deemed an outlier. This particular sample came from rat 2, which on decoding the data was discovered to be a different strain of rat (Wistar), and added to the experiment to determine whether there were physiological differences between strains. It is clear that this does indeed give different results from the others.

Looking at the other PCA data, the 18 month WAD rats appear to mostly be grouped together at the opposite side of the graph to the weanlings (Fig. 1). This could be expected as at 18 months there has been time for many differences to have arisen compared to the physiological state of the weanlings. An 18 month WAD sample was suggested to be another outlying sample. It was subsequently confirmed that this particular rat had an obvious pathology.

Exploratory data analysis with PCA was able to place the five rat cohorts in distinctive groups. The most important biomarkers suggested by the loadings were urea (UREA), adjusted



calcium (ACA) and alanine aminotransferase (ALT). The intake of protein can affect the rate and amount of urea output.²⁸ As the WAD diet contained a low level of protein, this could have been a factor resulting in substantial differences between the amount in the control rats and the WAD rats. Furthermore, low phosphate excretion has been associated with high calcium excretion,²⁹ which may explain why the two are anti-correlated in PC3 (Fig. 2).

Multivariate classification using partial least squares discriminant analysis (PLS-DA) further confirmed the initial inferences from the exploratory analysis. Overall classification accuracies were attained at 92% and 91% for the 12 month and 18 month time periods respectively. The PLS-DA biplots give more links between the classes and the various measurements affecting them (Fig. 3). Total cholesterol (CHOL) and triacylglycerol (TAG) are among those making a strong impact on the 18 month WAD data. This would be expected from a diet high in fat. It also showed the cholesterol:HDL (cHDL) ratio – the ratio of total cholesterol to HDL-cholesterol – to be a big factor in influencing this class. Iron (FE), alkaline phosphatase (AP), and glucose (GLC) have an impact on the weanlings. Iron is normally found in large quantities in suckling rat pups.³⁰ High-density lipoprotein (HDL) cholesterol and aspartate aminotransferase (A) appear to influence the data for the 12 month control and 18 month control rats.

Low-density lipoprotein (LDL) was also shown to be a substantial factor with the 18 month WAD and also a little with the 12 month WAD. LDL cholesterol, along with other lipoproteins including HDL, enables transport of fat molecules. Elevated LDL itself is associated with health problems such as cardiovascular disease.³¹ This is in contrast to HDL, which has a cardio-protective effect. Therefore, it is not surprising to find such a significant increase in LDL measurements for the group of rats on high fat diet compared to those on the normal rat diet. HDL cholesterol is more abundant in the 12 and 18 month control rats as they are on a balanced diet likely to promote increased HDL, and indeed they have more HDL than the weanlings. Creatinine (CRT) seems to affect both 12 month and 18 month WAD. Creatinine is a good indicator of kidney health. In humans, creatinine is used alongside measurements such as a person's sex, age, weight and race to detect kidney disease.³² Increased levels are associated with renal disease.

An experiment by Nakasa and co-workers had similar findings; their research found giving rats a high fat diet resulted in an accumulation of cholesterol that was 1.6 times that of the rats fed the normal rat diet.³³ The plasma lipoproteins of this high fat group also showed an increase in LDL cholesterol and a decrease in HDL cholesterol levels. Our raw biochemical data confirm this finding. Isoprene is a common volatile constituent of many body secretions (particularly breath) and because it is linked to cholesterol synthesis it might be expected to show variations in our samples.³⁴ However isoprene is not conventionally measured by clinical analysers so we have no data relating to its levels in the blood samples of the rats used here.

Comparing one experimental class against another gave a strong indication that there was a significant difference between the data for the two different diets at 18 months.

PLS-DA demonstrated that there may be a more significant difference between the two 12 month classes than originally suspected during exploratory analysis. The loadings and biplots for this comparison suggest that triacylglycerol (TAG) and total cholesterol (CHOL) differ the most between the two. It also shows HDL is more of a factor for the 12 month control group, which is expected as high HDL is beneficial. In fatty diets, such as that taken in by the WAD rats, HDL amounts are overtaken by LDL. ALT also influences the 12 month control. The 18 month comparisons showed similar results to the 12 month comparisons, indicating some predictive value for the 12 month data; however the biplots (Fig. 3) suggest there are more variables affecting the 18 month rats. Albumin (ALB), ALT, oestradiol (OST) and insulin (IN) all appear to be factors influencing both the 12 and 18 month controls.

Lower protein intake has been linked with a lower level of serum albumin.³⁵ The 18 month control group may be influenced by a high level of albumin compared to the 18 month WAD due to having a higher level of protein in their diet. High ALT levels have been linked with high fat diets.³⁶ Hence, it would seem that the 18 month control group should only be influenced by ALT if it was at a significantly low level compared to the 18 month WAD. The 18 month WAD group seems to have many factors influencing the class, including phosphate, creatinine, triacylglycerol, magnesium, cholesterol, adjusted calcium and urea. Total cholesterol, triacylglycerol and LDL would all be expected to influence the WAD rats due to being associated with high fat diets.

SIFT-MS data

Exploratory data analysis using PCA for the SIFT-MS data yielded poorer grouping of samples by time periods (Fig. 4) compared to the biochemical data (Fig. 1).

Comparison of one class against another provided very good classification accuracies. Comparing the 18 month control against the 18 month WAD produced an overall accuracy of 100% at LV5. This is a vast improvement from comparing the 12 month control against the 12 month WAD which produced an overall classification accuracy of only 71%. This could imply that 12 months is insufficient for rats on the two diets to be clearly distinctive from one another. Although the results are not as significantly strong, this concurs with the results from the first data set (biochemical data).

The PLS-DA biplots (Fig. 5 and 6) indicated that m/z 17, m/z 45, m/z 59, m/z 63, and m/z 77 as having an influence on the 18 month control samples (Fig. 6). Acetaldehyde is represented by m/z 45, m/z 63 and m/z 81, and these ions (63 & 81) are also shown to have influence in the 12 month control samples. Acetaldehyde is produced during digestion by the oxidation of ethylene³⁷ but more commonly from the oxidation of ethanol using the enzyme alcohol dehydrogenase. Curiously, ethanol was found to have an influence on the 12 and 18 month WAD samples. It is interesting that two related compounds influ-



ence either the WAD or control samples, and this poses some interesting questions. Acetaldehyde is a much more toxic compound than ethanol,³⁸ so it is not obvious why this should influence control diets more. However, it cannot be discerned from this analysis whether the levels of acetaldehyde in the faecal headspace are higher or lower in the WAD samples and of ethanol on the control samples, merely that ions from these compounds are significant in differentiating the groups. Smith and co-workers reported that ethanol along with acetone, methanol, and dimethyl sulphide varied across a narrow concentration range when faecal headspace from six female pigs was measured by SIFT-MS (H_3O^+) but ammonia varied greatly.³⁹

Levels of acetaldehyde could be related to the nature and number of microbes or their enzymes within the digestive tract of the animal. If the diet is affecting the composition of the gut microbiota, it may thereby affect the production of acetaldehyde. In an experiment, a 2–4 fold difference in acetaldehyde was detected in rats fed on two different standard commercial diets.⁴⁰ Although the rats were not fed the same kind of diets as in this experiment, this shows the possibility of the different diets having an effect on acetaldehyde level.

Conclusion

The use of multivariate data analysis methods applied to biochemical data and headspace VOC data can tease out differences in physiological markers in rats fed a control diet, a western affluent diet (WAD), or milk (weanlings); significant changes in such markers are difficult to identify individually. Although the numbers of samples in this study are relatively small, results demonstrate that the approach of combining analysis of a suite of biochemical markers and multivariate statistics can be used in identifying the most significant markers and thus those that may best be used in relating physiological differences, including preclinical disease, to diets. These results do show a clear difference between rats fed the two different diets, and it is likely that some of these changes are detrimental to the rats fed the WAD. It is possible that similar differences would be seen in humans fed different diets, however this study cannot show this, although it does offer a way of investigating these changes in humans.

The combination of exploratory data analysis and multivariate classification has revealed differences between the data classes involved in this study. In both the biochemical data and the SIFT-MS data, there appear to be substantial differences between the two diets. The multivariate classification has led to the conclusion that there are in fact significant differences between the two groups, clear at both 12 months and 18 months when one class was classified against another. Additionally, for both datasets, PLS-DA biplots were able to give a good indication of where differences lie between the two diet groups. Identification of metabolic disturbances at an early stage, well before overt clinical signs appear, opens the door to earlier interventions and better prognosis.

Overall, there is an effect on the levels of the measured variables within the body of the rat when different diets are given over time. There are levels of particular analytes that are more abundant in the WAD rats, especially seen by 18 months, than in the control diet rats. Some of these analytes have a less positive impact on the body and its functions, and could indicate a deleterious effect of diet on the ageing rat. Indeed we have observed that the administration of WAD for similar periods of time results in an increase in tumours and other pathologies, and an overall decrease in survival in Sprague-Dawley rats, in line with the substantial clinical evidence from human nutritional studies.

The results are showing promise for both methods of analysis, in that there are significant differences in the levels of the measurements for the control diet rats and the WAD rats. The analysis of VOCs from faeces by SIFT-MS offers a viable alternative for clinical diagnosis. Arguably collection of faecal samples is less invasive than taking a blood sample or applying an endoscopic procedure, and may be more acceptable to some patients. Indeed, faecal sampling is already used in mass screening for bowel cancer in the UK, so it is possible that this form of sampling and analysis might be extended to a wider range of clinical diagnoses. Faecal samples frozen at $-20\text{ }^{\circ}\text{C}$ appear to have a 'shelf-life' of several months and this could facilitate the handling and analysis of large numbers of samples, and it is feasible that faecal samples could be routinely collected from patients and sent to a central facility for analysis to enable mass screening of patients for GI conditions.

Acknowledgements

The Open University (OU) Research Development Fund provided some funding for this project, and Karen Evans, Agata Stramek, Julia Barkans and Claire Batty provided technical support.

References

- 1 R. A. McCance and E. Widdowson, *The Composition of foods*, RSC, 7th edn, 2014.
- 2 R. M. Welch and R. D. Graham, Breeding for micronutrients in staple food crops from a human nutrition perspective, *J. Exp. Bot.*, 2004, **55**(396), 353–364.
- 3 P. M. Ridker, Inflammatory biomarkers and risks of myocardial infarction, stroke, diabetes, and total mortality: implications for longevity, *Nutr. Rev.*, 2007, **65**, S253–S259.
- 4 H. J. Adroge and N. E. Madias, Sodium and potassium in the pathogenesis of hypertension, *N. Engl. J. Med.*, 2007, **356**, 1966–1978.
- 5 K. He, k. Liu, M. L. Daviglus, *et al.*, Associations of dietary long-chain n-3 polyunsaturated fatty acids and fish with biomarkers of inflammation and endothelial activation, *Am. J. Cardiol.*, 2009, **103**, 1238–1243.



- 6 H. A. MacQueen, D. A. Sadler, S. Moore, *et al.*, Deleterious effects of a cafeteria diet on the livers of non-obese rats, *Nutr. Res.*, 2007, **27**, 38–47.
- 7 S. N. Meydani and D. Wu, Age-associated inflammatory changes: role of nutritional intervention, *Nutr. Rev.*, 2007, **65**, S213–S216.
- 8 J. H. Feng and A. M. Graham, How Far Should Salt Intake Be Reduced?, *Hypertension*, 2003, **42**, 1093–1099.
- 9 A. L. Kau, P. P. Ahern, N. W. Griffin, A. L. Goodman and J. I. Gordon, Human nutrition, the gut microbiome and the immune system, *Nature*, 2011, **474**(7351), 327–336.
- 10 B. V. Howard and J. Wylie-Rosett, Sugar; Cardiovascular Disease: A Statement for Healthcare Professionals From the Committee on Nutrition of the Council of Nutrition, Physical Activity; Metabolism of the American Heart Association, American Heart Association, Inc., *Circulation*, 2002, (106), 523–527.
- 11 D. C. Willcox, B. J. Willcox, H. Todoriki and M. Suzuki, The Okinawan Diet: Health Implications of a Low-Calorie, Nutrient-Dense, Antioxidant-Rich Dietary Pattern Low in Glycemic Load, *J. Am. Coll. Nutr.*, 2009, **28**(4), 500–516.
- 12 B. J. Willcox, C. D. Willcox, H. Todoriki, A. Fujiyoshi, K. Yano, Q. He, J. D. Curb and M. Suzuki, Caloric Restriction, the Traditional Okinawan Diet; Healthy Aging: The Diet of the World's Longest-Lived People and Its Potential Impact on Morbidity and Life Span, *Annu. N. Y. Acad. Sci.*, 2007, **1114**(10), 434–455.
- 13 H. A. MacQueen, W. S. Wassif, I. Walker, D. A. Sadler and K. Evans, Age-related biomarkers can be modulated by diet in the rat, *Food Nutr. Sci.*, 2011, **2**, 884–890, DOI: 10.4236/fns.2011.28120.
- 14 A. Amann and D. Smith, *Volatile Biomarkers: Non-invasive Diagnosis in Physiology and Medicine*, Elsevier, 2013.
- 15 C. Turner, Potential of breath and skin analysis for monitoring blood glucose concentration in diabetes, *Expert Rev. Mol. Diagn.*, 2011, **11**(5), 497–503, DOI: 10.1586/ERM.11.31.
- 16 D. Smith and P. Španěl, Selected ion flow tube mass spectrometry (SIFT-MS) for on-line trace gas analysis, *Mass Spectrom. Rev.*, 2005, **24**, 661–700.
- 17 P. Španěl and D. Smith, Progress in SIFT-MS; breath analysis and other applications, *Mass Spectrom. Rev.*, 2011, **30**, 236–267.
- 18 P. Španěl, P. Rolfe, B. Rajan and D. Smith, The Selected Ion Flow Tube (SIFT) – A Novel Technique For Biological Monitoring, *Ann. Occup. Hyg.*, 1996, **40**(6), 615–626.
- 19 C. Walton, D. Fowler, C. Turner, W. Jia, R. Whitehead, L. Griffiths, C. Dawson, R. Waring, D. B. Ramsden, J. A. Cole, M. Cauchi, C. Bessant and J. O. Hunter, Analysis of volatile organic compounds of bacterial origin in chronic gastrointestinal diseases, *Inflammatory Bowel Dis.*, 2013, **19**(10), 2069–2078.
- 20 M. Hulsmans and P. Holvoet, MicroRNAs as Early Biomarkers in Obesity and Related Metabolic and Cardiovascular Diseases, *Curr. Pharm. Des.*, 2013, **19**(32), 5704–5717.
- 21 K. K. Pasikanti, K. Esuvaranathan, P. C. Ho, R. Mahendran, R. Kamaraj, Q. H. Wu, E. Chiong and E. C. Y. Chan, Noninvasive Urinary Metabonomic Diagnosis of Human Bladder Cancer, *J. Proteome Res.*, 2010, **9**(6), 2988–2995.
- 22 M. Cauchi, C. Walton, D. Fowler, C. Turner, W. Jia, R. Whitehead, L. Griffiths, C. Dawson, R. Waring, D. B. Ramsden, J. A. Cole, C. Bessant and J. O. Hunter, Application of gas chromatography mass spectrometry (GC-MS) in conjunction with multivariate classification for the diagnosis of gastrointestinal diseases, *Metabolomics*, 2014, **10**(6), 1113–1120, DOI: 10.1007/s11306-014-0650-1.
- 23 S. de Jong, SIMPLS: An alternative approach to partial least squares regression, *Chemom. Intell. Lab. Syst.*, 1993, **18**, 251–263.
- 24 R. G. Brereton, *Applied Chemometrics for Scientists*, Wiley, Chichester, 2007.
- 25 R. van den Berg, H. Hoefsloot, J. Westerhuis, A. Smilde and M. van der Werf, Centering, scaling, and transformations: improving the biological information content of metabolomics data, *BMC Genomics*, 2006, **7**, 142.
- 26 S. Wold, K. Esbensen and P. Geladi, Principal component analysis, *Chemom. Intell. Lab. Syst.*, 1987, **2**, 37–52.
- 27 M. Barker and W. Rayens, Partial least squares for discrimination, *J. Chemom.*, 2003, **17**, 166–173.
- 28 T. K. Das and J. C. Waterlow, The rate of adaption of urea cycle enzymes, aminotransferases and glutamic dehydrogenase to changes in the dietary protein intake, *Br. J. Nutr.*, 1974, **32**(2), 353–373.
- 29 S. V. Shah, S. A. Kempson, T. E. Northrup and T. P. Dousa, Renal Adaptation to a Low Phosphate Diet in Rats, *J. Clin. Invest.*, 1979, **64**(4), 955–966.
- 30 H. H. Wouter and S. Wamberg, Milk Intake of Suckling Kittens Remains Relatively Constant from One to Four Weeks of Age, *J. Nutr.*, 2000, **130**, 77–82.
- 31 L. F. van Gaal, I. L. Mertens and C. E. De Block, Mechanisms linking obesity with cardiovascular disease, *Nature*, 2006, **444**, 875–880.
- 32 J. L. Gross, M. J. de Azevedo, S. P. Silveiro, L. H. Canani, M. L. Caramori and T. Zelmanovitz, Diabetic nephropathy: diagnosis, prevention, and treatment, *Diabetes Care*, 2005, **28**(1), 164–176.
- 33 T. Nakasa, M. Yamaguchi, O. Okinaka, K. Metori and S. Takahashi, Effects of du-zhong [*Eucommia ulmoides*] leaf extract on plasma and hepatic lipids in rats fed on a high fat plus high cholesterol diet, *J. Agric. Chem. Soc. Jpn.*, 1995, **69**(11), 1491–1498.
- 34 B. deLacy Costello, A. Amann, H. Al-Kateb, C. Flynn, W. Filipiak, T. Khalid, D. Osborne and N. Ratcliffe, A review of volatiles from the healthy human body, *J. Breath Res.*, 2014, **8**, 014001.
- 35 M. M. Jibani, L. L. Bloodworth, E. Foden, K. E. Griffiths and O. P. Galpin, Predominantly Vegetarian Diet in Patients with Incipient and Early Clinical Diabetic



- Nephropathy: Effects on Albumin Excretion Rate and Nutritional Status, *Diabetic Med.*, 1991, **8**(10), 949–953.
- 36 M. Carmiel-Haggai, A. I. Cederbaum and N. Neito, A high-fat diet leads to the progression of non-alcoholic fatty liver disease in obese rats, *FASEB J.*, 2005, **19**(1), 136–138.
- 37 K. Sano, H. Uchida and S. Wakabayashi, A new process for acetic acid production by direct oxidation of ethylene, *Catal. Surv. Asia*, 1999, **3**(1), 55–60.
- 38 C. S. Lieber, Metabolism of alcohol, *Clin. Liver Dis.*, 2001, **9**, 1–35.
- 39 D. Smith and P. Španěl, Analysis of volatile emissions from porcine faeces and urine using selected ion flow tube mass spectrometry, *Bioresour. Technol.*, 2000, **75**(1), 27–33.
- 40 H. Marchner and O. Tottmar, Influence of the Diet on the Metabolism of Acetaldehyde in Rats, *Acta Pharmacol. Toxicol.*, 2009, **38**(1), 59–71.

