RESEARCH ARTICLE

Use of metabotyping for the delivery of personalised nutrition

Clare B. O'Donovan[1], Marianne C. Walsh[1], Anne P. Nugent[1], Breige McNulty[1], Janette Walton[2], Albert Flynn[2], Michael J. Gibney[1], Eileen R. Gibney[1] and Lorraine Brennan[1,3]

[1] Institute of Food & Health, University College Dublin (UCD), Belfield, Dublin, Ireland
[2] School of Food and Nutritional Sciences, University College Cork, Cork, Ireland
[3] UCD Conway Institute of Biomolecular Research, UCD, Belfield, Dublin, Ireland

Received: August 25, 2014
Revised: November 4, 2014
Accepted: November 13, 2014

Scope: Personalised nutrition can be defined as dietary advice that is tailored to an individual. In recent years, the concept of targeted nutrition has evolved, which involves delivering specific dietary advice to a group of phenotypically similar individuals or metabotypes. This study examined whether cluster analysis could be used to define metabotypes and developed a strategy for the delivery of targeted dietary advice.

Method and results: K-means clustering was employed to identify clusters based on four markers of metabolic health (triacylglycerols, total cholesterol, direct HDL cholesterol and glucose) ($n = 896$) using data from the National Adult Nutrition Survey. A decision tree approach was developed for the delivery of targeted dietary advice per cluster based on biochemical characteristics, anthropometry and blood pressure. The appropriateness of the advice was tested by comparison with individualised dietary advice manually compiled ($n = 99$). A mean match of 89.1% between the methods was demonstrated with a 100% match for two-thirds of participants.

Conclusion: Good agreement was found between the individualised and targeted methods demonstrating the ability of this framework to deliver targeted dietary advice. This approach has the potential to be a fast and novel method for the delivery of targeted nutrition in clinical settings.

Keywords:
Cluster analysis / Decision trees / Metabotypes / Personalised nutrition / Targeted nutrition

Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

Since the mapping of the human genome in 2001 [1, 2], the concept of personalised health has evolved where an individual receives health and lifestyle advice that is dependent on their genes. In more recent years, this idea has diverged into many disciplines including personalised nutrition [3,4]. More recently, personalised nutrition has been described in terms of levels where one level builds on the foundations of another [5]. Within this model of personalised nutrition, level 1 refers to dietary advice based on assessment of an individual's diet and level 2 involves advice based on the individual's diet and phenotypic parameters such as anthropometric measures and clinical biochemistry. Level 3 is the ultimate personalisation as it refers to the delivery of dietary advice based on an individual's diet, phenotypic parameters and genetic profile [6].

Concomitant to the development of the concept of personalised dietary advice, the concept of the delivery of tailored dietary advice to a group of similar individuals has evolved [7–9], which is often referred to as targeted nutrition. Metabotyping refers to the process of grouping or stratifying individuals based on their phenotypic or metabolic profiles [10, 11]. It has previously been suggested that the identification of

Correspondence: Dr. Lorraine Brennan, UCD School of Agriculture & Food Science, Science Centre South, Belfield, Dublin 4, Ireland
E-mail: lorraine.brennan@ucd.ie
Fax: +353-1-7161147

Abbreviations: HOMA, homeostasis model assessment; **NANS**, National Adult Nutrition Survey; **QUICKI**, quantitative sensitivity check index; **TAGs**, triacylglycerols

Table 1. Identification of the three clusters

Clustering variables (mmol/L)	Cluster 1 ($N = 274$)	Cluster 2 ($N = 423$)	Cluster 3 ($N = 178$)	p Value
Triacylglycerol	**0.97** ± 0.33[3]	1.03 ± 0.38[3]	2.27 ± 0.84[1,2]	1.83×10^{-102}
Total cholesterol	5.39 ± 0.77[2,3]	**4.28** ± 0.68[1,3]	5.83 ± 0.93[1,2]	4.78×10^{-86}
HDL cholesterol	2.01 ± 0.35[2,3]	1.34 ± 0.24[1]	**1.31** ± 0.23[1]	1.21×10^{-114}
Glucose	**5.07** ± 0.65[3]	5.09 ± 0.67[3]	5.82 ± 1.13[1,2]	4.67×10^{-12}

Clusters were determined by K-means cluster analysis. Values are presented as means (±) SD. p Values adjusted for age, gender and BMI and corrected for multiple comparisons. Underlined values represent the highest value across the clusters; the bold values represent the lowest. Superscript numbers denote where the differences lie across the clusters. For example, [1] means significantly different from cluster 1.

these metabotypes may pave the way forward in terms of personalised healthcare [3, 11, 12]. A successful approach to the identification of metabotypes is the use of cluster analysis; application of this approach has clustered individuals based on similar characteristics such as biochemical markers [13]. This method of grouping individuals is commonly employed in nutrition research to identify dietary patterns in populations [14, 15]. In clinical settings, it has been applied to establish subtypes of patients with disease, for example, to identify subgroups of early Parkinson's disease patients with varying motor and non-motor symptoms [16]. Cluster analysis has also been used to classify clinical phenotypes of asthma [17], subtypes of smokers with chronic obstructive pulmonary disease [18], patients who suffer with gout [19] and chronic low back pain patient groups in primary care [20]. Moreover, this approach has been used to identify response to dietary intervention and drug treatments [13, 21, 22].

While the identification of clusters with different phenotypes has been successfully demonstrated, the use of such clusters for stratified interventions has been limited. The objectives of the present study were (i) to employ cluster analysis to define metabotypes in a population and (ii) to develop a strategy for the delivery of targeted dietary advice.

2 Materials and methods

2.1 Participants—Recruitment and sampling

The current analysis was carried out on data that were obtained from the Irish National Adult Nutrition Survey (NANS), which was a nationally representative food consumption survey. The NANS study was approved by the Human Ethics Research Committee of University College Dublin and the Clinical Research Ethics Committee of the Cork Teaching Hospitals and of University College Cork (ECM 3 (p) 4 September 2008). Adults aged 18–90 years were recruited in the Republic of Ireland (740 males, 760 females) between October 2008 and April 2010. Eligible persons were those over the age of 18 years, free-living and not pregnant or breast feeding. Written consent was obtained in accordance with the Helsinki declaration.

Anthropometric measurements were collected by trained fieldworkers and included weight, height, waist and hip circumferences and body composition in duplicate. Weight and body composition were measured using a Tanita body composition analyser BC-420MA (Tanita Ltd, GB) to the nearest 0.1 kg. Each participant also had their blood pressure measured in triplicate using an OMRON M6 Comfort blood pressure monitor. A detailed description of the data collection has been previously published [www.iuna.net, 23].

Seventy-six per cent of participants also provided a blood sample of which 79% were fasting ($n = 896$). Blood samples were collected by venepuncture by a trained phlebotomist. A total of 45 mL of blood was collected into six tubes (three serum tubes, two EDTA tubes and one lithium heparin tube) for each participant. Following collection, the tubes were gently inverted to ensure thorough mixing with anticoagulant where appropriate. Five of the tubes were kept chilled and transported to the lab for further processing and storage while the sixth tube was kept at room temperature for full blood count analysis. The blood samples reached the laboratory within 5 h of collection (time delays between 30 min and 5 h) and were processed and stored at −80°C until required for further analysis. Only serum samples were used for the present analysis and these samples were centrifuged at 1570 × g for 15 min at 4°C. Biochemistry values were assessed using a clinical bioanalyser (RX Daytona; Randox Laboratories). Details of the methods were as follows: triacylglycerol (TAG) (lipase/glycerol kinase colorimetric); total cholesterol (cholesterol oxidase); HDL (direct clearance); glucose (glucose oxidase); apolipoprotein A-1, apolipoprotein B, apolipoprotein C2, apolipoprotein C3 and apolipoprotein E (immunoturbidimetric); non-esterified fatty acids (colorimetric); serum calcium (colorimetric) and c-reactive protein (immunoturbidimetric). The following analytes were measured using ELISA kits: 25-hydroxyvitamin D (OCTEIAw25-Hydroxy Vitamin D, ImmunoDiagnostic Systems Limited), parathyroid hormone (MD Bioproducts kit), osteocalcin (Metra kit), cross-linked c-telopeptide (Nordic Bioscience kit), adiponectin (ALPCO Diagnostics kit) and leptin soluble receptor (RnD Systems kit). Cytokines and hormones (tumour necrosis factor alpha, IL1B, IL2, IL8, IL10, insulin, plasminogen activator inhibitor-1, interferon gamma, monocyte chemoattractant protein-1, vascular endothelial growth factor, vascular cell adhesion molecule 1, intercellular adhesion molecule 1, epidermal growth factor,

Table 2. Demographical information for participants per assigned cluster membership

Demographics	Cluster 1 ($N = 274$)	Cluster 2 ($N = 423$)	Cluster 3 ($N = 178$)	p Value
Age (years)	45 ± 17^{2}	**38** $\pm 17^{1,3}$	47 $\pm 16^{2}$	3.73×10^{-11}
BMI (kg/m^2)	**25.4** $\pm 4.1^{2,3}$	$27.1 \pm 5.1^{1,3}$	29.3 $\pm 4.7^{1,2}$	1.19×10^{-14}
Gender (M/F)	77/197	244/179	124/54	2.17×10^{-19}
W.C. (cm)	**85.5** $\pm 12.2^{3}$	91.5 ± 13.5	100.0 $\pm 13.4^{1}$	0.05
Sys BP (mmHg)	120.4 ± 17.9	123.5 ± 16.1	132.3 ± 17.6	0.51
Dias BP (mmHg)	76.9 ± 11.1	76.7 ± 10.1	81.7 ± 10.8	0.99
Percentage of participants with M.S.	**0.75**	6.4	35.5	6.92×10^{-32}

Clusters were determined by K-means cluster analysis. Values are presented as means (±) SD. p Values adjusted for age, gender and BMI with the exception of age, gender and BMI and corrected for multiple comparisons. Underlined values represent the highest value across the clusters; the bold values represent the lowest. Superscript numbers denote where the differences lie across the clusters. For example, [1] means significantly different from cluster 1. Dias BP, diastolic blood pressure; MS, metabolic syndrome; Sys BP, systolic blood pressure; W.C., waist circumference.

leptin, resistin, leptin soluble receptor, e-selectin, p-selectin, l-selectin, c-peptide) were measured using a biochip array system (Evidence Investigator; Randox laboratories). All samples were run in duplicate and cytokine concentrations were calculated from a calibration curve. Standard quality control procedures were followed on both analysers to ensure data integrity.

Homeostatic model assessment-IR (HOMA) score was calculated using the formula (fasting insulin μU/mL × fasting glucose mmol/L) / 22.5. Quantitative sensitivity check index (QUICKI) score was calculated as follows; 1 / (log(fasting insulin μU/mL) + log(fasting glucose mg/dL)). Diagnosis of metabolic syndrome was defined as having three or more of the following criteria as per the National Cholesterol Education Programme's Adult Treatment Panel III criteria for metabolic syndrome 2001 [24], including fasting blood glucose concentrations of 5.5—7.0 mmol/L, serum TAG concentrations of ≥1.5 mmol/, HDL < 1.0 mmol/L (men), <1.3 mmol/L (women), blood pressure ≥ 130/85 mmHg and waist circumference > 102 cm (men) and >88 cm (women). The percentage of participants with the metabolic syndrome was calculated based on the above criteria and those missing one or more the risk factors were excluded from this calculation.

2.2 Statistical analysis

Data were analysed using SPSS software package version 20.0 for Windows (SPSS, Inc). Four markers of metabolic health (TAGs, total cholesterol, HDL cholesterol and glucose) were chosen for clustering. These four variables were chosen as they are routinely measured and widely applicable markers of metabolic health. The cluster analysis was performed on fasting samples ($n = 896$), which were standardised using z-scores. Three participants were removed from the dataset due to outlying glucose values (>10 mmol/L). Two step cluster analysis revealed three clusters and K-means analysis was then used to characterise these clusters.

Table 3. Comparison of parameters relating to metabolic health across the clusters

Markers of metabolic health	Cluster 1 ($N = 274$)	Cluster 2 ($N = 423$)	Cluster 3 ($N = 178$)	p Value
C-peptide (ng/mL)	**1.47** $\pm 1.34^{2,3}$	$2.13 \pm 2.18^{1,3}$	3.31 $\pm 2.80^{1,2}$	1.86×10^{-05}
Insulin (μIU/mL)	**7.45** $\pm 6.48^{3}$	9.06 ± 6.76^{3}	13.53 $\pm 11.07^{1,2}$	1.12×10^{-03}
HOMA	**1.76** $\pm 2.06^{3}$	2.11 ± 1.83^{3}	3.71 $\pm 4.00^{1,2}$	2.29×10^{-04}
QUICKI	0.37 $\pm 0.04^{3}$	0.36 ± 0.04^{3}	**0.33** $\pm 0.03^{1,2}$	8.63×10^{-08}
NEFA (mmol/L)	**0.62** $\pm 0.33^{3}$	0.62 ± 0.33^{3}	0.80 $\pm 0.35^{1,2}$	1.33×10^{-03}
Leptin (ng/mL)	6.21 ± 7.90	5.93 ± 7.60	6.40 ± 7.33	1.00
Leptin soluble receptor (ng/mL)	30.45 $\pm 6.95^{2,3}$	26.00 ± 6.20^{1}	**25.54** $\pm 6.85^{1}$	4.01×10^{-08}
Adiponectin (μg/mL)	8.04 $\pm 3.97^{2,3}$	5.27 ± 2.15^{1}	**4.49** $\pm 2.15^{1}$	3.72×10^{-14}
Resistin (ng/mL)	5.95 ± 2.64	6.04 ± 2.75	6.30 ± 2.77	1.00
ApoA1 (mg/dL)	181.14 $\pm 42.90^{2,3}$	**144.38** $\pm 29.46^{1}$	148.05 ± 28.96^{1}	1.87×10^{-24}
ApoB (mg/dL)	$105.98 \pm 24.91^{2,3}$	**91.93** $\pm 24.63^{1,3}$	132.77 $\pm 25.54^{1,2}$	1.68×10^{-45}
ApoC2 (mg/dL)	$4.82 \pm 1.90^{2,3}$	**4.23** $\pm 1.93^{1,3}$	7.33 $\pm 2.94^{1,2}$	1.39×10^{-27}
ApoC3 (mg/dL)	$9.81 \pm 2.73^{2,3}$	**7.99** $\pm 2.57^{1,3}$	13.13 $\pm 4.71^{1,2}$	8.35×10^{-46}
ApoE (mg/dL)	$3.27 \pm 1.21^{2,3}$	**2.65** $\pm 0.86^{1,3}$	3.61 $\pm 1.29^{1,2}$	1.87×10^{-13}

Clusters were determined by K-means cluster analysis. Values are presented as means (±) SD. p Values adjusted for age, gender and BMI and corrected for multiple comparisons. Underlined values represent the highest value across the clusters; the bold values represent the lowest. Superscript numbers denote where the differences lie across the clusters. For example, [1] means significantly different from cluster 1. ApoA1, apolipoprotein A1; ApoB, apolipoprotein B; ApoC2, apolipoprotein C2; ApoC3, apolipoprotein C3; ApoE, apolipoprotein E; HOMA, homeostasis model assessment; NEFA, non-esterified fatty acids; QUICKI, quantitative insulin sensitivity check index.

Table 4. Inflammatory markers across the three clusters

Inflammatory markers	Cluster 1 ($N = 274$)	Cluster 2 ($N = 423$)	Cluster 3 ($N = 178$)	p Value
CRP (mg/L)	2.51 ± 2.77	2.68 ± 2.80	3.07 ± 2.93	1.00
TNF alpha (pg/mL)	**6.22** ± 2.29[2,3]	6.94 ± 2.15[1,3]	7.73 ± 2.74[1,2]	7.02×10^{-03}
IFNG (pg/mL)	1.17 ± 1.04	1.16 ± 1.02	1.08 ± 0.78	1.00
MCP1 (pg/mL)	228.23 ± 93.68	232.42 ± 92.51	252.29 ± 84.56	1.00
VEGF (pg/mL)	140.53 ± 114.65	131.99 ± 105.94	155.10 ± 125.15	1.00
VCAM1 (ng/mL)	**420.56** ± 124.67[2]	460.20 ± 136.70[1]	460.04 ± 147.76	0.05
ICAM1 (ng/mL)	**235.40** ± 69.24[3]	239.94 ± 81.26[3]	281.60 ± 108.82[1,2]	8.01×10^{-04}
IL1B (pg/mL)	1.54 ± 1.39	1.96 ± 2.75	1.48 ± 1.63	1.00
IL2 (pg/mL)	1.63 ± 1.68	1.71 ± 1.74	1.61 ± 1.41	1.00
IL8 (pg/mL)	10.48 ± 17.19	8.66 ± 9.08	10.45 ± 18.05	1.00
IL10 (pg/mL)	0.91 ± 1.17	1.05 ± 2.17	0.74 ± 0.48	1.00
EGF (pg/mL)	81.29 ± 38.99	84.10 ± 54.55	83.75 ± 39.34	1.00
PAI-1 (ng/mL)	**19.84** ± 7.31[3]	22.75 ± 8.51[3]	28.62 ± 9.15[1,2]	1.44×10^{-09}
E-selectin (ng/mL)	14.38 ± 6.60	15.61 ± 6.15	19.34 ± 12.83	0.29
P-selectin (ng/mL)	181.05 ± 61.84	179.11 ± 55.22	206.33 ± 67.24	1.79
L-selectin (ng/mL)	**1105.35** ± 281.34[2]	1225.46 ± 344.20[1]	1150.38 ± 323.91	0.02

Clusters were determined by K-means cluster analysis. p Values adjusted for age, gender and BMI and corrected for multiple comparisons. Values are presented as means (±) SD. Underlined values represent the highest value across the clusters; the bold values represent the lowest. Superscript numbers denote where the differences lie across the clusters. For example, [1] means significantly different from cluster 1. CRP, c-reactive protein; EGF, epidermal growth factor; ICAM1, intercellular adhesion molecule 1; IFNG, interferon gamma; MCP1, monocyte chemoattractant protein-1; PAI-1, plasminogen activator inhibitor-1; TNF alpha, tumour necrosis factor alpha; VCAM1, vascular cell adhesion molecule 1; VEGF, vascular endothelial growth factor.

Descriptive statistics (means and SDs) were performed to characterise the clusters in relation to demographical and clinical information. General linear models correcting for age, gender and BMI were performed to examine the differences across the clusters with Bonferonni post hoc tests. p Values were corrected for multiple comparisons using a Bonferonni correction where appropriate. Chi-square distributions were used to assess the gender distribution and percentage number of individuals with the metabolic syndrome across the clusters.

2.3 Development and testing of the targeted nutrition decision trees

Decision trees were manually developed based on the characteristics of the clusters with the addition of anthropometric (BMI, waist circumference) and blood pressure branches. The ability of the decision tree approach to deliver nutrition advice was tested by comparing the targeted approach with an individual-based approach delivered by a dietician, where a subset of participants were selected at random from the NANS cohort ($n = 99$) and individualised dietary advice was manually complied using the anthropometric and biochemical data available for each participant. The individual dietary advice (method 1) was then compared to the targeted advice (method 2) to assess the level of agreement between the two methods.

In order to compare the two methods effectively, the advice was categorised into themes such as bodyweight reduction, reduction of high fat foods and salt reduction (the full list of messages is shown in Table 7). Consistency between method 1 and method 2 was rated for each individual based on the themes of advice they received. Each theme of dietary advice was considered as a single message and the agreement/disagreement between the methods was expressed as a percentage of matching messages (no. of matched messages / no. of potential matched messages × 100). The actual number of messages given on each occasion using either method was also assessed (M2 / M1 × 100, where M1 = method 1 and M2 = method 2).

Table 5. Examination of markers of bone health across the clusters

Markers of bone health	Cluster 1 ($N = 274$)	Cluster 2 ($N = 423$)	Cluster 3 ($N = 178$)	p Value
Calcium (mmol/L)	2.45 ± 0.16	2.44 ± 0.13	2.46 ± 0.15	1.00
25(OH)D (nmol/L)	62.81 ± 25.53	60.83 ± 24.03	55.50 ± 21.27	1.00
PTH (pg/mL)	40.14 ± 23.69	37.33 ± 21.48	45.06 ± 24.13	1.00
OC (ng/mL)	10.88 ± 3.39	11.86 ± 4.20	9.94 ± 2.98	0.25
CTx (ng/mL)	0.44 ± 0.24	0.51 ± 0.33	0.39 ± 0.20	1.00

Clusters were determined by K-means cluster analysis. Values are presented as means (±) SD. p Values adjusted for age, gender and BMI and corrected for multiple comparisons. CTx, cross-linked c-telopeptide; OC, osteocalcin; PTH, parathyroid hormone; 25(OH)D, 25-hydroxyvitamin D.

3 Results

3.1 Identification and characterisation of the clusters

Two-step cluster analysis revealed the presence of three clusters in the dataset. Using this information, K-means cluster analysis was used to classify subjects into the three clusters. The clustering variables (TAG, total cholesterol, HDL cholesterol and glucose) were all found to be significantly different across the clusters (Table 1). Cluster 1 was characterised by subjects with the highest HDL cholesterol levels (2.01 ± 0.35 mmol/L) but the lowest glucose and TAG levels. Subjects in cluster 2 had the lowest total cholesterol levels (4.28 ± 0.68 mmol/L). Cluster 3 subjects were identified as having the highest levels in terms of TAGs, total cholesterol and glucose and the lowest HDL levels.

Demographical information for each of the three clusters is presented in Table 2. Subjects in cluster 2 were identified as the youngest (38 ± 17 years). In terms of anthropometric measures, BMI and waist circumference were found to be lowest in cluster 1 subjects (25.4 ± 4.1 kg/m^2 and 85.5 ± 12.2 cm, respectively). Subjects in cluster 3 were characterised as the oldest and had the highest BMI, with an average age of 47 ± 16 years and BMI of 29.3 ± 4.7 kg/m^2. The greatest prevalence of the metabolic syndrome was also found in cluster 3 subjects (35.5%). As age, gender and BMI were all found to be significantly different across the clusters, these were then controlled for throughout the analysis.

Assessment of a range of clinical biochemical parameters related to metabolic health revealed significant differences across the clusters (Table 3). Interestingly, parameters related to the metabolic syndrome were significantly different across the clusters such as non-esterified fatty acid ($p = 1.33 \times 10^{-03}$) and adiponectin ($p = 3.72 \times 10^{-14}$). The lowest levels of insulin resistance (HOMA 1.76 ± 2.06) and highest levels of insulin sensitivity (QUICKI 0.37 ± 0.04) were observed in cluster 1 subjects. In relation to cardiovascular health, cluster 2 subjects had the lowest levels of all the apolipoproteins examined (ApoA1, ApoB, ApoC2, ApoC3 and Apo E (where ApoA1, ApoB, ApoC2, ApoC3 and Apo E are apoliprotein A1, apolipoprotein B, apolipoprotein C2, apolipoprotein C3 and apolipoprotein E, respectively)). Subjects in cluster 3 were identified as having the most metabolically unfavourable profile with the highest levels of c-peptide (3.31 ± 2.80 ng/mL), insulin (13.53 ± 11.07 μIU/mL), insulin resistance (HOMA 3.71 ± 4.00) and the lowest levels of insulin sensitivity (QUICKI 0.33 ± 0.03).

A number of inflammatory markers were also investigated across the three clusters as shown in Table 4. Similar to the biochemical parameters, many of these were also found to be significantly different across the clusters including vascular cell adhesion molecule 1 ($p = 0.05$), intercellular adhesion molecule 1 ($p = 8.01 \times 10^{-04}$) and plasminogen activator inhibitor-1 ($p = 1.44 \times 10^{-09}$). Examination of parameters related to bone health metabolites revealed that there was no significant differences between the clusters as reported in Table 5.

3.2 Development and assessment of the targeted advice approach

For the delivery of targeted advice, a series of decision trees were manually constructed based on the biochemical characteristics of the clusters using the cutoffs presented in Supporting Information Table 1. Using this approach, cluster 1 subjects were given cholesterol lowering advice, cluster 2 subjects were relatively healthy and encouraged to continue a healthy lifestyle and cluster 3 subjects were given advice to help to lower TAGs, glucose and total cholesterol. The targeted dietary advice was further refined by the addition of extra branches for anthropometric and blood pressure measurements (Fig. 1). Each decision tree resulted in 12 possible dietary messages. An example of the decision trees is shown in Fig. 2. Dietary messages for each of the clusters were developed using the most up-to-date fact sheets available on the British Dietetic Association website (https://www.bda.uk.com/) and Irish Nutrition Dietetic Institute (INDI) website (https://www.indi.ie/). To assess the targeted approach a testing group was formed ($n = 99$), which was representative of the total cohort with a mean age of 45 ± 16 years and BMI of 27.2 ± 5.0 kg/m^2 and a gender distribution of 54 males to 46 females (data not shown). For each participant individualised dietary advice was composed and compared to the targeted dietary advice. Overall good agreement was found between the methods with a mean match between messages of 89.1% (range 20–100%) as depicted in Table 6. For 68 participants, there was a 100% match between the targeted and individual dietary advice.

In relation to the actual number of messages given on each occasion, good agreement was revealed between the methods with similar numbers of messages given for the majority of individuals as shown in Supporting Information Table 2 with an average of 109.8% where more than 100% meant that surplus information was given. Examination of the messages demonstrated excellent agreement for some of the messages (Table 7). Many of the dietary messages ($n = 9$) achieved complete agreement between the two methods including weight reduction advice, exercise and blood pressure lowering tips with an average of 94.4% (77–100%).

4 Discussion

This study describes a framework for the delivery of targeted nutrition using a combination of clustering techniques and a decision tree approach. The concept of targeted nutrition has emerged in recent years where individuals receive dietary advice at a group level; however, an effective method for its delivery has not yet been developed. Comparison of the

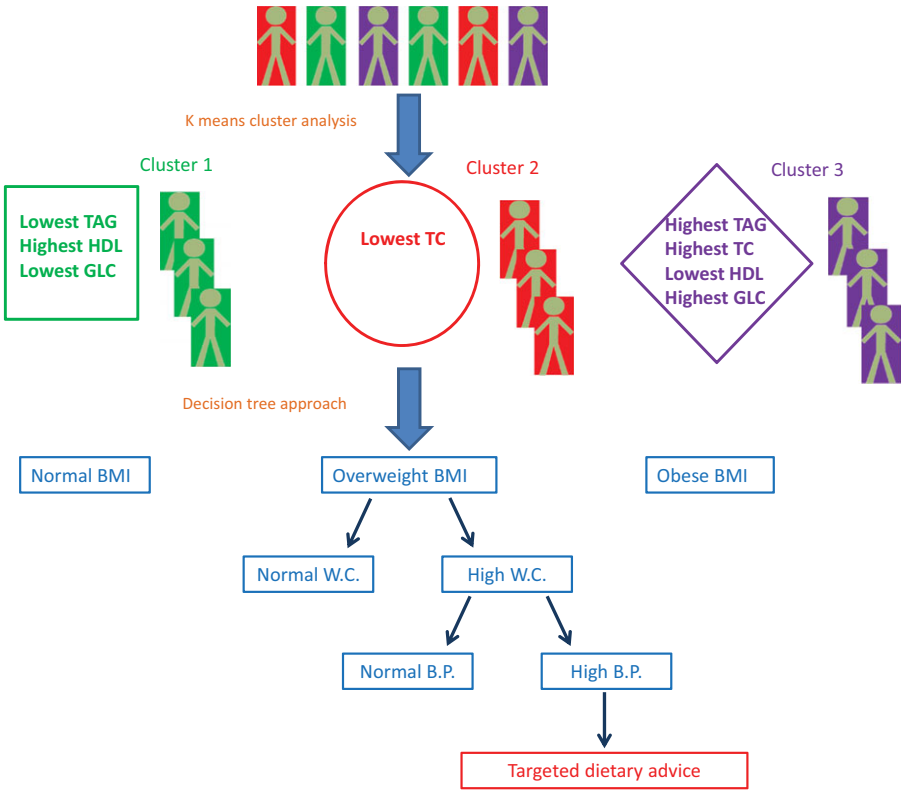

Figure 1. A framework for the delivery of targeted nutrition advice. Individuals are grouped into clusters using K-means cluster analysis. Dietary advice is then developed based on cluster specific characteristics and a decision tree approach of additional anthropometric (BMI, waist circumference) and clinical (blood pressure) branches. B.P., blood pressure; GLC, glucose; HDL, HDL cholesterol; TAGs, triacylglycerols; TC, total cholesterol; W.C., waist circumference.

targeted dietary advice to an individualised approach revealed good agreement between the methods and supports the use of the proposed framework.

The approach identified three metabolic clusters, which represented distinct phenotypes or metabotypes. The use of four commonly measured markers of metabolic health in defining these clusters indicates potential applicability for such a technique in both research and clinical settings. The application of a similar clustering approach has been successfully used to identify the response to dietary and drug interventions. Botelho and colleagues identified subgroups of dyslipidaemic individuals with varying levels of oxidative stress by using four biomarkers (malondialdehyde, ferric reducing power, 2,2-diphenyl-1-picrylhydrazyl radical and oxidised LDL) [22]. Cluster analysis has also been applied to NMR lipoprotein profiles for the identification of subgroups with varying responses to fenofibrate treatment [13]. In a dietary intervention study, a clustering approach was used to identify a cluster-specific response to vitamin D supplementation [21]. These examples together with the present results provide strong evidence for the use of clustering techniques for the identification of metabotypes.

The current metabotyping approach identified an "at-risk" cluster that had high fasting levels of TAGs, total cholesterol and glucose. Further analysis into this subset's metabolic phenotype indicates an adverse profile with high levels of insulin resistance and low levels of insulin sensitivity. Only four commonly measured biomarkers were needed to identify this at-risk group. The potential use of these four biomarkers to identify risk groups warrants further investigation for use in a clinical setting. Support for this type of approach comes from the fact that other studies have utilised cluster analysis for the identification of more specific at-risk groups in relation to cardiometabolic factors [25, 26].

While the identification of potential groups for the delivery of personalised healthcare has been previously demonstrated in a number of studies, few have proposed a method for the delivery of such group-based personalised health advice. The present paper is the first to our knowledge to develop a framework for the delivery of targeted nutrition. Decision tree algorithms have previously been used across various disciplines including healthcare settings [27, 28]. In the current paper, a decision tree method was manually employed to derive dietary advice for each of the three clusters identified. The ability of the method to deliver targeted dietary advice was then examined by comparison to individualised advice. Good agreement was found between the two methods. In the future these decision trees could be further tailored by the inclusion of food preferences such as likes/dislikes and cooking habits, which would allow the advice to be more specific to the individual. Moreover, the dietary advice could be further developed by the addition of information on other biochemical markers measured but not used for clustering purposes, for example, inflammatory markers.

The potential clinical applications of the framework described here are far reaching. Automation of the process

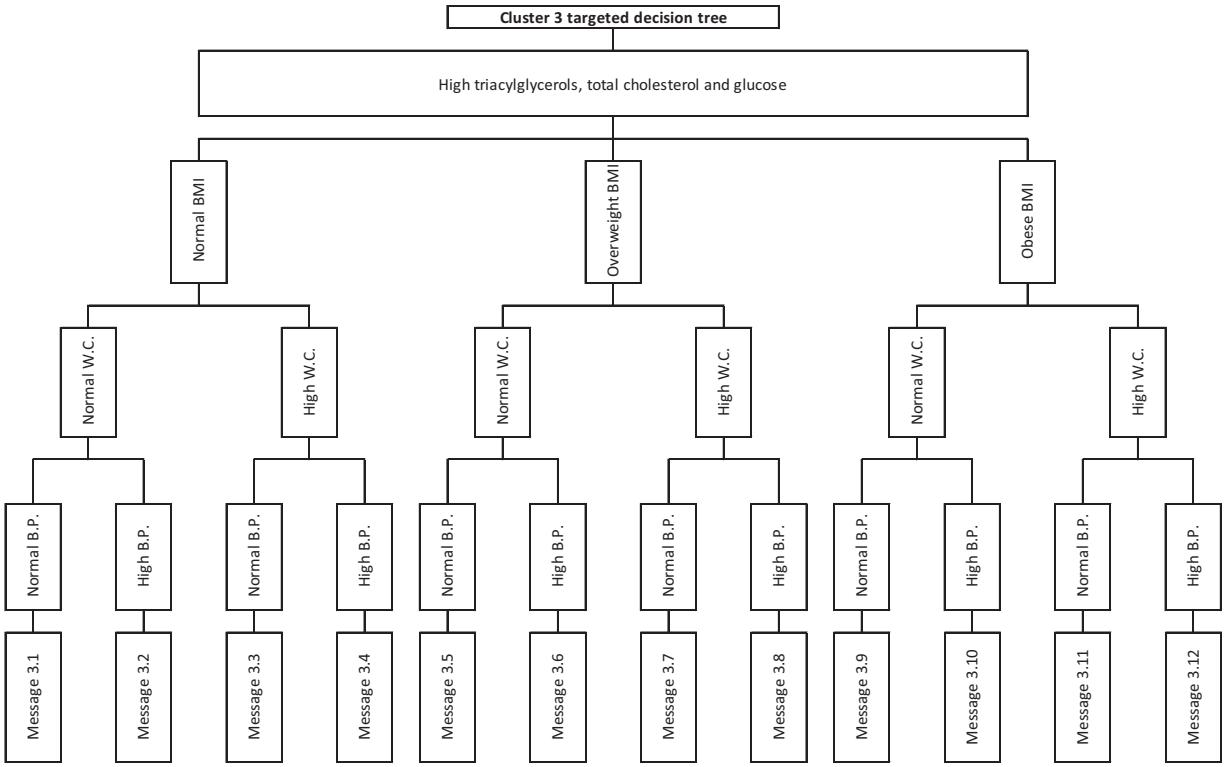

Figure 2. Example of the cluster 3 decision tree approach. Targeted dietary messages are devised based on the biochemical characteristics of cluster 3, anthropometric measures and blood pressure readings. B.P., blood pressure; W.C., waist circumference.

Table 6. Overall agreement between individualised and targeted dietary advice approaches

Percentage match	No. of participants
0–25	4
>25–50	5
>50–75	10
>75–100	80

The percentage match between method 1 (individualised) and method 2 (targeted) dietary advice presented in quartiles.

Table 7. Agreement between individualised and targeted dietary advice approaches at the message level

Dietary message	Percentage match
Maintain a healthy weight	100
Weight reduction	100
Exercise for 30 min/day	100
Exercise for 60–90 min/day	100
Follow a balanced diet	77
Portion size guidance	99
Reduce high fat foods	97
Choose low-fat products	93
Low-fat cooking advice	85
Choose lean meats	86
Eat oily fish	87
Reduce foods high in sugar	99
Reduce refined carbohydrates	89
Eat five portions of fruit and vegetables	88
Whole grains	88
Cooking tips for low salt	100
Remove salt from table	100
Avoid processed foods	100
Alcohol recommendations	100
Caffeine lowering	100

The percentage number of times each dietary message was given using both individual (method 1) and targeted dietary advice (method 2).

could allow rapid delivery of nutrition advice in a variety of clinical settings. For example, it could facilitate patient access to tailored dietary advice where access to a dietician is limited. Furthermore, it could provide a framework for faster delivery of dietary advice to individuals who are not high priority dietetic patients. The use of routinely measured markers would enable transfer of this approach to the primary care setting where in conjunction with routine blood checks one could obtain tailored dietary advice. The primary care setting is an ideal location for the delivery of nutrition advice where two-thirds of the population visit their general practitioner (GP) at least once a year and with 90% visiting every 5 years [29]. There is a large body of evidence in the literature describing health

professionals' opinions and attitudes to delivering dietary advice [30–32]. A recent study investigating patient attitudes to lifestyle counselling reported that the majority of patients would like to receive more support from their GP in relation to nutrition [29]. However, in a study of more than 2000 GPs across Europe, significant gaps were highlighted between GPs' knowledge and practices in the use of evidence-based recommendations for health promotion and disease prevention [31]. The same study reported that more than half of GPs believe that carrying out prevention and health-promotion activities are difficult [31]. The authors reported that two most important barriers to delivering nutrition advice were heavy workload/lack of time and no reimbursement [31]. Other barriers previously reported include lack of training, skills and patients motivation [33]. However, the framework described in the present study would allow the fast delivery of nutrition advice that is tailored to patients without the requirement of vast amounts of training or in-depth nutrition knowledge.

One of the important strengths of this work is that it has been performed on a large sample size, which was representative of a population group. The clustering variables used are routinely measured biomarkers of metabolic health making the framework easily transferable to a clinical or primary care setting. One of the limitations of the study is that this cohort was essentially quite a healthy sample. Further demonstration of this approach in a diseased or an at-risk cohort will be an important future step.

In conclusion, a framework for the delivery of targeted nutrition has been developed and tested successfully. Good agreement with an individualised approach demonstrates clearly the utility of this approach for delivering more tailored dietary advice. Future work will allow automation of this framework and enable successful delivery of nutrition advice in a fast automated fashion. Translation into use in clinical practice or primary care will be an important future development with the potential of improving health promotion and disease prevention.

C.O.D. and M.C.W. are funded by the European Union FP7 project Food4Me (contract KBBE.2010.2.3–02, Project no. 265494). Data used in the analysis were supported by joint funding from the Irish Department of Agriculture, Fisheries and Food and the Health Research Board under the Food for Health Research Initiative (2007–2012).

The authors have declared no conflict of interest.

5 References

[1] Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C. et al., Initial sequencing and analysis of the human genome. *Nature* 2001, *409*, 860–921.

[2] Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W. et al., The sequence of the human genome. *Science* 2001, *291*, 1304–1351.

[3] Kaput, J. Nutrigenomics research for personalized nutrition and medicine. *Curr. Opin. Biotechnol.* 2008, *19*, 110–120.

[4] Fallaize, R., Macready, A. L., Butler, L. T., Ellis, J. A. et al., An insight into the public acceptance of nutrigenomic-based personalised nutrition. *Nutr. Res. Rev.* 2013, *26*, 39–48.

[5] Ronteltap, A., van Trijp, H., Berezowska, A., Goossens, J., Nutrigenomics-based personalised nutritional advice: in search of a business model? *Genes Nutr.* 2013, *8*, 153–163.

[6] Gibney, M. J., Walsh, M. C., The future direction of personalised nutrition: my diet, my phenotype, my genes. *Proc. Nutr. Soc.* 2013, *72*, 219–225.

[7] German, J. B., Bauman, D. E., Burrin, D. G., Failla, M. L. et al., Metabolomics in the opening decade of the 21st century: building the roads to individualized health. *J. Nutr.* 2004, *134*, 2729–2732.

[8] Brennan, L., Session 2: personalised nutrition. Metabolomic applications in nutritional research. *Proc. Nutr. Soc.* 2008, *67*, 404–408.

[9] McNiven, E. M., German, J. B., Slupsky, C. M., Analytical metabolomics: nutritional opportunities for personalized health. *J. Nutr. Biochem.* 2011, *22*, 995–1002.

[10] Morris, C., O'Grada, C., Ryan, M., Roche, H. M. et al., Identification of differential responses to an oral glucose tolerance test in healthy adults. *PLoS One* 2013, *8*, e72890.

[11] Nicholson, J. K., Holmes, E., Kinross, J. M., Darzi, A. W. et al., Metabolic phenotyping in clinical and surgical environments. *Nature* 2012, *491*, 384–392.

[12] Nicholson, J. K., Global systems biology, personalized medicine and molecular epidemiology. *Mol. Syst. Biol.* 2006, *2*, 52.

[13] van Bochove, K., van Schalkwijk, D. B., Parnell, L. D., Lai, C. Q. et al., Clustering by plasma lipoprotein profile reveals two distinct subgroups with positive lipid response to fenofibrate therapy. *PLoS One* 2012, *7*, e38072.

[14] Bailey, R. L., Gutschall, M. D., Mitchell, D. C., Miller, C. K. et al., Comparative strategies for using cluster analysis to assess dietary patterns. *J. Am. Diet. Assoc.* 2006, *106*, 1194–1200.

[15] Lo Siou, G., Yasui, Y., Csizmadi, I., McGregor, S. E. et al., Exploring statistical approaches to diminish subjectivity of cluster analysis to derive dietary patterns: the Tomorrow Project. *Am. J. Epidemiol.* 2011, *8*, 956–967.

[16] Erro, R., Vitale, C., Amboni, M., Picillo, M. et al., The heterogeneity of early Parkinson's disease: a cluster analysis on newly diagnosed untreated patients. *PLoS One* 2013, *8*, e70244.

[17] Haldar, P., Pavord, I. D., Shaw, D. E., Berry, M. A. et al., Cluster analysis and clinical asthma phenotypes. *Am. J. Respir. Crit. Care Med.* 2008, *178*, 218–224.

[18] Castaldi, P. J., Dy, J., Ross, J., Chang, Y. et al., Cluster analysis in the COPDGene study identifies subtypes of smokers with distinct patterns of airway disease and emphysema. *Thorax* 2014, *69*, 415–422.

[19] Richette, P., Clerson, P., Perissin, L., Flipo, R.M. et al., Revisiting comorbidities in gout: a cluster analysis. *Ann Rheum. Dis.* 2015, *74*, 142–147.

[20] Viniol, A., Jegan, N., Hirsch, O., Leonhardt, C. et al., Chronic low back pain patient groups in primary care—a cross sectional cluster analysis. *BMC Musculoskelet. Disord.* 2013, *14*, 294.

[21] O'Sullivan, A., Gibney, M. J., Connor, A. O., Mion, B. et al., Biochemical and metabolomic phenotyping in the identification of a vitamin D responsive metabotype for markers of the metabolic syndrome. *Mol. Nutr. Food Res.* 2011, *55*, 679–690.

[22] Botelho, P. B., Fioratti, C. O., Abdalla, D. S., Bertolami, M. C. et al., Classification of individuals with dyslipidaemia controlled by statins according to plasma biomarkers of oxidative stress using cluster analysis. *Br. J. Nutr.* 2010, *103*, 256–265.

[23] Cashman, K. D., Muldowney, S., McNulty, B., Nugent, A. et al., Vitamin D status of Irish adults: findings from the National Adult Nutrition Survey. *Br. J. Nutr.* 2013, *109*, 1248–1256.

[24] Adult Treatment Panel III, Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults. *JAMA* 2001, *285*, 2486–2497.

[25] Zubair, N., Kuzawa, C. W., McDade, T. W., Adair, L. S., Cluster analysis reveals important determinants of cardiometabolic risk patterns in Filipino women. *Asia Pac. J. Clin. Nutr.* 2012, *21*, 271–281.

[26] Frazier-Wood, A. C., Glasser, S., Garvey, W. T., Kabagambe, E. K. et al., A clustering analysis of lipoprotein diameters in the metabolic syndrome. *Lipids Health Dis.* 2011, *10*, 237.

[27] Jarvis, S. W., Kovacs, C., Badriyah, T., Briggs, J., Mohammed, M. A. et al., Development and validation of a decision tree early warning score based on routine laboratory test results for the discrimination of hospital mortality in emergency medical admissions. *Resuscitation* 2013, *84*, 1494–1499.

[28] Zhang, T., Fulk, G. D., Tang, W., Sazonov, E. S., Using decision trees to measure activities in people with stroke. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2013, *2013*, 6337–6340.

[29] Brotons, C., Drenthen, A. J., Durrer, D., Moral, I., Beliefs and attitudes to lifestyle, nutrition and physical activity: the views of patients in Europe. *Fam. Pract.* 2012, *29*(Suppl 1), i49–i55.

[30] Brotons, C., Ciurana, R., Pineiro, R., Kloppe, P. et al., Dietary advice in clinical practice: the views of general practitioners in Europe. *Am. J. Clin. Nutr.* 2003, *77*(4 Suppl), 1048s–1051s.

[31] Brotons, C., Bjorkelund, C., Bulc, M., Ciurana, R. et al., Prevention and health promotion in clinical practice: the views of general practitioners in Europe. *Prev. Med.* 2005, *40*, 595–601.

[32] Pineiro, R., Brotons, C., Bulc, M., Ciurana, R. et al., Healthy diet in primary care: views of general practitioners and nurses from Europe. *Eur. J. Clin. Nutr.* 2005, *59*, S77–S80.

[33] Hiddink, G. J., Hautvast, J. G., van Woerkum, C. M., Fieren, C. J. et al., Nutrition guidance by primary-care physicians: perceived barriers and low involvement. *Eur. J. Clin. Nutr.* 1995, *49*, 842–851.

