Title: Classifying individuals into a dietary pattern based on metabolomic data.

Authors:

Orla Prendiville[1,2], Janette Walton[3,4], Albert Flynn[3], Anne P. Nugent[5], Breige A. McNulty[1], Lorraine Brennan[1,2]

Authors Affiliations:

[1]UCD School of Agriculture and Food Science, UCD Institute of Food and Health, University College Dublin, Belfield, Dublin 4, Republic of Ireland

[2]UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield, Dublin 4, Republic of Ireland

[3]School of Food and Nutritional Sciences, University College Cork, Cork, Republic of Ireland

[4]Dept. Biological Sciences, Cork Institute of Technology, Cork, Ireland

[5]Institute for Global Food Security, School of Biological Sciences, Queens University Belfast, Northern Ireland.

Keywords:

Biomarkers, dietary patterns, metabolomics, reproducibility

Corresponding Author:

Professor Lorraine Brennan[1,2]

Institute of Food and Health and Conway Institute, School of Agriculture and Food Science, University College Dublin, Belfield, Dublin 4, Republic of Ireland

+35317166815

lorraine.brennan@ucd.ie

Abbreviations:

ADD, Average Danish Diet; AHEI, Alternative Healthy Eating Index; AMDS, Alternative Mediterranean Diet Score; DASH, Dietary Approaches to Stop Hypertension; DMA, dimethylamine; EGRac, erythrocyte glutathione reductase activation coefficient; FDR, false discovery rate; FFQ, food frequency questionnaire; NANS, National Adult Nutrition Survey; NND, New Nordic Diet; NOESY, nuclear overhauser effect spectroscopy; ppm, parts per million; TMAO, trimethylamine n-oxide; TSP, sodium trimethyl propionate.

Abstract:

Scope: The objectives are to develop a metabolomic-based model capable of classifying individuals into dietary patterns and to investigate the reproducibility of the model.

Methods and Results: *K*-means cluster analysis was employed to derive dietary patterns using metabolomic data. Differences across the dietary patterns were examined using nutrient biomarkers. The model was used to assign individuals to a dietary pattern in an independent cohort, A-DIET Confirm (n=175) at four time points. The stability of participants to a dietary pattern was assessed. Four dietary patterns were derived: Moderately Unhealthy, Convenience, Moderately Healthy and Prudent. The Moderately Unhealthy and Convenience patterns had lower adherence to the Alternative Healthy Eating Index (AHEI) and the Alternative Mediterranean Diet Score (AMDS) compared to the Moderately Healthy and Prudent patterns (AHEI = 24.5 and 22.9 versus 26.7 and 28.4, $p < 0.001$). The dietary patterns were replicated in A-DIET Confirm, with good reproducibility across four time points. The stability of participants' dietary pattern membership ranged from 25.0-61.5%.

Conclusion: The multivariate model classified individuals into dietary patterns based on metabolomic data. In an independent cohort, the model classified individuals into dietary patterns at multiple time points furthering the potential of such an approach for nutrition research.

1 Introduction

Dietary pattern analysis has emerged as an important tool in nutritional epidemiology, with many authors favouring its use, concluding that dietary patterns are capable of capturing the variation of eating patterns in a population. [1, 2] Dietary patterns can provide descriptions of dietary intakes and exposures that correspond to more representative real-life scenarios compared to looking at a single food or nutrient effect. [3, 4] However, dietary pattern analysis is underpinned by data derived from traditional dietary assessment methods which have well documented limitations. [5] Dietary biomarkers were identified as a potential objective measure to enhance dietary assessment. [6] Dietary pattern biomarkers are a subclass of dietary biomarkers which comprise of a set of food intake biomarkers and food compound intake biomarkers that are reflective of an individual's habitual dietary intake and have the capacity to distinguish between a range of dietary habits. [7]

Ideally, dietary pattern biomarkers could classify individuals into a distinct dietary pattern and allow compliance to *a priori* dietary patterns to be examined. [6] There are a limited number of studies published in the area of classifying individuals into dietary patterns based on metabolite data. O'Sullivan and colleagues successfully identified metabolomic profiles reflective of three dietary patterns through identification of metabolites associated with intakes of certain food groups. [8] Recently published work assessed metabolite profiles of four predefined healthy diet pattern scores in American post-menopausal women. [9] The work resulted in the identification of similar metabolites predictive of the four healthy diet pattern scores, with the majority of the top 5 metabolites reflecting the emphasis of fish and omega 3 consumption in the diet pattern scores. Of the studies that used metabolite data (urinary or plasma) to classify individuals into dietary patterns, most succeeded in creating multivariate models reflective of previously identified

dietary patterns. [10-14] Using urinary metabolite profiles, a multivariate model classified participants into a healthy or unhealthy dietary pattern in a cross-sectional study with 567 adults. [14] Dietary data supported the derived metabolomic-based dietary patterns. Other research was successful in validating a multivariate model in two independent populations by assessing participants' characterisation to the Dietary Approaches to Stop Hypertension (DASH) diet score based on urinary metabolite profiles. [12] Here, both independent populations supported the associations between urinary metabolite profiles and the DASH diet score.

The potential for metabolomic-based dietary pattern analysis demonstrated in the literature to-date is very promising, however, research is needed to ensure metabolomic-based models are supported by data such as nutritional status biomarkers and food intake data and applicable across populations. In addition, the generalizability of dietary patterns in terms of reproducibility over time is emerging as an important aspect and one that is essential for the future development of dietary patterns in Precision Nutrition. [15-16] Thus, the objective of the present study was to develop a multivariate model using urinary metabolomic data to classify participants into dietary patterns and employ it in an independent population group. The ability of this model to classify individuals into dietary patterns at four time points was examined.

2 Materials and Methods:

2.1 The National Adult Nutrition Survey

The study design is described elsewhere, and ethical approval was granted by the University College Cork Clinical Research Ethics Committee of the Cork Teaching Hospitals and University College Dublin Human Ethics Research Committee (ECM 3 (p) 04/11/08). [17] In brief, NANS recruited 1500 free-living males and females, aged 18 and older, and who were not pregnant or lactating. All eligible participants gave written consent according to the Helsinki declaration. A four-day semi-weighed food diary was collected to assess habitual dietary intake and anthropometric measurements including height, weight, waist and hip circumferences, body composition and blood pressure were also obtained. Upon completion of the four-day semi-weighed food diary, fasting blood and urine samples were collected. The dietary data was analysed using WISP software (Tinuviel Software). Nutrient data in WISP was based on data from McCance and Widdowson's The Composition of Foods, fifth and sixth editions and all supplementary editions. Dietary and urinary data from a subset of 567 participants who took part in the National Adult Nutrition Survey (NANS) were obtained for the present study. 600 participants were randomly selected to represent equal numbers of males and females, with 33 participants being excluded due to high ethanol and acetaminophen peaks. [14] The dietary data for this subset was collapsed into a set of food groups based on previous research. [18] The Alternative Healthy Eating Index (AHEI) and Alternative Mediterranean Diet Score (AMDS) were calculated to assess adherence to both diet pattern scores. The AHEI was scored based on work by McCullough and colleagues, while the AMDS scoring was adapted from work by Fung and colleagues. [19, 20] Details for measurements of nutrient biomarkers was previously described elsewhere. [14, 21-23]

2.2 A-DIET Confirm

A-DIET Confirm was a longitudinal study examining the habitual dietary intakes of participants across a four-month period and involved the collection of biological samples. Ethical approval was granted for A-DIET Confirm by University College Dublin Sciences Human Research Ethics Committee (LS-16-91-Gibbons-Brennan). Participants included healthy males and females, between 18 and 60 years old, who were not consuming supplements or prescribed medication (the oral contraceptive pill was allowed) and had a BMI between 18.5 and 30 kg/m^2. Exclusion criteria included pregnancy/lactation and any diagnosed health condition. Once consent was obtained, participants completed four study visits, once a month, for four consecutive months. Anthropometric measurements were collected in duplicate at each study visit including height, weight, waist, and hip circumference measurements. Dietary data was collected by administrating a 24-hour dietary recall at each visit, to assess the participants' habitual dietary intake, based on the US Department of Agriculture Automated Multiple-Pass method and followed a protocol previously described. [24] A food atlas was used to verify portion sizes if the exact amount was not known by the participant. [25] Participants collected fasting first void urine samples into 50 ml collection vessels the morning of each of their study visits. The samples were inverted, centrifuged at 1800 x g for 10 minutes at 4 ºC, aliquoted into five 1 ml eppendorfs and were stored at -80 ºC.

The 24-hour dietary recalls (n=673) were coded according to the food atlas used during data collection in relation to portion size. The dietary data was double entered independently by two researchers and cross-checked for any discrepancies in Nutritics (Dublin, Ireland). For food group analysis, each food or drink item was assigned to one of 32 predefined food groups based on previous studies (**Supporting Information Table S1**). [18, 27] As participants were asked to

avoid alcoholic beverages the day prior to each visit and exclusion criteria included the consumption of nutritional supplements, dietary intakes were analysed in A-DIET Confirm participants based on 30 food groups. In total, 191 participants were recruited for the A-DIET Confirm study. Following the first study visit, 16 participants dropped out. As a result, 175 participants completed two, three or four visits and were included for analysis.

2.3 Metabolomic analysis

Metabolomic analysis of urine samples collected from A-Diet Confirm was carried out using NMR spectroscopy. Urine samples (500 µl) were defrosted at room temperature for 30 minutes and prepared by addition of 250 µl potassium phosphate buffer (0.2 mol KH_2PO_4/l, 0.8 mol K_2HPO_4/l). After centrifugation at 5360 x g for 5 minutes at 4 ºC, 60 µl deuterium oxide, and 10 µl sodium trimethyl [2,2,3-$2H_4$] proprionate (TSP) were added to 540 µl of the supernatant. Spectra were acquired on a 600 MHZ Varian NMR Spectrometer (Varian Limited, Oxford, United Kingdom) by using the first increment of a nuclear overhauser effect spectroscopy (NOESY) pulse sequence at 25 ºC. Spectra were acquired with 16384 data points and 128 scans. Water suppression was achieved during the relaxation delay (2.5 s) and the mixing time (100 ms). ^{1}H NMR spectra were referenced to TSP and were processed manually with Chenomx NMR Suite (version 8.3, Chenomx Edmonton, Canada) by using a line broadening of 0.2 Hz, and all spectra were phase and baseline corrected. All spectra were converted into 550 spectral regions of 0.01 parts per million (ppm). The spectral regions from 0.505 ppm to 7.995 ppm were included for analysis, with the exclusion of the water region (4.0ppm to 6.0 ppm), and data were normalized to the total area of the spectral integral. Urinary metabolomic analysis was performed on the NANS samples in a similar manner to that described above, as previously detailed elsewhere. [14] In brief, urine samples (500 µl) were prepared by the addition of 250 µl potassium

phosphate buffer and centrifugation. A total of 60 µl deuterium oxide, and 10 µl TSP were added to 540 µl of the supernatant. ^{1}H-NMR spectra were acquired using the first increment of a NOESY pulse sequence with the same parameters as described above. Metabolite identification was performed using Chenomx NMR suite.

2.4 Statistical analysis

Statistical analysis was performed using IBM SPSS Statistics version 24. A total of 567 NANS urinary spectra were available and analysis of the data revealed four outlying samples which were removed for the remainder of the analysis. Using z-scores of the NANS urinary metabolomic data the number of clusters was determined using a hierarchical clustering analysis approach in SIMCA: a total of four potential clusters were identified (**Supporting Information Figure S3**). *K*-means cluster analysis was performed to derive the four clusters and roughly equal numbers of participants occupied the four clusters. This method of cluster analysis creates clusters by assuming a certain number of clusters, *k*, and using Euclidean distances, assigning participants to non-overlapping, mutually exclusive clusters. Convergence was achieved following 11 iterations with convergence defined as a maximum absolute coordinate change for any centre of 0 (**Supporting Information Table S2**). Initial cluster centres were determined using the default approach in SPSS. Analysis of the derived clusters was carried out to identify the characteristics of each cluster in relation to food group percentage energy contribution, nutrient intakes, nutrient status, diet quality scores and cluster demographics. Means and SD were derived for all data and one-way ANOVAs and chi-square test were used to identify differences. A one-way ANOVA was performed in Metaboanalyst 4.0 to identify significant spectral features (bins) from the urinary metabolomic data between the four clusters, with a false

179 discovery rate (FDR) adjusted *p*-value (based on the Benjamini-Hochberg method) ≤ 0.05
180 considered statistically significant.

181 Using cluster centers from the NANS four cluster model, the participants in A-DIET Confirm
182 (n=175) were classified using the metabolomic data at each visit. Individuals with missing data
183 for a particular timepoint were not included at that timepoint. Food group means and SD for
184 each cluster were derived at each visit. A Kruskal-Wallis test was performed to identify
185 differences in food group intakes across the four clusters, based on the average percentage
186 energy contribution means and SD. Statistical significance was determined in SPSS using *p*-
187 value ≤ 0.05. The stability of the dietary patterns in A-DIET Confirm were examined by
188 identifying the movement of participants from visit 1 to visit 2, 3, and 4. The strength of
189 association or the level of agreement of cluster membership between visit 1 (reference visit) and
190 visit 2, 3 or 4 was also assessed by obtaining Cohen's kappa, Goodman and Kruskal's gamma
191 and Kendall's tau-b correlation coefficient in SPSS. We define reproducibility of dietary patterns
192 as the similarity of the dietary patterns obtained over multiple timepoints. [28]

3 Results

3.1 Defining four dietary patterns using metabolite data

Urinary metabolomic data from 567 NANS participants were included for analysis. A total of four outliers were identified and removed from further analysis due to high hippurate, trimethylamine N-oxide and acetylsalicylate peaks. *K*-means cluster analysis was performed on the remaining participants (n= 563) using a four-cluster solution. Comparison of demographics and anthropometric characteristics between participants in each cluster identified significant differences ($p < 0.05$) in age, social class, smokers, and supplement users (**Table 1**). Analysis of the food intake data revealed that several food groups were significantly different across clusters **(Table 2**). Furthermore, analysis of nutrient status, nutrient intakes, diet quality scores and demographics indicated that the clusters identified using the metabolomic data were significantly different in food group intake, nutrient intake, and dietary pattern scores (**Table 1, Tables 3-4**). In line with the food intake data and nutrient intakes, we considered cluster 1 and cluster 2 as 'unhealthy' dietary patterns. We defined cluster 1 as Moderately Unhealthy. The participants in cluster 1 had the highest intakes of processed white meat, and low intakes of fish, fish dishes and products and potatoes, however, they had high intakes of wholemeal, brown bread and rolls, and fruit. In addition, the participants had a lower adherence to the AHEI and AMDS compared to the Moderately Healthy and Prudent dietary patterns indicating participants consumed a less healthy diet, as well as having the lowest intake of protein and vitamin B6. The participants in cluster 2 were characterised by an unfavourable nutritional profile with low folate status along with high levels of erythrocyte glutathione reductase activation coefficient (EGRac), osteocalcin, cross-linked c-telopeptide, γ-tocopherol, urinary sodium, and urinary creatinine. Food group analysis demonstrated that participants in cluster 2 had the highest intakes of

processed red meat, chips and processed potatoes, savouries, and high energy beverages. These participants also had the lowest intakes of vegetables and vegetables dishes, fruit, low fat and skimmed milks, and wholemeal, brown bread and rolls, consistent with a Convenience dietary pattern. [29, 30] Furthermore, participants in cluster 2 had the highest sodium intake and the lowest adherence to the AHEI and AMDS. Interestingly, participants in cluster 2 had the lowest mean age and the highest rate of smoking.

We described clusters 3 and 4 as 'healthy' dietary patterns, again supported by analysis of dietary data and nutrient status. Food group intakes for participants in cluster 3 included the highest contribution to energy from vegetables and vegetable dishes, fruit, and fish, fish dishes and products. In addition, those in cluster 3 had the lowest intakes of processed red meat and processed white meat. However, compared to cluster 4, participants had high intakes of chips and processed potatoes and lower dietary scores, therefore cluster 3 was identified as a Moderately Healthy dietary pattern. The characteristics of participants in cluster 4 included the lowest levels of urinary sodium along with EGRac, cross-linked c-telopeptide, γ-tocopherol and urinary creatinine, as well as the highest serum levels of folate. Cluster 4 was defined as a Prudent dietary pattern, in line with the literature. [29, 31] Food groups that significantly contributed to the intakes of participants in cluster 4 included the highest intakes of low fat and skimmed milks, high intakes of vegetables and vegetable dishes, fruit, and wholemeal, brown bread and rolls and lowest intakes of processed red meat, similar to cluster 3, savouries, chips and processed potatoes, and high energy beverages. Adherence to the AHEI and AMDS was highest in those classified into cluster 4. The participants in cluster 4 had the highest mean age, showed higher intakes of supplements, and had the lowest rate of smokers.

There was a total of 351 NMR features discriminating, significantly, across the four clusters (**Figure 1**)**.** The metabolites that differed across the four clusters included alanine, citrate, pantothenic acid, creatinine, isobutyrate, hippurate, trimethylamine n-oxide (TMAO), and leucine (**Supporting Information Table S3**).

3.2 Reproducibility and stability of dietary patterns in a longitudinal study over 4 months

Repeated measures of urine samples (two to four) were obtained from 175 participants in a longitudinal study over four months. Three participants had their urinary metabolomic data from visit 3 excluded due to poor quality data. Using the cluster centers from the NANS model, participants from A-DIET Confirm were classified into one of the four derived dietary patterns at each of the four visits using the metabolomic data. Significant differences were observed for sex between the four clusters ($p = 0.020$) (**Supporting Information Table S4**). Analysis of the dietary data demonstrated similar results in the food groups contributing to the four patterns compared to the results of the NANS dietary data analysis (**Supporting Information Table S5**). Furthermore, similar dietary patterns were obtained across the four timepoints: the foods with the highest contribution to energy were similar across time (**Figure 2-3).** Analysis of the food groups for each cluster across time revealed no major differences: only one food group in cluster 3 changed across time (**Supporting Information Tables S6-S9**). The stability of A-DIET Confirm participants' dietary pattern membership was examined based on the participants' classification at visit 1. Membership remained stable for 25 to 61.5% of participants between visit 1 and visit 2, 3, or 4 (**Table 5**). On average 29.3% of participants classified into cluster 1 at visit 1 retained their cluster membership at visit 2, 3 or 4. This was similar in cluster 3, whereby 32.1% of participants remained in the same cluster at visit 2, 3, or 4. The average level of stability for cluster 2 was 45.8%. The highest level of stability was observed in cluster 4 with on average

51.9% of participants classified at visit 1 remaining in cluster 4 at visit 2, 3, or 4. Using cluster 1 membership as the reference, Kappa statistics revealed the best agreement between visit 1 and 2 (**Table 5**).

4 Discussion

The present study successfully developed a metabolomic-based multivariate model which was capable of classifying participants into one of four dietary patterns. The dietary patterns were supported by significant differences in nutritional status, food group intakes and diet quality scores across the clusters. Importantly, this model was employed in an independent population across four time points.

K-means cluster analysis was successful in classifying 563 participants into one of four dietary patterns based on their urinary metabolomic data independent of dietary data. Paramount to this, the classification of the participants into each pattern was supported using biomarkers of nutritional status and food intake data. Of note, the classification was supported by the adherence to both the AHEI and AMDS dietary pattern scores as well as nutrient status biomarkers. The importance of this work is emphasised by the fact that there are only a limited number of studies to-date where metabolomic data has been used to derive dietary patterns, without the use of the participants' dietary intakes. In particular, the present work derived dietary patterns comparable to those of a *posteriori* approach, while previous research used metabolomic data to examine the adherence to various predefined diets. Research by Andersen and colleagues used urinary metabolomic data to develop models to assess adherence to the New Nordic diet (NND) and the Average Danish diet (ADD). [11] The study successfully classified 81% of participants into the correct dietary group. Classifying individuals into metabolomic-based dietary patterns offers

great potential in nutritional epidemiology. This approach may provide an objective method and allows us to overcome issues relating to under-reporting of dietary intakes.

An important aspect of our work is that assessment of food intake and nutrient status corroborate the metabolomic-based dietary pattern classification. Significant differences were observed across all four clusters with respect to percentage energy contribution of food groups and nutrient intakes. Clusters 1 and 2 were representative of 'unhealthy' dietary patterns, defined as a Moderately Unhealthy dietary pattern and a Convenience dietary pattern, respectively. We categorized clusters 3 (Moderately Healthy) and 4 (Prudent) as 'healthy' dietary patterns, with distinct aspects differentiating them from one another in terms of food group intakes, nutrient intakes and nutritional status. For example, cluster 3 had higher intakes of fruit and vegetable dishes compared to cluster 4, while cluster 4 had the lowest intakes of chips and processed potatoes. It is noteworthy that findings from the dietary analysis demonstrated that clusters 1 and 2 had a lower adherence to the AHEI and AMDS scores compared to clusters 3 and 4. High scores of the AHEI corresponds with a healthier diet while lower scores represent individuals following a less healthy diet. [32] This is supportive of our distinction between 'healthy' and 'unhealthy' dietary patterns. The fact that both the 'healthy' and 'unhealthy' dietary patterns, derived in previous research, could be further broken down demonstrates that there can be many variations of healthy and unhealthy diets within a population. [14] The literature to-date supports this concept of more variations of dietary patterns. Walthouwer and colleagues examined three dietary patterns, with two representing healthier diets compared to the third: healthy, moderately healthy, and unhealthy. [33] Pryer and colleagues also identified four diets within a male population ranging from "Convenience" (unhealthy) to a "Healthier" dietary pattern. [34] These distinctions may be important in guiding individuals towards an improved dietary intake. Our

study was successful in reproducing the metabolomic-based dietary patterns in an independent population. The multivariate metabolomic model was capable of classifying participants of the A-DIET Confirm study into one of the four dietary patterns, at four different time points. Additionally, the dietary data collected in the A-DIET Confirm study further supported the model developed in NANS with similar differences in food group contribution across the four clusters evident between both studies. It should be noted that there were fewer significant food group differences in the A-DIET clusters probably due to the lower number of individuals in the study. Metabolomic analysis identified metabolites including hippurate, creatinine, citrate, and tryptophan to be discriminating across the four clusters. These metabolites were also identified as discriminating metabolites between a healthy and unhealthy cluster in previously published work, with hippurate also identified to be present in higher concentrations of those following a Mediterranean diet compared with a Western diet. [14, 35] It is also important to acknowledge that the metabolomic profile will capture biomarkers that are not dietary biomarkers but potentially as a metabolic consequence of a certain diet. We consider this is the case for the endogenous metabolites, creatinine, citrate, and tryptophan. It should also be noted that the full profile was used as opposed to individual metabolites and consider that the success of the model is due in part to combination of multiple metabolites.

With the increased interest in the use of dietary patterns to demonstrate the links between diet and disease, the stability of dietary pattern membership is important because it will permit studies to account for changes in dietary habits. Understanding that individual's dietary pattern membership may fluctuate over time is key for the development of more effective public health policies and interventions, especially relating to dietary intakes associated with health outcomes. [36-38] Previous research demonstrated that examining dietary intakes at multiple time points

would help to understand the nature of changes that occurs over time in an individual's dietary pattern. [28] Our study demonstrated that stability of participants differed across the dietary patterns with stability higher for two of the dietary patterns. Participants in dietary pattern 1 and 3 displayed a greater variability in their dietary patterns: future exploitation of such variability could be used to promote a healthier diet. In the literature, the level of stability that individuals retain depends on the number of time points used in the longitudinal studies. When two timepoints were examined higher stability (66 to 73%) was reported compared to studies where three timepoints were used (41.8%). [28, 33, 39] Furthermore, these previous studies were based on two or three dietary patterns, compared to the present study, where membership of four dietary patterns was examined. Taking these observations into account the range observed in our study for dietary pattern 2 and 4 compares well to the previous literature. Stability of dietary pattern 1 and 3 was lower than previously reported but may result from a combination of use of more timepoints and dietary patterns in the present study. However, one must be cognisant of this lower stability and overall, it supports the collection of multiple samples to capture the variation in dietary intake of individuals in longitudinal studies.

A limitation of the present study relates to use of two different dietary intake instruments in the two studies. The A-DIET Confirm study used 24-hour dietary recalls which are generally reflective of recent dietary intakes compared, for example to the use of food frequency questionnaires (FFQs) (assess habitual dietary intakes over a period of time) or food diaries that offer a more detailed description of recent dietary intakes. This may have limited the identification of food group differences to support dietary pattern classification using metabolomic data. However, in the development of the model in the NANS study we used four-day food diaries and were able to identify a large number of differences in food group intake

across the clusters. Self-reported data is prone to a number of well-defined bias and in an effort to overcome this we used nutrient status biomarkers in conjunction with the dietary data to confirm the dietary patterns. *K*-means cluster analysis was selected as the method to derive dietary patterns because it can group individuals into non-overlapping clusters. [40] This was particularly useful for tracking of individuals across the timepoints. However, a limitation of k-means is that it is necessary to predefine the number of clusters. The ability to classify individuals into dietary patterns using urinary metabolomic data and the reproducibility of the model in an independent population across four time points are major strengths of our study. The results add to the limited research in the area of dietary pattern biomarkers as an objective method of dietary pattern classification. There is limited knowledge to-date surrounding the stability of dietary pattern membership and our results support the requirement for multiple time points to be examined. The ability to distinguish the derived dietary patterns supported by nutritional status and dietary data is another strength of the present study.

The present work derived a metabolomic-based multivariate model that was successful in classifying individuals into dietary patterns and was reproducible in an independent study at 4 time points. While the overall dietary patterns were reproducible, our work indicated that there was movement between dietary patterns over time. Further work is warranted to establish the importance of such movement between dietary patterns and to perform similar work in larger more diverse population groups. The ability to utilize the metabolomic-based model in a different study is an important step forward in the development of the potential of such approaches. Further development of this approach in different population groups will be important for the development of metabolomic-based dietary assessment. To conclude, the present study demonstrates the potential that metabolomic-based dietary patterns offers in the

search for an objective method of dietary assessment, however, it is now imperative that the reproducibility and stability of such dietary patterns is extensively investigated.

5 References:

[1] P.F. Jacques, K.L. Tucker, *Am. J. Clin. Nutr.* **2001**, *73*, 1.

[2] H. Barkoukis, *J. Am. Diet. Assoc.* **2007**, *107*, 234.

[3] R.L. Bailey, M.D. Gurschall, D.C. Mitchell, C.K. Miller, F.R. Lawrence, H. Smiciklas-Wright, *J. Am. Diet. Assoc.* **2006**, *106*, 1194.

[4] J.M. Harrington, D.L. Dahly, A.P. Fitzgerald, M.S. Gilthorpe, I.J. Perry, *Public Health Nutr*. **2014**, *17*, 2674.

[5] A.F. Subar, L.S. Freedman, J.A. Tooze, S.I. Kirkpatrick, C. Boushey, M.L. Neuhouser, F.E. Thompson, N. Potischman, P.M. Guenther, V. Tarasuk, J. Reedy, S.M. Krebs-Smith, *J. Nutr*. **2015**, *145*, 2639.

[6] L. Brennan, F.B. Hu, *Mol. Nutr. Food. Res.* **2019**, *63*, 449.

[7] Q. Gao, G. Praticò, A. Scalbert, G. Vergères, M. Kolehmainen, C. Manach, L. Brennan, L.A. Afman, D.S. Wishart, C. Andres-Lacueva C, M. Garcia-Aloy, H. Verhagen, E.J.M. Feskens, L.O. Dragsted, *Genes Nutr.* **2017**, *12*, 34.

[8] A. O'Sullivan, M.J. Gibney, L. Brennan, *Am. J. Clin. Nutr.* **2011**, *93*, 314.

[9] M.L. Mccullough, M.L. Maliniak, V.L. Stevens, B.D. Carter, R.A. Hodge, Y. Wang, *Am. J. Clin. Nutr.* **2019**, *109*, 1439.

[10] H.M. Lindqvist, M. Radjursoga, D. Malmodin, A. Winkvist, L. Ellegard, *Am. J. Clin. Nutr*. **2019**, *110*, 53.

[11] M.B. Andersen, A. Rinnan, C. Manach, S.K. Poulsen, E. Pujos-Guillot, T.M. Larsen, A. Astrup, L.O. Dragsted, *J. Proteome Res.* **2014**, *13*, 1405.

[12] I. Garcia-Perez, J.M. Posma, E.S. Gibson Rm Chambres, T.H. Hansen, H. Vestergaard, T. Hansen, M. Beckmann, O. Pedersen, P. Elliott, J. Stamler, J.K. Nicholson, J. Draper, J.C. Mathers, E. Holmes, G. Frost, *Lancet Diabetes Endocrinol.* **2017**, *5*, 184.

[13] T. Esko, J.N. Hirschhorn, H.A. Feldman, Y.H. Hsu, A.A. Deik, C.B. Clish, C.B. Ebbeling, D.S. Ludwig, *Am. J. Clin. Nutr.* **2017**, *105*, 547.

[14] H. Gibbons, E. Carr, B.A. McNulty, A.P Nugent, J. Walton, A. Flynn, M.J. Gibney, L. Brennan, *Mol. Nutr. Food. Res.* **2017**, *61*, 1601050.

[15] V. Edefonti, R. De Vito, A. Salvatori, F. Bravi, L. Patel, M. Dalmartello, M. Ferraroni, *Adv. Nutr.* **2020**, *11*, 1255.

[16] V. Edefonti, R. De Vito, M. Dalmartello, L. Patel, A. Salvatori, M. Ferranoni, *Adv. Nutr.* **2020**, *11*, 293.

[17] K.D. Cashman, S. Muldowney, B. McNulty, A. Nugent, A.P. Fitzgerald, M. Kiely, J. Walton, M.J. Gibney, A. Flynn, *Br. J. Nutr.* **2013**, *109*, 1248.

[18] Y.M. Lenighan, A.P. Nugent, K.F. Li, L. Brennan, J. Walton, A. Flynn, H.M. Roche, B.A. McNulty, *Br. J. Nutr.* **2017**, *118*, 222.

[19] M.L. McCullough, D. Feskanich, M.J. Stampfer, E.L. Giovannucci, E.B. Rimm, F.B. Hu, D. Spiegelman, D.J. Hunter, G.A. Colditz, W.C. Willett, *Am. J. Clin. Nutr.* **2002**, *76*, 1261.

[20] T.T. Fung, M.L. McCullough, P.K. Newby, J.E. Manson, J.B. Meigs, N. Rifai, W.C. Willett, F.B. Hu, *Am. J. Clin. Nutr.* **2005**, *82*, 163.

[21] C.B. O'Donovan, M.C. Walsh, A.P. Nugent, B. McNulty, J. Walton, A. Flynn, M.J. Gibney, E.R. Gibney, L. Brennan, *Mol. Nutr. Food. Res.* **2015**, *59*, 377.

[22] Y. Zhao, F.J. Monahan, B.A. McNulty, M.J. Gibney, E.R. Gibney, *Br. J. Nutr.* **2014**, *112*, 1575.

[23] B.A. McNulty, A.P. Nugent, J. Walton, A. Flynn, C. Tlustos, M.J. Gibney, *Br. J. Nutr.* **2017**, *117*, 422.

[24] A.J. Moshfegh, D.G. Rhodes, D.J. Baer, T. Murayi, J.C. Clemens, W.V. Rumpler, D.R. Paul, R.S. Sebastian, K.J. Kuczynski, L.A. Ingwersen, R.C. Staples, L.E. Cleveland, *Am. J. Clin. Nutr.* **2008**, *88*, 324.

[25] M. Nelson, M. Atkinson, J. Meyer. A Photographic Atlas of Food Portion Sizes, Great Britain, Food Standards Agency, **2002**.

[26] H. Gibbons, C.J.R. Michielsen, M. Rundle, G. Frost, B.A. McNulty, A.P. Nugent, J. Walton, A. Flynn, M.J. Gibney, L. Brennan, *Mol. Nutr. Food Res.* **2017**, *61*, 1700037.

[27] A.P. Hearty, M.J. Gibney, *Br. J. Nutr.* **2009**, *101*, 598.

[28] L.H. Dekker, J.M. Boer, M.D. Stricker, W.B. Busschers, M.B. Snijder, M. Nicolaou, W.M. Verschuren, *J. Nutr.* **2013**, *143*, 1728.

[29] R. Villegas, A. Salim, M.M. Collins, A. Flynn, I.J. Perry, *Public Health Nutr.* **2004**, *7*, 1017.

[30] K.M. Lara, E.B. Levitan, O.M. Gutierrez, J.M. Shikany, M.M. Safford, S.E. Judd, R.S. Rosenson, *J. Am. Coll. Cardiol.* **2019**, *73*, 2036.

[31] A. Castelló, E. Boldo, B. Pérez-Gómez, V. Lope, J.M. Altzibar, V. Martín, G. Castaño-Vinyals, M. Guevara, T. Dierssen-Sotos, A. Tardón, V. Moreno, M. Puig-Vives, C. Llorens-Ivorra, J. Alguacil, I. Gómez-Acebo, J. Castilla, E. Gracia-Lavedán, V. Dávila-Batista, M. Kogevinas, N. Aragonés, P. Amiano, M. Pollán, *Maturitas* **2017**, *103*, 8.

[32] J.L. Fargnoli, T.T. Fung, D.M. Olenczuk, J.P. Chamberland, F.B. Hu, C.S. Mantzoros, *Am. J. Clin. Nutr.* **2008**, *88*, 1213.

[33] M.J. Walthouwer, A. Oenema, K. Soetens, L. Lechner, H. De Vries, *Appetite* **2014**, *82*, 154.

[34] J.A. Pryer, R. Nichols, P. Elliott, B. Thakrar, E. Brunner, M. Marmot, *J. Epidemiol. Community Health* **2001**, *55*, 29.

[35] L. Gonzalez-Guardia, E.M. Yubero-Serrano, J. Delgado-Lista, P. Perez-Martinez, A. Garcia-Rios, C. Marin, A. Camargo, N. Delgado-Casado, H.M. Roche, F. Perez-Jimenez, L. Brennan, J. López-Miranda, *J. Gerontol. A. Biol. Sci. Med. Sci.* **2015**, *70*, 78.

[36] G.L. Ambrosimi, P.M. Emmett, K. Northstone, S.A. Jebb, *Obesity (Silver Spring)* **2014**, *22*, 458.

[37] P.M. Emmett, L.R. Jones, *Nutr. Rev.* **2015**, *73*, 175.

[38] C.E. Gasser, J.A. Kerr, F.K. Mensah, M. Wake, *Br. J. Nutr.* **2017**, *117*, 1137.

[39] A.N. Funtikova, A.A. Benitez-Arciniega, M. Fito, H. Schroder, *Nutr. Res.* **2015**, *35*, 265.

[40] C.M. Berg, G. Lappas, E. Strandhagen, A. Wolk, K. Toren, A. Rosengren, N. Aires, D.S. Thelle, L. Lissner, *Am. J. Clin. Nutr.* **2008**, *88*, 289.

Author's Contributions:

OP conducted research, analysed data and wrote paper; JW, AF and APN provided essential materials and dietary data analysis from NANS; BAM provided essential materials and analysed data; LB designed and conducted research, and wrote paper. All authors read and approved the final version of the manuscript.

Acknowledgements:

The authors thank all the NANS and A-DIET Confirm participants for their commitment to both studies. LB acknowledges funding from The European Research Council ERC (647783).

Conflict of Interest:

All authors declare no conflicts of interest.

Tables:

Table 1. Demographics and Anthropometric Characteristics across the four clusters in the National Adult Nutrition Survey (n=563)

Characteristic	Cluster 1 (n=131)	Cluster 2 (n=112)	Cluster 3 (n=167)	Cluster 4 (n=153)	*P-value*
Age (years)	47.0 ± 17.2[a]	37.7 ± 14.4[b]	48.2 ± 14.9[a,c]	**52.1 ± 15.9[c]**	<0.001
Sex (% males)	53.4	58.9	49.7	42.5	0.055
Social Class (%PMT)	44.8	40.9	**58.1**	51.3	0.017
Smokers (%)	22.5	**26.8**	17.6	5.2	<0.001
Supplement users (%)	40.3	42.3	46.7	**60.1**	0.030
BMI (kg/m^2)	28.0 ±6.0	27.0 ± 4.9	27.5 ± 4.9	27.3 ± 4.7	0.491
Systolic Blood Pressure (mmHg)	126.9 ± 18.7	123.4 ± 9.4	126.7 ± 18.7	125.8 ± 18.6	0.462
Diastolic Blood Pressure (mmHg)	78.8 ± 10.6	77.9 ± 10.8	79.3 ± 11.2	78.7 ± 11.3	0.809

Data is mean ± SD, n; number of participants, %; percentage, %PMT; percentage of participants employed in a professional, managerial or technical capacity, mmHg; millimetres of mercury, *P*-value determined using a one-way ANOVA; *P*-value <0.05 was considered significant. Bold values describe the clusters with the highest demographic and anthropometric characteristics. Underlined values describe the clusters with the lowest demographic and anthropometric characteristics. [a,b,c] Mean values with unlike superscript letters are significantly different between clusters (P< 0·05). Cluster 1 = Moderately Unhealthy dietary pattern; Cluster 2 = Convenience dietary pattern; Cluster 3 = Moderately Healthy dietary pattern; Cluster 4 = Prudent dietary pattern.

Table 2. Diet quality scores and food group intake (%TE) across the four clusters in the National Adult Nutrition Survey (n=563)

Diet quality score/Food Group	Cluster 1 (n=131)	Cluster 2 (n=112)	Cluster 3 (n=167)	Cluster 4 (n=153)	*P-value*
Diet quality scores:					
AHEI	24.5 ± 9.1[a,b]	22.9 ± 8.8[b]	26.7 ± 10.5[a,c]	**28.4 ± 12.1[c]**	<0.001
AMDS	3.16 ± 1.49[a,b]	2.70 ± 1.53[b]	3.39 1.59[a,c]	**3.57 ± 1.66[c]**	<0.001
Food groups:					
Rice, Pasta, Flours and Starches	1.6 ± 2.8	2.2 ± 3.6	2.0 ± 2.9	2.2 ± 3.1	0.291
Savouries	2.7 ± 5.2	**4.3 ± 6.7**	2.7 ± 5.3	2.6 ± 5.1	0.042
White Bread, Rolls, Scones and Croissants	7.4 ± 7.0	7.3 ± 6.5	7.7 ± 7.1	7.5 ± 6.8	0.948
Wholemeal, Brown Bread and Rolls	**7.5 ± 6.5**	5.3 ± 6.3	7.4 ± 6.7	7.4 ± 6.4	0.026
Ready-to-eat Breakfast Cereals	3.8 ± 4.8	4.1 ± 4.5	4.1 ± 4.9	4.6 ± 5.6	0.612
Other Breakfast Cereals	1.7 ± 3.4	1.3 ± 3.4	1.9 ± 4.0	2.4 ± 4.2	0.100
Biscuits, Cakes and Pastries	5.6 ± 6.3	6.2 ± 7.1	6.2 ± 6.0	6.0 ± 5.9	0.862
Whole-milk	2.6 ± 4.2	2.6 ± 3.8	2.5 ± 3.7	2.6 ± 4.5	0.992
Low Fat and Skimmed Milks	2.4 ± 3.3[a,b]	2.1 ± 3.0[a,b]	2.1 ± 3.1[a]	**3.1 ± 3.8[b]**	0.024
Other Milk and Milk-based Beverages	0.2 ± 0.8	0.3 ± 1.1	0.8 ± 2.8	0.4 ± 2.1	0.095
Cream, Ice-creams, Rice Puddings, Custards & Desserts	1.9 ± 3.0	1.8 ± 2.6	2.1 ± 3.2	2.2 ± 3.4	0.794

Cheese	2.3 ± 3.0	2.5 ± 3.3	2.4 ± 2.9	2.1 ± 2.7	0.640
Yogurts	1.7 ± 2.3	1.2 ± 2.2	1.5 ± 2.4	1.8 ± 2.9	0.188
Eggs and Egg Dishes	1.5 ± 1.9	1.1 ± 1.7	1.5 ± 2.5	1.6 ± 2.5	0.179
Butter, Fat Spreads and Hard Cooking Fats	4.0 ± 4.7	3.0 ± 3.8	3.3 ± 5.1	3.2 ± 4.2	0.316
Low Fat Spreads and Oils	0.8 ± 1.6	0.9 ± 1.7	1.3 ± 2.4	1.1 ± 2.0	0.103
Potatoes	2.6 ± 2.7[a]	2.7 ± 2.8[a,b]	3.1 ± 3.2[a,b]	**3.7 ± 3.0[b]**	0.012
Chips and Processed Potatoes	4.7 ± 4.9[a]	**5.5** ± 5.5[a]	4.2 ± 5.1[a]	2.7 ± 3.0[b]	<0.001
Vegetables and Vegetable Dishes	3.3 ± 2.5[a,c]	2.7 ± 2.2[a]	**4.3 ± 3.6[b]**	3.8 ± 3.1[b,c]	<0.001
Fruit Juices and Smoothies	1.0 ± 1.5	1.0 ± 1.8	0.9 ± 1.7	1.2 ± 1.9	0.224
Fruit	3.3 ± 3.8[a]	2.0 ± 2.9[b]	**3.5 ± 3.6[a]**	3.4 ± 3.2[a]	0.002
Savoury Snacks	1.9 ± 3.4	2.9 ± 4.8	1.7 ± 2.9	2.2 ± 4.4	0.090
Fish, Fish Dishes and Products	2.2 ± 3.3[a,b]	2.0 ± 3.1[a]	**3.3 ± 4.3[b]**	3.1 ± 3.6[a,b]	0.004
Unprocessed White Meat	3.4 ± 3.9	3.7 ± 4.3	3.9 ± 4.4	3.7 ± 4.9	0.827
Processed White Meat	**2.1 ± 4.6[a]**	1.7 ± 2.7[a,b]	1.1 ± 2.4[b]	1.4 ± 3.2[a,b]	0.050
Unprocessed Red Meat	7.4 ± 6.2	7.5 ± 6.7	7.4 ± 6.2	7.2 ± 6.2	0.982
Processed Red Meat	5.6 ± 5.9[a,c]	**6.2 ± 6.2[a]**	4.0 ± 4.3[b]	4.0 ± 4.7[b,c]	<0.001
Alcoholic Beverages	6.1 ± 8.1	5.4 ± 7.9	5.3 ± 6.7	4.7 ± 5.9	0.463
Sugars, Syrups, Preserves and Sweeteners	1.7 ± 2.4	1.4 ± 2,3	1.7 ± 2.4	1.9 ± 2.7	0.455

Confectionary	2.6 ± 3.6	3.3 ± 4.6	2.2 ± 3.4	2.5 ± 3,8	0.113
Soups, Sauces and Condiments	2.9 ± 3.3	2.8 ± 3.1	2.3 ± 2.3	2.3 ± 2.6	0.180
Low Energy Beverages	0.2 ± 0.5	0.1 ± 0.4	0.2 ± .06	0.1 ± 0.3	0.317
High Energy Beverages	1.4 ± 2.7[a]	**2.6 ± 3.5[b]**	1.1 ± 2.3[a]	1.0 ± 2.3[a]	<0.001

Data is mean ± SD of percentage energy contribution to total energy intakes (%TE), n; number of participants, AHEI; Alternative Healthy Eating Index, AMDS; Alternative Mediterranean Dietary Score, *P*-value determined using a one-way ANOVA; *P*-value <0.05 was considered significant. Bold values describe the clusters with the highest energy contribution to significant food groups. Underlined values describe the clusters with the lowest energy contribution to significant food groups. [a,b,c] Mean values with unlike superscript letters are significantly different between clusters (P <0·05). Cluster 1 = Moderately Unhealthy dietary pattern; Cluster 2 = Convenience dietary pattern; Cluster 3 = Moderately Healthy dietary pattern; Cluster 4 = Prudent dietary pattern.

Table 3. Nutrient intake across the four clusters in the National Adult Nutrition Survey (n=563)

Nutrient	Cluster 1 (n=131)	Cluster 2 (n=112)	Cluster 3 (n=167)	Cluster 4 (n=153)	*P-value*
Energy (kcal)	2023 ± 616	2137 ± 643	2047 ± 611	1942 ± 591	0.086
Macronutrients					
Protein (%TE)	16.6 ± 3.2[a]	16.8 ± 4.4[a,b]	17.3 ± 3.4[a,b]	**17.9 ± 3.7[b]**	0.013
Carbohydrates (%TE)	45.2 ± 7.2	45.6 ± 7.3	45.7 ± 6.4	46.4 ± 7.2	0.557
Total Sugars (%TE)	18.1 ± 5.9	17.8 ± 6.0	17.9 ± 5.7	18.9 ± 5.8	0.422
Total Fat (%TE)	34.4 ± 6.4	33.8 ± 6.2	34.1 ± 7.3	33.0 ± 5.7	0.283
Saturated Fat (%TE)	13.4 ± 3.5	13.3 ± 3.1	13.3 ± 4.4	12.9 ± 3.3	0.648
Monounsaturated Fat (%TE)	12.6 ± 2.9	12.3 ± 2.5	12.2 ± 2.7	11.8 ± 2.6	0.085
Polyunsaturated Fat (%TE)	6.6 ± 2.4	6.1 ± 2.6	6.5 ± 2.7	6.0 ± 2.1	0.378
Dietary Fibre (g/day)	19.4 ± 7.5	18.8 ± 8.7	21.2 ± 8.5	20.0 ± 7.4	0.069
Micronutrients					
Vitamin A (µg)	1079 ± 695[a,b]	985 ± 909[a]	1192 ± 856[a,b]	**1351 ± 1125[b]**	0.008
Vitamin C (mg)	125 ± 190	113 ± 170	127 ± 195	154 ± 248	0.393
Vitamin D (µg)	3.82 ± 3.47	5.53 ± 13.32	4.90 ± 4.69	6.00 ± 5.62	0.077
Vitamin E (mg)	12.4 ± 19.6	11.5 ± 10.0	16.6 ± 45.3	17.7 ± 44.8	0.396
Vitamin B6 (mg)	2.98± 1.86[a]	3.71 ± 5.08[a,b]	**5.29 ± 11.76[b]**	3.91 ± 4.70[a,b]	0.047

Vitamin B12 (µg)	6.58 ± 22.09	5.95 ± 7.09	7.22 ± 11.73	6.33 ± 6.72	0.877
Thiamin (mg)	2.02 ± 2.06	2.49 ± 4.98	4.14 ± 11.57	3.24 ± 6.23	0.083
Riboflavin (mg)	2.22 ± 1.86	2.65 ± 4.90	4.35 ± 12.10	3.18 ± 4.62	0.074
Total Folate (µg)	428 ± 635	353 ± 185	376 ± 207	395 ± 283	0.436
Total Niacin (mg)	28.7 ± 34.8	27.7 ± 13.6	29.0 ± 18.1	30.4 ± 25.5	0.848
Sodium (mg)	2512 ± 820[a,b]	**2777 ± 958[a]**	2555 ± 900[a,b]	2389 ± 765[b]	0.004
Potassium (mg)	3096 ± 980	3033 ± 1014	3215 ± 928	3194 ± 1214	0.444
Calcium (mg)	899 ± 350	931 ± 338	941 ± 369	1018 ± 521	0.084
Magnesium (mg)	293 ± 113	287 ± 109	307 ± 107	310 ± 121	0.296
Phosphorus (mg)	1384 ± 447	1401 ± 450	1432 ± 437	1409 ± 432	0.819
Iron (mg)	14.56 ± 13.45	14.92 ± 16.44	13.17 ± 5.13	16.02 ± 18.00	0.329
Zinc (mg)	10.82 ± 6.69	10.99 ± 5.08	11.62 ±9.43	11.75 ± 7.12	0.662
Iodine (µg)	149 ± 76	144 ± 72	154 ± 85	**170 ± 91**	0.046

Data is mean ± SD of nutrient intakes, n; number of participants, kcal; kilocalories, %TE; percentage of total energy, g/day; grams per day, µg; micrograms, mg; milligrams, *P*-value determined using a one-way ANOVA; *P*-value <0.05 was considered significant. Bold values describe the clusters with the highest intake of significant nutrients. Underlined values describe the clusters with the lowest intake of significant nutrients. [a,b,c] Mean values with unlike superscript letters are significantly different between clusters (P< 0·05). Cluster 1 = Moderately Unhealthy dietary pattern; Cluster 2 = Convenience dietary pattern; Cluster 3 = Moderately Healthy dietary pattern; Cluster 4 = Prudent dietary pattern.

Table 4. Biochemical and Nutrient Status across the four clusters in the National Adult Nutrition Survey (n=563)

Biomarker	Cluster 1 (n=131)	Cluster 2 (n=112)	Cluster 3 (n=167)	Cluster 4 (n=153)	*P-value*
Serum Glucose (mmol/l)	5.3 ± 0.9	5.3 ± 0.7	5.3 ± 0.8	5.3 ± 1.0	0.910
Serum Calcium (mmol/l)	2.4 ± 0.1	2.4 ± 0.2	2.5 ± 0.2	2.4 ± 0.2	0.254
TIBC (mmol/l)	61.7 ± 9.0	60.9 ± 8.9	60.0 ± 7.9	59.9 ± 8.1	0.236
Serum Ferritin (ng/ml)	145 ± 148	136 ± 126	113 ± 82	119 ± 115	0.082
Haemoglobin (g/dl)	14.2 ± 1.4[a,b]	**14.5 ± 1.5[a]**	14.2 ± 1.4[a,b]	14.0 ± 1.3[b]	0.037
Serum Folate (mmol/l)	33.8 ± 37.9[a,b]	28.5 ± 17.4[a]	31.0 ± 18.9[a,b]	**38.4 ± 25.0[b]**	0.012
Red Cell Folate (nmol/l)	1004 ± 492	985 ± 436	972 ± 396	1096 ± 492	0.079
Serum Vitamin B12 (pmol/l)	313 ± 204	347 ± 293	327 ± 171	328 ± 169	0.663
EGRac	1.35 ± 0.16[a,b]	**1.39 ± 0.17[a]**	1.37 ± 0.17[a,b]	1.33 ± 0.15[b]	0.026
PLP (nmol/l)	85.7 ± 68.2	111.1 ± 108.0	108.6 ± 92.3	107.7 ± 95.8	0.091
Serum 25-Hydroxyvitamin D (nmol/l)	59.0 ± 21.6	61.9 ± 25.8	59.9 ± 23.7	65.7 ± 27.0	0.088
Osteocalcin (ng/ml)	11.0 ± 3.7[a,b]	**11.6 ± 3.6[a]**	10.2 ± 3.2[b]	11.1 ± 3.6[a,b]	0.005
Cross-linked C-telopeptide (mg/ml)	0.40 ± 0.22[a]	**0.52 ± 0.27[b]**	0.40 ± 0.21[a]	0.39 ± 0.20[a]	<0.001
α-Tocopherol mmol/ml	26.3 ± 7.4	25.3 ± 7.7	27.4 ± 7.4	26.9 ± 6.9	0.114
γ-Tocopherol mmol/ml	1.81 ± 0.90[a,b]	**1.94 ± 0.90[a]**	1.79 ± 0.98[a,b]	1.59 ± 0.83[b]	0.017
Urinary Sodium (mmol/l)	95.9 ± 45.1[a,b]	**105.7 ± 45.0[a]**	96.5 ± 38.3[a,b]	90.1 ± 39.1[b]	0.030

Formatted: Polish

Urinary Potassium (mmol/l)	41.6 ± 27.2	46.6 ± 27.2	50.1 ± 28.5	47.3 ± 26.5	0.069
Urinary Iodine (ng/ml)	128.0 ± 96.0	114.8 ± 69.7	129.5 ± 102.7	127.0 ± 80.1	0.763
Urinary Creatinine (umol/l)	10877 ± 6512[a]	**13753 ± 5429[b]**	10161 ± 5457[a]	9913 ± 4880[a]	<0.001

Data is mean ± SD, n; number of participants, TIBC; total iron binding capacity, EGRac; erythrocyte glutathione reductase activity coefficient, PLP; pyridoxal-5'-phosphate, *P*-value determined using a one-way ANOVA; *P*-value <0.05 was considered significant. Bold values describe the clusters with the highest intake of significant nutrients. Underlined values describe the clusters with the lowest intake of significant nutrients. [a,b,c] Mean values with unlike superscript letters are significantly different between clusters (P< 0·05). Cluster 1 = Moderately Unhealthy dietary pattern; Cluster 2 = Convenience dietary pattern; Cluster 3 = Moderately Healthy dietary pattern; Cluster 4 = Prudent dietary pattern.

Table 5. Stability of A-DIET Confirm participants cluster membership across four visits (n=175)

Cluster	Visit 1 Number of Participants	Visit 2	Visit 3	Visit 4	Average
1	30	33.3%	29.6%	25.0%	29.3%
2	52	61.5%	41.7%	34.1%	45.8%
3	43	27.9%	34.1%	34.1%	32.1%
4	52	60.0%	50.0%	45.7%	51.9%
Kappa		0.300*	0.192*	0.144*	
Gamma		0.381*	0.319*	0.293*	
Kendall's tau-b		0.298*	0.240*	0.222*	

Number of Participants refers to the number classified in a cluster a visit 1. Percentage refers to the percentage of participants from visit 1 that stay within cluster 1, 2, 3 or 4 at visit 2, visit 3, visit 4. Average refers to the average percentage of stability across visit 2, visit 3, and visit 4. Cluster 1 = Moderately Unhealthy dietary pattern; Cluster 2 = Convenience dietary pattern; Cluster 3 = Moderately Healthy dietary pattern; Cluster 4 = Prudent dietary pattern. Kappa; Cohen's kappa statistic, Gamma; Goodman and Kruskal's gamma, Kendall's tau-b; Kendall's tau-b correlation coefficient. Kappa, Gamma and Kendall's tau-b refer to the agreement between cluster membership at visit 1 and at visit 2, 3 and 4. * denotes statistical significance (P-value < 0.05).

Figures

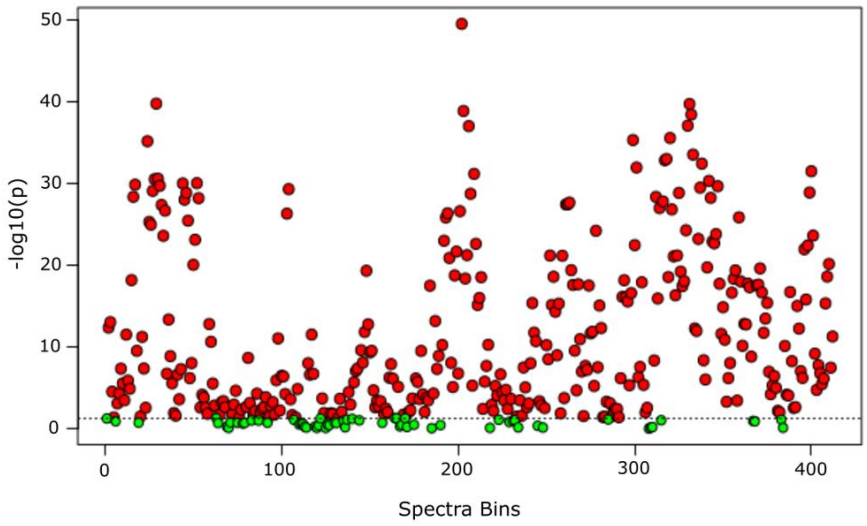

Figure 1. Plot of the log *p*-values obtained in a one-way ANOVA showing significant (red, n=351) and non-significant (green, n=199) NMR spectral bins (features) across the four clusters. The plot depicts the regions of interest in the metabolomic data that are key to defining the multivariate model. NMR analysis was performed on urine samples from the National Adult Nutrition Survey

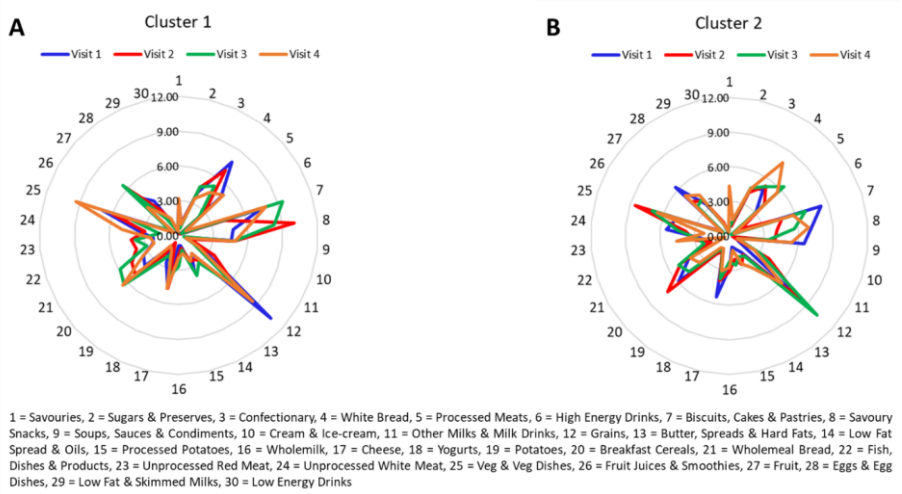

Figure 2. Reproducibility of the metabolomic-based dietary patterns in A-DIET Confirm across four visits. Radar plots are of the food group percentage energy contribution across the four visits for A cluster 1; B cluster 2. Cluster 1 = Moderately Unhealthy dietary pattern; Cluster 2 = Convenience dietary pattern.

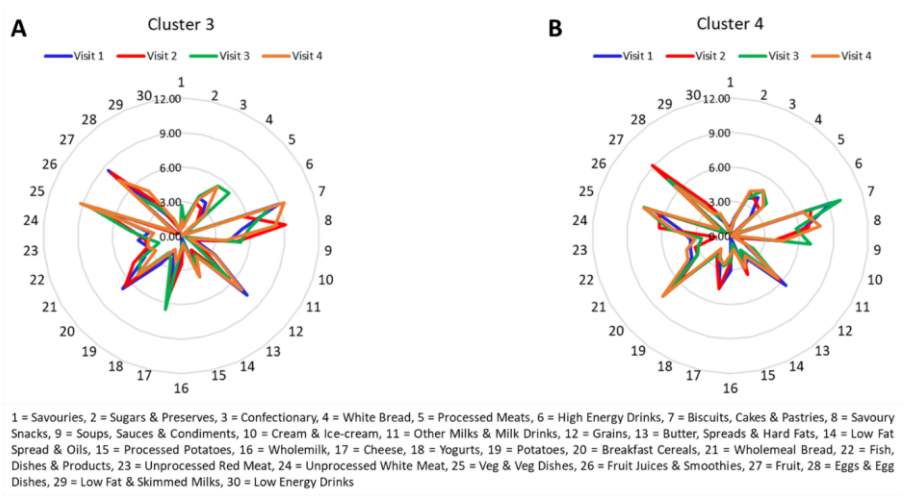

Figure 3. Reproducibility of the metabolomic-based dietary patterns in A-DIET Confirm across four visits. Radar plots are of the food group percentage energy contribution across the four visits for A cluster 3; B cluster 4. Cluster 3 = Moderately Healthy dietary pattern; Cluster 4 = Prudent dietary pattern.