



---

# Identifying Food Consumption Patterns among Young Consumers by Unsupervised and Supervised Multivariate Data Analysis

Ulf Hammerling<sup>1\*</sup>, Eva Freyhult<sup>2</sup>, Anna Edberg<sup>1,3</sup>, Salomon Sand<sup>4</sup>,  
Sisse Fagt<sup>5</sup>, Vibeke Kildegaard Knudsen<sup>5</sup>, Lene Frost Andersen<sup>6</sup>,  
Anna Karin Lindroos<sup>4</sup>, Daniel Soeria-Atmadja<sup>1,7</sup>  
and Mats G. Gustafsson<sup>1</sup>

<sup>1</sup>*Cancer Pharmacology and Computational Medicine, Department of Medical Sciences, Uppsala University, Uppsala Academic Hospital, SE-75185 Uppsala, Sweden.*

<sup>2</sup>*Cancer Pharmacology and Computational Medicine, Department of Medical Sciences, Bioinformatics Infrastructure for Life Sciences, Science for Life Laboratory, Uppsala University, SE-75185 Uppsala, Sweden.*

<sup>3</sup>*Råd and Rön, P.O. Box 38001, SE-10064 Stockholm, Sweden.*

<sup>4</sup>*National Food Agency, SE-75126 Uppsala, Sweden.*

<sup>5</sup>*Department of Nutrition, National Food Institute, Technical University of Denmark, DK-2860 Søborg, Denmark.*

<sup>6</sup>*Department of Nutrition, University of Oslo, NO-0316, Norway.*

<sup>7</sup>*Reveal, P.O. Box 22500, SE-10422 Stockholm, Sweden.*

## **Author's contributions**

*This work was carried out in collaboration between all authors. Authors UH, DSA and MGG designed the study. Authors AE, EF and DSA conducted the multivariate data analyses. Authors UH, AE, DSA, EF and MGG wrote the paper. Authors SS, SF, VKK, LFA and AKL suggested study refinements and reviewed manuscript drafts. All authors read and approved the final manuscript.*

**Short Communication**

**Received 17<sup>th</sup> January 2014**

**Accepted 31<sup>st</sup> May 2014**

**Published 7th July 2014**

---

## ABSTRACT

Although computational multivariate data analysis (MDA) already has been employed in the dietary survey area, the results reported are based mainly on classical exploratory (descriptive) techniques. Therefore, data of a Swedish and a Danish dietary survey on young consumers (4 to 5 years of age) were subjected not only to modern exploratory MDA, but also modern predictive MDA that via supervised learning yielded predictive classification models. The exploratory part, also encompassing Swedish 8 or 11-year old Swedish consumers, included new innovative forms of hierarchical clustering and bi-clustering. This resulted in several interesting multi-dimensional dietary patterns (dietary prototypes), including striking difference between those of the age-matched Danish and Swedish children. The predictive MDA disclosed additional multi-dimensional food consumption relationships. For instance, the consumption patterns associated with each of several key foods like bread, milk, potato and sweetened beverages, were found to differ markedly between the Danish and Swedish consumers. In conclusion, the joint application of modern descriptive and predictive MDA to dietary surveys may enable new levels of diet quality evaluation and perhaps also prototype-based toxicology risk assessment.

*Keywords: Dietary surveys; young consumers; unsupervised MDA; supervised MDA; dietary prototypes; dietary patterns.*

## 1. INTRODUCTION

Surveys of food consumption are undertaken on a regular basis in many developed countries. The overall purpose of these investigations is to gain insight in food habits among the general national populations and outstanding segments thereof, e.g. women, men, children, elderly as well as socio-demographic or regional consumer categories. Dietary surveys are commonly designed as either a 24-h recall or a dietary record scheme, as proposed by the European Food Safety Authority (EFSA) [1]. Traditionally, statistical analysis of food consumption data has been focused on intake levels of single nutrients or foods at a time, i.e. using classical univariate statistical analyses. This approach is, however, not quite appropriate to disclose consumption patterns across several dimensions, i.e. poorly adapted to the fact that people eat a variety of foods rather than single nutrients. Thus, the last decade has witnessed a gradual transition in research from a nutrient to a food level perspective, translating to an emphasis on dietary patterns [2-5]. Typically, patterns of food intake are discovered by means of exploratory dimension-reducing techniques, which fall into two separate categories: Cluster Analysis, on the one hand, and Factor Analysis - commonly appearing as Principal Component Analysis (PCA) - on the other. Cluster analysis is, however, not widely reported in dietary surveys of young consumers. Nonetheless, a few papers elaborating on this sort of interrogation have appeared in the literature [6-8]. Another set of methods, collectively referred to as Mixture Models (MMs), have lately also found usage in explorative nutritional epidemiology. MMs can help identifying latent sub-populations within an overall population without a need for data to directly recognise the underlying categories or to which an entity belongs. Certain variants of MMs can therefore be perceived as clustering procedures. Intricate dietary patterns within a large fraction of the EPIC (The European Prospective Investigation into Cancer and Nutrition) cohort, by means of a variant MM, have recently been reported [9]. Moreover, Latent Class Analysis (LCA) - closely related to MMs - identifies subgroups within a population. Constructs are

subsequently used for regression analysis [10,11]. By means of LCA dietary prototypes have been identified in cohorts of modest and large sizes [12,13].

The above-mentioned MDA techniques are, however, all confined to the descriptive modelling category of data analysis, also known as unsupervised multivariate analysis. Although this sort of modelling is indeed helpful to the disentanglement of embedded associations within dietary surveys, the class of MDA techniques known as predictive modelling can enable the unveiling of even deeper entrenched patterns in multi-dimensional data sets [14-16]. Prediction models based on supervised learning are built on observed examples that include a response variable, which may incorporate either a discrete or a continuous value. Some well established examples of such algorithmic families are Artificial Neural Networks (ANN) [17], Decisions Trees (DT) [18,19], Random Forests (RF) [20,21] and Nearest Shrunken Centroids (NSC) [22-25]. Using ANNs and Decision Trees, Hearty and Gibney modelled an Irish food consumption survey to predict each individual's Health Eating Index, in anyone of five separate categories, with reasonably high accuracy [26]. Nutritional epidemiology, however, let alone dietary surveys, rarely employs predictive modelling.

With a view to this overall background, we applied modern and novel innovative MDA methods to disclose pertinent dietary patterns in data on pre-adolescent consumers and compared them between two countries. Therefore, data used in this study were compiled from two national food consumption surveys of young consumers, one in Sweden and the other in Denmark. The former data set incorporated three age classes, one preschool and two elementary school children, whereas the latter included preschool children only. The material was subjected to various statistical techniques, including two sorts of cluster (unsupervised) analysis as well as two varieties of supervised analysis, RF and NSC, but only results of the latter supervised technique are shown in the present paper. Recently, this work - financed by the Nordic Council of Ministers, Nordic Working Group for Diet, Food and Toxicology (NKMT) - appeared in the form of a TemaNord report [27]. Below, pertinent findings of this study are outlined.

## **2. DATA SETS AND METHODOLOGY**

### **2.1 Data Sets**

The analyses were based on two dietary survey data sets collected in Denmark and Sweden: (i) A fraction of data encapsulating preschool consumers (children of 4-5 years of age, in total 318 subjects), compiled in 2000 through 2008 of *the Danish National Survey of Diet and Physical Activity*, an ongoing project at the National Food Institute, Technical University of Denmark. (ii) Data collected in *Riksmaten - barn 2003*, which is an open and estimated food diary across four consecutive days, conducted by the (Swedish) National Food Agency. It encapsulates 2496 children of three age classes: 4, 8 and 11 years of age, with 590 individuals of preschool (4 years) age [28].

### **2.2 Assignment of Foods into Groups**

The Swedish and Danish consumption data were aggregated into 35 and 25 separate food groups, respectively. Prior to Danish-Swedish comparison, data from a Swedish age-matched excerpt were harmonised with those of Denmark, i.e. transformed to 25 food

categories. Mean intakes of each food group were input variables to the various MDA settings. Below, accordingly defined food groups appear in italic font.

## 2.3 Data Pre-processing

Intake levels were consistently expressed as weight percent, i.e. the consumption of each food group was divided by the total intake of all food groups. Certain analyses were preceded by an additional normalisation-type transformation, either Z-score normalisation (all food group values will have zero mean and unit variance) or by dividing each food group value with its arithmetic mean value.

## 2.4 MDA

Descriptive/exploratory modelling (unsupervised MDA) and predictive modelling (supervised MDA) were applied to the data sets. A significant part of the descriptive analyses was conducted using a top-down multi-branching hierarchical clustering algorithm developed in-house (OMB-DHC) [29,30]. Each accordingly identified major cluster was represented by its centroid (average), which subsequently was visualized together all the others using bi-clustering (yielding a “heatmap”) and multi-dimensional Scaling [31]. Finally, the Danish and Swedish age-matched data sets (the original data and not the prototypes) were each subjected to variable selection and training of the Nearest Shrunken Centroid (NSC) classifier, thereby enabling multivariate pattern recognition. The classifiers were built to discriminate between low (below the 20<sup>th</sup> percentile) and high (above the 80<sup>th</sup> percentile) consumers of each food category studied, based on their dietary patterns (excluding the food category considered). The MDA was performed in a repeated holdout procedure using 80% of the data to train the model and the remaining 20% to test model performance. The holdout procedure was repeated 100 times and an average performance over all repeats was computed. The NSC modelling included a variable selection step, which yielded counts of how many times each variable (food type) was selected. Null distribution for performance measures and variable selection counts (for each variable) was determined by means of 200 permutations.

## 2.5 Computational Environments

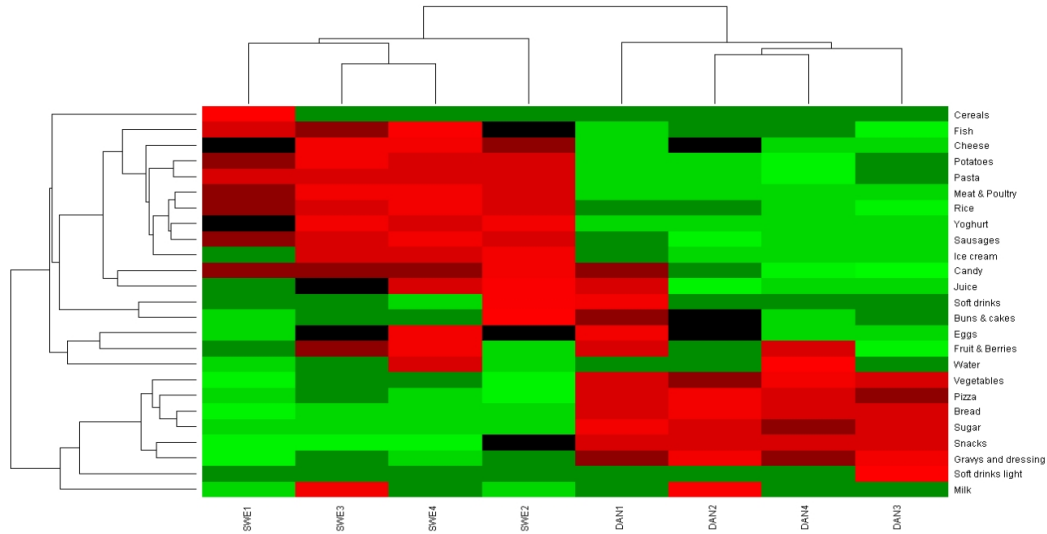
Data pre-processing and unsupervised MDA were conducted in the Matlab development suite (Mathworks, Inc, Natick, USA), whereas the supervised MDA analysis was conducted in R [32]. The NSC algorithm used was available in the R package *pamr* [22].

## 3. RESULTS AND DISCUSSION

### 3.1 Descriptive Modelling by Means of Cluster Analysis

The multi-branching hierarchical clustering (OMB-DHC) output allows for the multi-tier analysis of clusters, appearing at gradually lower segregation levels. Upon analysing the age-matched trans-national data sets with OMB-DHC, four groups of each of the Swedish and Danish dataset of preschool children emerged as the most appropriate representations of the respective data sets. These clusters appeared at the highest branching level using the Danish data set, but emerged at several hierarchical levels in the corresponding Swedish data (Appendix Fig. 1). The term dietary prototype was assigned to the centroids (mean) of the accordingly identified groups. Two such prototypes from each nation, named by

outstanding food groups, showed some mutual resemblance: *Milk* and *Soft beverages (sweetened)/Buns & cakes* in the Swedish data set, on the one hand, and *Milk* and *Soft beverages (sweetened)/Juice* among the Danish children, on the other. Nonetheless, subsequent (hierarchical) bi-clustering of the accordingly derived dietary prototypes revealed a striking discrepancy between age-matched Danish and Swedish children's consumption patterns. The resulting "heatmap" discloses an approximately inverse relationship between the two countries (Fig. 1). Another downstream multidimensional analysis of the prototypes unveiled an appreciable data distribution within each national population, but showed nonetheless much larger distance across the nations (Appendix Fig. 2).



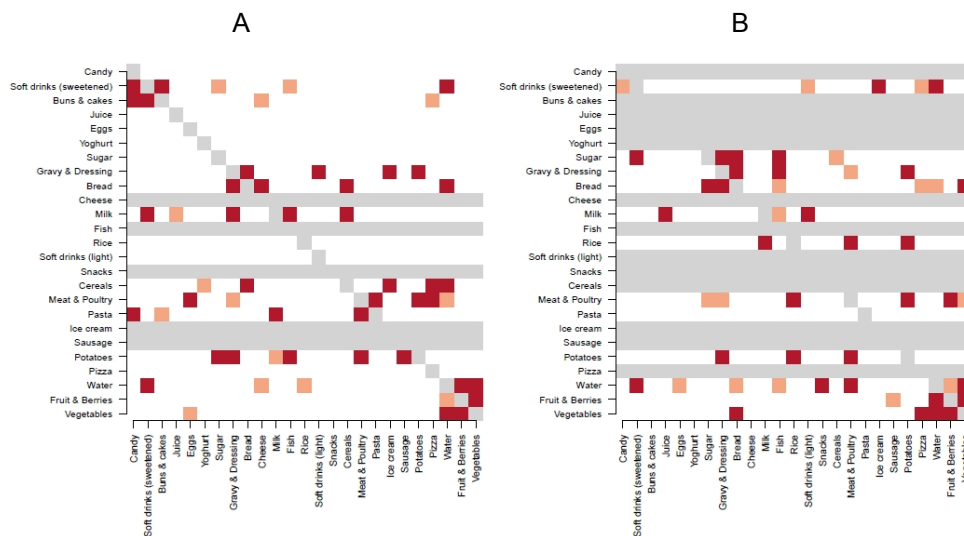
**Fig. 1. A second level exploratory MDA using 8 dietary prototype profiles obtained in a previous clustering analysis (OMB-DHC) step. Four prototypes were obtained from Swedish and four from Danish age-matched preschool consumer data. The panel displays results from a conventional bi-clustering procedure. Colours indicate high (red), intermediate (black) or low (green) consumption levels. The panel was reproduced with permit from the Nordic Council of Ministers**

A multidimensional scaling analysis of dietary prototypes, as derived from the entire Swedish data set, identified one divergent prototype in each of the two older age groups (8 and 11 years old consumers, respectively). The two outstanding prototypes, each encapsulating about 20 % of the respective elementary school age population, exhibit obvious mutual similarity and feature low-fat foods - largely with respect to dairy products - and otherwise generally prudent eating profiles. Moreover, each of the Swedish age group housed another and fairly well defined dietary prototype. These prototypes (encompassing 20 % of preschool children and nearly 25 % among each age-class of elementary school consumers in the data set) featured relatively low intake of *Vegetables* and *Fruit and berries* as well as rather high consumption of sugar-sweetened soft drinks. A dietetic prototype of the latter was identified also in the Danish data material (representing about 27 % of the children in the national data set), but devoid of association with low consumption of *Vegetables* or *Fruit and berries* (not shown).

### 3.2 Predictive Modelling Using Classifiers

Using predictive modelling, data of age-matched preschool Danish and Swedish children were scrutinized for food preferences attached to each food group. This undertaking was

based on NSC classification, using low and high consumers (i.e. the 20% fractions of subjects with lowest and highest consumption) for each food category in the absence of the particular food. For example, a *Bread* classification model was built to classify High/Low consumers of *Bread*, based on the consumption pattern across the remaining foods. This procedure was conducted to unveil associations with the respective foods, not readily disclosed by any exploratory methodology. Models were also built in a similar fashion by means of RF, providing results that concurred with those derived from NSC analyses (data not shown). NSC was used in a repeated holdout setting and permutations were conducted to assess results quality [27]. The eating behaviour attached to several food groups showed marked disparity across the national border (Fig. 2). A few observations from the NSC-modelling exercises are as follows: Models for *Potato* depend heavily on *Rice* and *Meat & chicken* among Danish children, but are rather linked to *Sausage* and *Fish* in Sweden. The Danish *Bread* models depend strongly on the consumption of *Sugar* (i.e. marmalade) and *Vegetables*, whereas the Swedish *Bread* models depend mainly on *Cheese*, *Spreads & Dressing* and *Cereals* for consumers of this age group (Fig. 3). Moreover, predictive classification of *Milk* showed associations with (fruit) *Juice* and *Soft beverages (without sugar)* among the Danish children, but were rather associated with *Cereals*, *Fish* and *Soft beverages (sweetened)* in the Swedish data set. Large cross-national difference in terms of food group associations with *Meat & chicken* was also observed (not shown).



**Fig. 2. Significant variables identified using NSC predictive modelling of the Swedish (A) and Danish (B) dietary survey data sets on preschool children. The NSC technique applied here attempts, irrespectively of the targeted food (one in each model), to reliably classify the respective (low/high consumption) populations. Thus, each model was built to predict class Y based on consumption pattern of the remaining food groups X. Notably, NSC performs simultaneous variable selection and model construction. For each food group data was partitioned in training (80 %) and test (20 %) sets. The entire sequence of model-building, followed by testing, was iterated 100 times, ultimately generating a selection frequency of each variable, as appearing in panels A or B, above. Colour-coded squares indicate significance at either of two separate levels (dark red: 0.99 level; light red: 0.95 level; white: non-significant), whereas consistent non-significant models appear as grey bars. Panels reproduced with permit from the Nordic Council of Ministers**

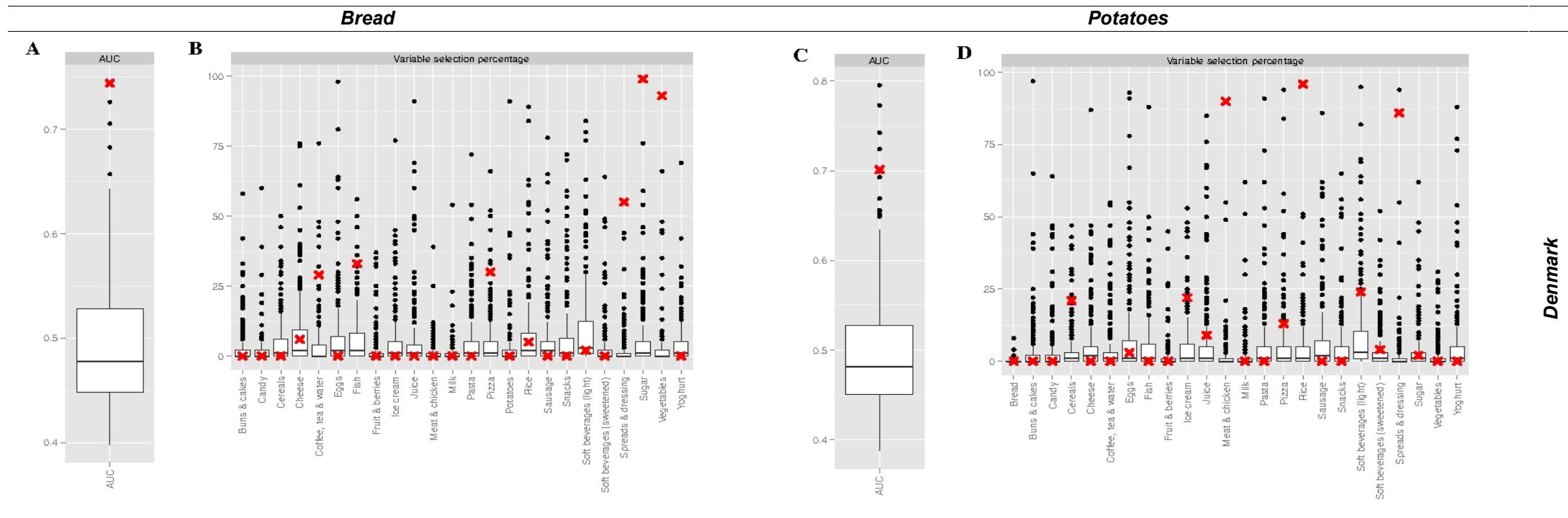
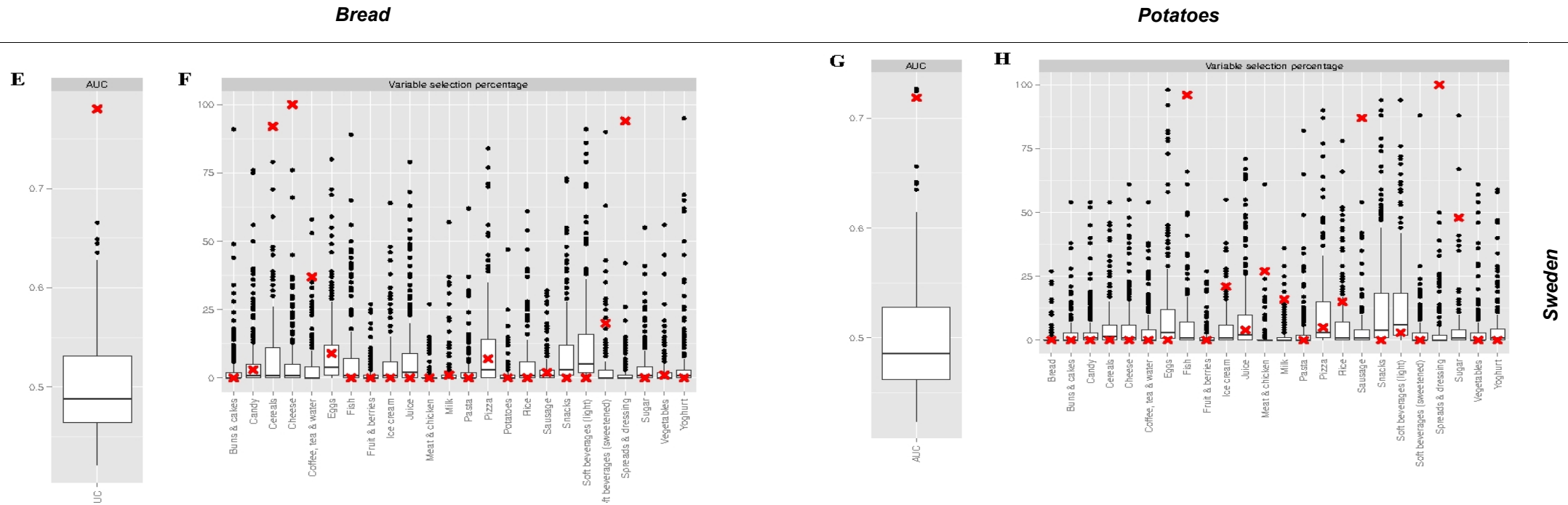


Fig. 3. Continued in Next Page.....



**Fig. 3. Graphical display of results derived from Nearest Shrunken Centroid (NSC) modelling for each of two selected pairs of Danish-Swedish consumption patterns, as attached to either of the Bread or Potato food groups. Performance is expressed as AUC (A, C, E, G) and variable selection frequency appears as percent (B, D, F, H), for Danish and Swedish Bread (A, B and E, F, respectively) as well as Danish and Swedish Potato (C, D and G, H, respectively) models. Null distribution, as determined by 200 permutations, is shown as a boxplot for each AUC and variable (food group). The original (un-permuted) model results are indicated by red X marks. Panels A, C, E, and G show that the performance estimates of the four displayed models are significantly higher than expected by chance (at the 0.05 significance level). Panel B clearly shows that Sugar, and Vegetables are the most important variables in the Danish Bread model, whereas panel F shows that Cereals, Cheese and Spreads & dressing are most essential to the Swedish Bread model. As for the Potato models, panel D highlights Meat & chicken and Rice as most important to the Danish Potato model, whereas panel H indicate that Fish, Sausage and Spreads & dressing are fundamental to the Swedish Potato model**



#### 4. CONCLUSION

Embedded multi-dimensional dietary patterns in data from two Nordic dietary surveys of young consumers have been disclosed through joint application of modern descriptive and predictive MDA modelling [27]. The descriptive modelling involved cluster analysis and multi-dimensional scaling, which were used to identify interesting consumer prototypes in the dietary data sets. More intricate patterns were unveiled through a second round of exploratory data analysis based on the consumer prototypes identified in the first step.

In the current analysis, two well-known algorithms for supervised MDA (random forest and nearest shrunken centroid) were employed to identify eating preferences tied to each of a panel of food groups. Several such dietary associations were found to be strikingly dissimilar between Danish and Swedish young consumers. For instance, *Potato, Bread, Milk and Meat and poultry* emerged as connected to quite distinct food groups among the respective national preschool consumers, as disclosed by supervised data interrogation [27].

Outstanding observations made in this study include: i) Two prudent and mutually similar dietary prototypes, located remotely from all counterparts in each age group, appeared among elementary school children in the Swedish material. ii) Seemingly, Danish and Swedish children (4-5 years) consumption patterns are strikingly different, in part owing to many individual foods being linked to different eating habits and iii) dietary prototypes with significant elements of *Soft beverages (sweetened)* appeared in the Danish data on preschool children and in each age class of the Swedish material.

Findings summarized in the present paper, outlined in more detail in a recently published TemaNord report, show that the joint application of clustering techniques and supervised (machine) learning techniques provide a promising framework for deepened understanding of intricate eating habit patterns [27]. Rather than relying solely on assessments based mainly on independent uni-variate distributions, reflecting the consumption of particular food categories across the whole population (e.g., WHO/IPCS, 2009) [33], the consumer prototypes (and/or associated multivariate correlation patterns) should also be taken into account. A consumer prototype view, which allows for e.g. prototype specific exposure assessment, seemingly enables more adequate analysis of the explicit impact of differences in diet on the intake of nutritional or hazardous substances, thus potentially allowing for more accurate delineation of potential risk groups. Lastly, predictive modelling enables exhaustive characterisation of eating habits, thereby providing a powerful technique in dietary pattern analysis.

#### ACKNOWLEDGEMENTS

The study was funded by the Nordic Council of Ministers, Nordic Working Group for Diet, Food and Toxicology (NKMT).

#### COMPETING INTERESTS

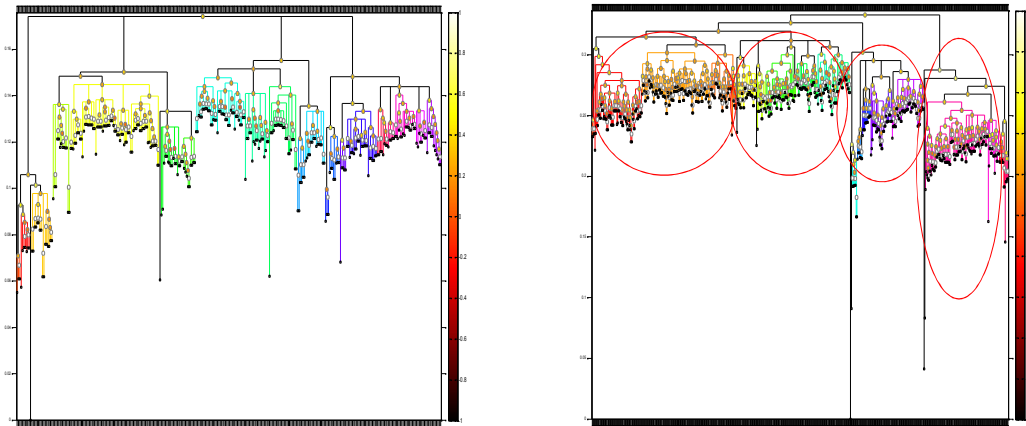
Authors have declared that no competing interests exist.

## REFERENCES

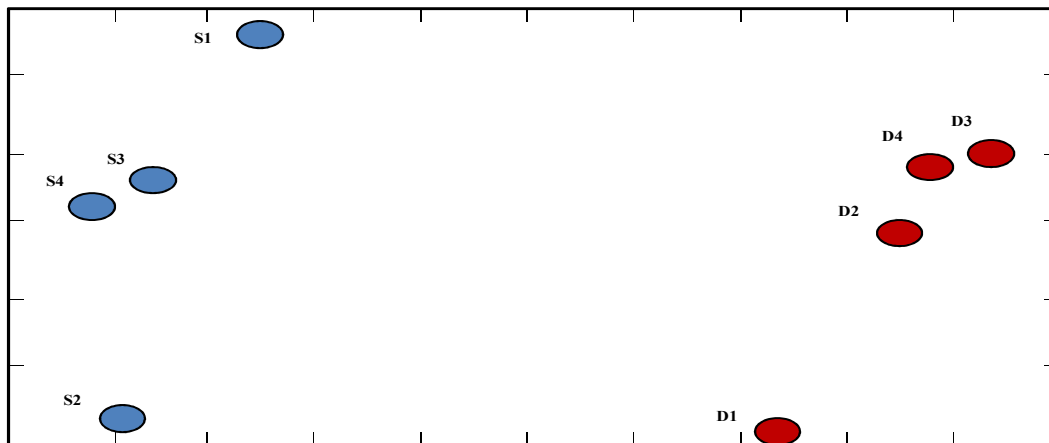
1. EFSA. General principles for the collection of national food consumption data in the view of a pan-European dietary survey. *The EFSA Journal*. 2009;7(12):1435.
2. Newby PK, Tucker KL. Empirically derived eating patterns using factor or cluster analysis: A review. *Nutr Rev*. 2004;62(5):177-203.
3. Moeller SM, Reedy J, Millen AE, Dixon LB, Newby PK, Tucker KL, et al. Dietary patterns: challenges and opportunities in dietary patterns research an Experimental Biology workshop, April 1, 2006. *J Am Diet Assoc*. 2007;107(7):1233-9.
4. Reedy J, Wirfalt E, Flood A, Mitrou PN, Krebs-Smith SM, Kipnis V, et al. Comparing 3 dietary pattern methods-cluster analysis, factor analysis and index analysis--With colorectal cancer risk: The NIH-AARP Diet and Health Study. *Am J Epidemiol*. 2010;171(4):479-87.
5. Tucker KL. Dietary patterns, approaches and multicultural perspective. *Appl Physiol Nutr Metab*. 2010;35(2):211-8.
6. Pryer JA, Rogers S. Dietary patterns among a national sample of british children aged 1 1/2-4 1/2 years. *Public Health Nutr*. 2009;12(7):957-66.
7. Smith AD, Emmett PM, Newby PK, Northstone K. A comparison of dietary patterns derived by cluster and principal components analysis in a UK cohort of children. *Eur J Clin Nutr*. 2011;65(10):1102-9.
8. Grieger JA, Scott J, Cobiac L. Cluster analysis and food group consumption in a national sample of Australian girls. *J Hum Nutr Diet*. 2012;25(1):75-86.
9. Fahey MT, Ferrari P, Slimani N, Vermunt JK, White IR, Hoffmann K, et al. Identifying dietary patterns using a normal mixture model: Application to the EPIC study. *J Epidemiol Community Health*. 2012;66(1):89-94.
10. Patterson BH, Dayton CM, Graubard BI. Latent class analysis of complex sample survey data: Application to dietary data. *J Am Stat Assoc*. 2002;97(459):721-9.
11. Vermunt JK, Magidson J. Latent class models for classification. *Comput Stat & Data Anal*. 2003;41:531-7.
12. Padmadas SS, Dias JG, Willekens FJ. Disentangling women's responses on complex dietary intake patterns from an Indian cross-sectional survey: A latent class analysis. *Public Health Nutr*. 2006;9(2):204-11.
13. Sotres-Alvarez D, Herring AH, Siega-Riz AM. Latent class analysis is useful to classify pregnant women into dietary patterns. *J Nutr*. 2010;140(12):2253-9.
14. Soeria-Atmadja D, Lundell T, Gustafsson MG, Hammerling U. Computational detection of allergenic proteins attains a new level of accuracy with in silico variable-length peptide extraction and machine learning. *Nucleic Acids Res*. 2006;34(13):3779-93.
15. Andersson CR, Gustafsson MG, Strombergsson H. Quantitative chemogenomics: machine-learning models of protein-ligand interaction. *Curr Top Med Chem*. 2011;11(15):1978-93.
16. McGuire MF, Iyengar MS, Mercer DW. Computational approaches for translational clinical research in disease progression. *J Investig Med*. 2011;59(6):893-903.
17. Zou J, Han Y, So SS. Overview of artificial neural networks. *Methods Mol Biol*. 2008;458:15-23.
18. Hammann F, Drewe J. Decision tree models for data mining in hit discovery. *Expert Opin Drug Discov*. 2012;7(4):341-52.
19. Stiglic G, Kocbek S, Pernek I, Kokol P. Comprehensive decision tree models in bioinformatics. *PLoS One*. 2012;7(3):33812.
20. Breiman L. Random forests. *Machine Learning*. 2001;45:5-32.

21. Chen X, Ishwaran H. Random forests for genomic data analysis. *Genomics*. 2012;99(6):323-9.
22. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci USA*. 2002;99(10):6567-72.
23. Geurts P, IRRthum A, Wehenkel L. Supervised learning with decision tree-based methods in computational and systems biology. *Mol Biosyst*. 2009;5(12):1593-605.
24. Suarez-Farinas M, Shah KR, Haider AS, Krueger JG, Lowes MA. Personalized medicine in psoriasis: Developing a genomic classifier to predict histological response to alefacept. *BMC Dermatol*. 2010;10:1.
25. Christin C, Hoefsloot HC, Smilde AK, Hoekman B, Suits F, Bischoff R, et al. A critical assessment of feature selection methods for biomarker discovery in clinical proteomics. *Mol Cell Proteomics*. 2013;12(1):263-76.
26. Hearty AP, Gibney MJ. Analysis of meal patterns with the use of supervised data mining techniques-artificial neural networks and decision trees. *Am J Clin Nutr*. 2008;88(6):1632-42.
27. Edberg A, Freyhult E, Sand S, Fagt S, Kildegaard-Knudsen V, Frost Andersen L, et al. Discovery and characterisation of dietary patterns in two Nordic countries. Using non-supervised and supervised multivariate statistical techniques to analyse dietary survey data. Copenhagen: Nordic Council of Ministers. 2013;548.
28. Enghardt Barbieri H, Pearson M, Becker W. Riksmaten - barn 2003. Livsmedels- och näringsintag bland barn i Sverige. Uppsala: Livsmedelsverket; 2006.
29. Soeria-Atmadja D, Onell A, Borga A. IgE sensitization to fungi mirrors fungal phylogenetic systematics. *J Allergy Clin Immunol*. 2010;125(6):1379-86e1.
30. Maddah F, Soeria-Atmadja D, Malm P, Gustafsson MG, Hammerling U. Interrogating health-related public databases from a food toxicology perspective: Computational analysis of scoring data. *Food Chem Toxicol*. 2011;49(11):2830-40.
31. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. New York: Springer-Verlag; 2001.
32. R. R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical computing. Vienna; 2012. Available from: <http://www.R-project.org/>.
33. WHO/IPCS. Principles and methods for the risk assessment of chemicals in foods. EHC 240. Environmental Health Criteria. 2009;240.

APPENDIX



**Fig. 1. Multi-branching cluster dendrograms, produced by means of the OMB-DHC algorithm. Left panel: Output based on dietary survey data on Danish consumers (4-5 years old). Four distinct aggregations appear at the highest hierarchical level and were labelled dietary prototypes. Right panel: Output based on an excerpt of *Riksmaten - barn 2003*, encompassing 4 year old subjects. The data precipitate in two overarching clusters. Owing to segregation of one such cluster (left-centre part of the image), three aggregations at subordinate hierarchical levels - jointly with the rightmost top level cluster - were assigned dietary prototypes**



**Fig. 2. A second tier exploratory analysis, using classical multi-dimensional scaling (MDS; reduced to two dimensions in the display), of 8 dietary prototype profiles, which were obtained in a previous step via clustering (OHM-DHC) analysis. Four prototypes were obtained from Swedish and four from Danish age-matched preschool consumer data**

© 2014 Hammerling et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Peer-review history:*  
 The peer review history for this paper can be accessed here:  
<http://www.sciencedomain.org/review-history.php?iid=513&id=30&aid=5227>