2009

Identification of food groups for use in a self-administered, computer-assisted diet history interview for use in Australia

S. Burden
University of Wollongong, sburden@uow.edu.au

Y. C. Probst
University of Wollongong, yasmine@uow.edu.au

D. G. Steel
University of Wollongong, dsteel@uow.edu.au

Linda C. Tapsell
University of Wollongong, ltapsell@uow.edu.au

Publication Details

This article was originally published as Burden, S, Probst, YC, Steel, DG & Tapsell, LC, Identification of food groups for use in a self-administered, computer-assisted diet history interview for use in Australia, Journal of Food Composition and Analysis, 22(2), 2009, 130-136.

Identification of food groups for use in a self-administered, computer-assisted diet history interview for use in Australia

Abstract
To develop a set of food groups for use in a self-administered, computer-assisted diet history interview for use in Australia by combining foods into groups so as to minimize database error in the macronutrient values for the groups. The program needs to appropriately balance the level of detail used with the load on respondents and errors associated with categorization of foods into groups.

Keywords
diet history, food groups, cluster analysis, stepwise regression, association rules

Publication Details
This article was originally published as Burden, S, Probst, YC, Steel, DG & Tapsell, LC, Identification of food groups for use in a self-administered, computer-assisted diet history interview for use in Australia, Journal of Food Composition and Analysis, 22(2), 2009, 130-136.

Identification of food groups for use in a self-administered, computer-assisted diet history interview for use in Australia

Sandy Burden

PhD Candidate, School of Mathematics and Applied Statistics, University of Wollongong, NSW 2522, Australia. Fax: +61 2 42214845, Email: alh98@uow.edu.au

Yasmine C. Probst*

Research Fellow, Smart Foods Centre, University of Wollongong, NSW, Australia. Ph: +61 2 4221 5302, Fax: +61 2 4221 4844, Email: yasmine@uow.edu.au

David G. Steel

Associate Dean (Research), Faculty of Informatics and Director, Centre for Statistical and Survey Methodology, University of Wollongong, NSW 2522, Australia. Ph: +61 2 42213823, Fax: +61 2 42214845, Email: dsteel@uow.edu.au

Linda C. Tapsell

Professor and Director, National Centre of Excellence in Functional Foods, University of Wollongong, NSW, Australia. Ph: +61 2 4221 3152, Fax: +61 2 4221 4844, Email: ltapsell@uow.edu.au

* Corresponding author

ABSTRACT

Objective: To develop a set of food groups for use in a self-administered, computer-assisted diet history interview for use in Australia by combining foods into groups so as to minimize database error in the macronutrient values for the groups. The program needs to appropriately balance the level of detail used with the load on respondents and errors associated with categorization of foods into groups.

Materials and Method: Various statistical techniques were utilised to aggregate a large number of food items into compositionally and conceptually similar groups.

Statistical analyses performed: Exploratory statistical analysis; cluster analysis; stepwise regression analysis; and association rule analysis.

Results: A database containing 433 food groups was created which minimised the level of database error in the resulting data collection.

Conclusions: Although some database error was introduced by aggregating food items into groups, the magnitude of the errors was reasonable considering other error sources.

Applications: Collection of an individual's diet history and measurement of energy and macro-nutrient intake.

Keywords:

Diet history

Food groups

Cluster analysis

Stepwise regression

Association rules

INTRODUCTION

Traditional pen and paper dietary assessment methodology

The diet history interview is a common method for assessing dietary habits (Burke, 1947, Tapsell et al., 2002).. The interview is begins in an open-ended manner to allow patients to freely recall foods eaten, and is followed up with closed questions to gain necessary detail. Disadvantages of the diet history interview is the time taken to administer and the need for a skilled interviewer to facilitate the process (Monnier et al., 2001). In population survey research, a closed version known as a food frequency questionnaire (FFQ) is often used (Ocke et al., 1997). Participants indicate the frequency of consumption of a defined list of foods. This technique has the advantage of requiring less time and fewer resources, however, accurate estimation of energy and nutrient intake is difficult due to the limited number of foods discussed and the error associated with the increased cognitive demands. A FFQ usually considers around 100 foods while in Australia nutrient data is available for over 4500 foods.

Automated dietary assessment methodology

Today, computing technology allows for various components of a dietary assessment to be automated. The use of computer-assisted interviews has the advantage of incorporating some of the benefits of both face-to-face interviews and the FFQ format while improving the time efficiency and processing speed. Computer-assisted interviews are also well acceptance by participants (Knapp and Kirk, 2003, Weber et al., 2003) and can range from an interviewer-administered format where a dietitian or researcher asks all the questions and enters the responses into the

computer, through to a self-administered format where the patient or participant reads or listen to the questions and enters the responses into the computer themselves.

The creation of a computer-assisted diet history website, known as DietAdvice, which contains a database of foods and associated nutrient profiles, was undertaken from 2003-2005 (Probst et al., 2007, Probst and Tapsell, 2007b). The website was developed to capture self-administered reporting of usual dietary intake. Using the website, users complete an interview containing a large but fixed number of foods. The website was developed using a multiple pass approach in which the user is guided through food group based questions of increasing detail followed by portion size and frequency of consumption in the final pass.

The output variables of interest in this project were energy and macronutrient intakes. Using the system, individuals can complete an interview without the need for an interviewer. Upon completion of the assessment a dietitian is able to remotely access the site and develop individualised dietary advice. An integral part of this project was the choice of foods to include in the database and the nutrient profiles to be associated with them. The website needed to allow easy navigation by the participant and a reasonable interview time, whilst maintaining sufficient detail for analysis and minimising database error.

Database error from grouped food data

Although many types of error are associated with diet history interview, in this paper we focus on the error introduced when assigning a nutrient profile to a grouping of foods rather than using the nutrient composition information for individual food items (database error). This type of error occurs when several food items with similar but different nutrient profiles are grouped into a single aggregate item. The optimal

nutrient profile for the group reflects the approximate composition of the component food items. However, as each user consumes different amounts of the foods within each food group and only the average composition of the food group is reported, database error is resultantly introduced.

To minimise database error, all ~4500 foods contained in the Australian nutrient database (NUTTAB 1995) were required. To an untrained user such a list of foods would be unworkable. Consequently, the aim of this study was to develop a hierarchical database containing a set of food groups for use in the DietAdvice database. The website needed to appropriately balance the level of detail used with the load on respondents and the database error in the energy and macronutrient intakes. This goal was achieved by grouping similar food items and discarding rarely eaten items using several statistical techniques. This paper will also compare the food grouping system created with the one on which it was modelled (NNS95).

METHODS

A hierarchical database of foods was developed by grouping the food items from NNS95 into a smaller number of groups for use in the DietAdvice database. The food groups were created using iterations of hierarchical cluster analysis and stepwise regression to identify compositionally similar food items and to identify those groups which were important for understanding variation in the diets of individuals. After each iteration, the results were combined with the professional judgement of several dieticians to create a database which maintained sufficient detail to discriminate between different diets. Finally, association analysis was used to identify foods commonly reported with one another, to help identify a logical database structure. The database error associated with the final groups was then calculated.

1995 National Nutrition Survey (NNS95) food database

The 1995 NUTTAB database was the basis of the National Nutrition Survey (NNS95) food database. The DietAdvice food database was developed using the NNS95 conducted by the Australian Bureau of Statistics (ABS) (Australian Bureau of Statistics, 1995). As NNS95 is the most recent nationwide survey of dietary intake for Australia, it was seen as the most appropriate basis for the development of the database to be used in the DietAdvice project.

The NNS95 includes detailed records of the 24-hour recall interview for 13,858 individuals, along with adjustments for usual intake, various survey weights, personal characteristics and demographic variables. The 24-hour recalls are used in this study 'as-reported' in the survey. No allowances have been made for under-reporting, over-reporting or recall errors, or for the differences between the 24-hour recall results and the usual intake of an individual. The unadjusted values are considered more representative of the results expected to be obtained during future diet history interviews. Survey weights have been applied so the results are representative of the Australian population within the scope of the NNS95.

The NNS95 food database is hierarchical in nature, with three nested levels of groups containing 21, 106 and 370 groups respectively. Approximately 4500 individual food items are included in the database. The actual food database of the NNS95 was not considered appropriate for use as the DietAdvice database due to the differences in their methodology. NNS95 was administered by trained interviewers who used the hierarchy to find foods while the DietAdvice food groups would be used in place of individual foods for the computer user to select from. The third level

NNS95 groups were used as a starting point for analysis and hence for comparison with the final DietAdvice groups.

Statistical framework

In this study, individual rather than population estimates of the macronutrient and energy variables are required because the aim is to minimise the error associated with the use of food group based nutrient profiles for each individual completing the interview. The framework assumes that all foods which are eaten by the Australian population are categorised into P distinct items which can be partitioned into M separate groups in some fashion.. The k^{th} food item in group j is denoted p_{jk}, j=1,…,M; k=1,…,p_j, creating a 2-level hierarchy of foods. The following sections describe how the P food items in NNS95 were divided into M groups for use in the DietAdvice database.

A standard 1g quantity of any food item contains a fixed number of kilojoules (kJ) of energy and a defined proportion of macronutrients, which contribute a fixed proportion of energy to the total energy of the food. Let the energy contained in 1g of food be denoted x_{jk}, and assume the quantity is measured exactly. Presuming that for each individual i, i=1,...,n from the NNS95, the quantity of food item p_{jk} eaten by individual i in a given 24 hour period is denoted w_{ijk}. This quantity is also considered to be measured exactly for each individual in the sample, however in reality measurement and reporting errors will be present. Ignoring these potential errors, the total daily energy intake for person i, denoted Y_i, is given by (1).

$$Y_i = \sum_{k=1}^{P_j} x_{jk} w_{ijk} \quad (1)$$

When food items are grouped for data collection purposes in the DietAdvice database, the energy content of each food item x_{jk} in group j must be replaced with some kind of proxy X_j for the whole group. One such proxy - the population unbiased, minimum variance estimator - of the average energy per group X_j, is estimated by (2) for each group j=1,...,M. The term π_i is the selection probability of the i^{th} person in the NNS95 sample dataset. The estimator is the average of the energy content of a gram of each food in the group, weighted by the estimated population consumption of that food in grams. Using this estimator, the total daily energy intake for person i, is estimated by $\hat{Y}_i$ using (3).

$$X_j = \frac{\sum_{i=1}^{n}\sum_{k=1}^{P_j} x_{jk} w_{ijk} / \pi_i}{\sum_{i=1}^{n}\sum_{k=1}^{P_j} w_{ijk} / \pi_i} \quad (2) \qquad\qquad Y_i = \sum_{j=1}^{M} X_j \sum_{k=1}^{P_j} w_{ijk} \quad (3)$$

The database error associated with aggregating foods into groups can be measured as a bias: a measure of the precision of estimation. For an estimator $\hat{Y}_i$ of Y_i the bias is:

$$Bias(\hat{Y}_i) = \hat{Y}_i - Y_i = \sum_{j=1}^{M}\sum_{k=1}^{P_j} (X_j - x_{jk}) w_{ijk} \quad (4)$$

Similar results are obtained for each macronutrient and the percentage energy derived from each macronutrient. In these cases, x_{jk} represents the grams or the proportion of energy from the macronutrient of interest respectively.

A complex sample design was utilised in NNS95 to obtain a sample of the Australian population, with varying selection probabilities. To obtain an unbiased estimator of population intake, the intake for each individual in the sample is weighted by the inverse of the probability of that person being selected, denoted π_i i.e. $E\left[\sum_{i=1}^{n} \frac{Y_i}{\pi_i}\right] = \sum_{i=1}^{N} Y_i$, where N is the number of individuals in the population. The

resulting estimate of the population intake is given by (5) which can be decomposed into an estimator of the population intake for each food group as shown in (6). Ignoring the various measurement and reporting errors, the population estimates are unbiased and the mean square error is the same as the sampling variance of $\hat{Y}$.

$$\hat{Y} = \sum_{i=1}^{n} \frac{1}{\pi_i} Y_i = \sum_{i=1}^{n} \sum_{j=1}^{M} \sum_{k=1}^{P_j} \frac{x_{jk} w_{ijk}}{\pi_i} \quad (5)$$

$$\hat{Y}_j = \sum_{i=1}^{n} \frac{1}{\pi_i} Y_{ij} = \sum_{i=1}^{n} \sum_{k=1}^{P_j} \frac{x_{jk} w_{ijk}}{\pi_i} \quad (6)$$

Cluster analysis

Cluster analysis is a widely used statistical technique that has been used for defining both food exposure patterns (Wirfalt et al., 2000) and for combining foods into groups (Akbay et al., 2000, Windham et al., 1985). In this study, hierarchical cluster analysis was used as a tool to help inform decision-makers of the compositional similarity between foods, so that foods which were compositionally most similar were grouped together. As some foods which are compositionally similar are conceptually dissimilar, for example custard and egg noodles, professional judgement was then used to identify conceptually similar foods (Probst and Tapsell, 2007a).

There is no single approach to clustering techniques. Different clustering algorithms will give different results, some of which may be more appropriate for the given data. Cluster results are also dependent on the relative magnitude of the variables included in the analysis. For this project, three hierarchical clustering algorithms were applied to the food item data: average linkage, complete linkage and Ward's method (Johnson and Wichern, 2002). For all techniques, Euclidean distance was used to measure similarity for each of the variables: total energy (kJ), protein (g), carbohydrate (g), PUFA (g), MUFA (g) and SFA (g),.All variables were standardised across the dataset to have a mean of zero and standard deviation of one.and were

weighted equally with respect to the analysis. The inclusion of the three main types of fatty acids (rather than total fat) was necessary to help understand the relative fat composition in different groups. Due to the size of the NNS95 database, clustering was performed using subsets of data containing conceptually similar items. For example all types of cheese or bread.

Stepwise regression

Stepwise regression analysis (Johnson and Wichern, 2002) has been used successfully to identify food items to include in a FFQ (Shai et al., 2004). In this project it was utilised to identify the food groups which are the best predictors of variability in intake for inclusion in the DietAdvice database.

In each analysis, the response variable was the total energy or macronutrient intake for each survey respondent (in kilojoules for energy and grams for the other macronutrients - carbohydrate, protein, total fat, SFA, MUFA and PUFA), or the percentage of energy derived from the macronutrient for each respondent. The predictor variables were all the food items or groups in the database (as the analysis was repeated for both the NNS95 and DietAdvice databases). For each individual, the value of each predictor variable was calculated as the total weight (in grams) of foods in the group that was consumed in the 24-hour recall interview.

In stepwise regression the initial model contains no predictors. During computation, predictor variables are added or deleted in a stepwise fashion using rules for inclusion until no more variables meet the criteria. The final regression equation then contains a subset of V statistically significant variables from the set of M possible predictors for that response.

New predictors were added if the predictor will produce the largest comparative increase in the F-statistic and a p-value ≤ 0.15.. Predictors were deleted when they produced a smaller increase in the F-statistic and a high p-value ≥ 0.15. For each regression, the R^2 value has been used to measure the strength of the relationship between a group of predictors and the response. This value is a measure of the correlation between Y and the best linear combination of the set of predictors W.

Association rules

Finally, to aid navigation through the database and to create reminders and prompts, an association analysis was conducted using the NNS95 food groups. Association analysis (Witten and Frank, 2000) can identify the food items which are regularly eaten together at the same meal. Using the results of this analysis, foods commonly consumed together can be placed nearby one another in the food list, making them easy to find. Similarly, within the interview if one food (say cereal) is eaten, the associated foods (such as milk) can be highlighted in some way as a reminder. The use of association analysis in nutrition is not widely reported. Being a data mining tool, it is generally used to identify associations between large numbers of items contained in a database. In this analysis sample weights were not included, so the results reflect food habits in the sample rather than the population.

To generate association rules for the three main meals defined in the NNS95 (breakfast, lunch and dinner), the modified apriori algorithm was utilised (Witten and Frank, 2000). The apriori algorithm comprises two main steps. Firstly it finds all groups of food items whose support is greater than the minimum support level (in this case 1% of transactions). These combinations are called frequent item sets. The algorithm uses the frequent item sets to generate the desired rules. The rules hold

only if confidence is above the minimum confidence level chosen (50% in this case). For the NNS95 data, the rules were generated for food groups. All analyses were completed using SAS/Stat Software (Version 8, Release 8.02, SAS Institute Inc. Cary, NC, USA).

RESULTS

Initial statistical analysis for food exclusion and hierarchy formation

The initial framework of the DietAdvice database was created from the NNS95 database. Of the 4500 food items included in the NNS95 database, only 4175 were consumed by individuals within the survey. The remaining 325 foods were immediately discarded. Food items in the NNS95 were categorised into various meals and snack groups and aggregated into the 370 groups from the third level of the NNS95 food hierarchy. Food groups consumed at each meal by less than 1% of the NNS95 sample were also discarded.

The basic key groups were then chosen from the most commonly consumed food groups in the NNS95 by energy content and by frequency of consumption for each meal. The top 10 groups for the three main meals by energy and frequency are shown in Table 1. These tables were also created for individual food items to ensure that all important food items were included in the groups. Comparisons were made between the frequency based and energy based food lists to ensure that groups regularly appearing in lower positions in each list were also included in the new database. As a comparison, using the final DietAdvice database which contains 433 groups, shows the equivalent top food groups for the three main meals. [TABLE 1]

Cluster analysis results

The results from analyses were presented in several ways. Dendrograms were produced to help identify food items or groups of food items within the 370 NNS95 food groups which did not appear to fit in the original (Probst and Tapsell, 2007a). Tables showing possible options for splitting a single group into a given number of sub-groups based on the different clustering techniques (data not shown) were also produced. The table included here is indicative only. It contains the energy, carbohydrate, protein and total fat content of each food item in the group (components of fat are not included here for space considerations). It also contains the suggested groupings (1-6) identified by each of the clustering techniques and the final groups (coded A-F) chosen for the project. The table exemplifies how foods in the single original group have been split into separate groups based on both conceptual differences between the foods (quiche, pies and vol-au-vents) and on differences in composition identified using the clustering techniques (vegetable and meat based quiches). Other foods originally within different groups in NNS95 may also be included in each of the final groups shown.

Regression results

In addition to capturing the main groups contributing to a persons intake of each macronutrient and energy, it was also desired that sufficient detail be maintained in the database, so that an understanding of the variation in intake between individuals could be identified.

To identify the foods that accounted for the greatest variation in individuals intakes of energy or the defined macronutrients, stepwise regression analysis for energy and each macronutrient intake was undertaken. Total energy or macronutrient intake were included as response variables and both food items and the 370 NNS95

groups defined previously were used as predictors. As an example of the results, Table 2 shows the 20 food groups from the NNS95 database which account for the greatest variation in energy intake. The food groups in Table 2 account for more than 50% of the variation in energy intake between individuals, as measured by the cumulative R^2. For comparative purposes, the groups are also ranked on the basis of percentage contribution total intake. The ranks show that the foods which account for variation between individuals are also important in measuring total intake and so have generally already been included in the DietAdvice database. Similar results were created for each macronutrient of interest. The top 20 DietAdvice food groups which account for the greatest variation in energy intake. Compared with the NNS95 groups both the cumulative R^2 and total intake are lower for the DietAdvice food groups, which is to be expected with a greater number of groups in the database. However, the new food groups are more specific, generally containing more similar items than the NNS95 groups, so the DietAdvice groups better explain the food choices accounting for variation in diet. For example, it is the consumption of full fat ice-cream (rank 5 in the DietAdvice database) rather than ice-cream in general (rank 10 in NNS95) which explains a lot of variation in energy intake [INSERT TABLE 2].

Table 3 shows the cumulative R^2 obtained from stepwise regression using macronutrients and percentage energy from each macronutrients as a response. Cumulative R^2 is a measure of the percentage variation in consumption of each macronutrient and energy which can be explained by the 370 NNS95 food groups compared with the original ~4500 individual food items. In all cases variations in diet with regard to the components of fat are the most difficult to capture when using food groups instead of individual items. However, considerable variability can still be explained when only the 50 groups accounting for the greatest variability in

individuals diets are included in the analysis. This suggests that some groups act as a proxy measure for a particular type of diet. Table 3 also shows the cumulative R^2 value obtained for each stepwise regression analysis using the final 433 DietAdvice food groups. In all cases, analyses were repeated using only the top 50 food groups which explained variation in that response. The table shows that variability in all macronutrients between individuals diets can be detected using the DietAdvice groups. In total, the DietAdvice groups account for 98% of the variability in energy intake in peoples diets compared with using the ~4500 individual food items. The new groups explain variation in individual diets better than the original groups from NNS95, although once again the variation in the components of total fat are the least well explained. [INSERT TABLE 3]

Assessment of error due to food grouping

An assessment of the error associated with grouping foods into the 433 groups in the DietAdvice database and the 370 groups in the NNS95 database rather than including all 4500 food items was undertaken. Box plots of the percentage bias and absolute bias for each macronutrient of interest for each decile of the population estimated from the NNS95 were created.

Comparing the percentage bias introduced by using food groupings several interesting features can be identified. The levels of error are all lower for the DietAdvice groups compared with NNS95 groups, suggesting that regrouping the data has been successful in reducing this error. Also, average percentage bias remains relatively constant for each decile of the population, suggesting that those with generally low or high intakes of particular macronutrients are not adversely affected by grouping. Finally, it can be seen that for all deciles except decile 1 the average

level of bias is less than 5% for energy, protein and carbohydrate and less than 10% for total fat.

The mean, median, standard deviation and 95% quantile of the percentage bias are shown in Table 4 for each macronutrient. In all cases the difference between the mean and median show that the distribution of bias is skewed and there are some large outliers. In general the bias is improved when using the DietAdvice groups. The reduction in database size from 4500 items to 433 groups has in many cases introduced a bias of less than 10%, although for some individuals significant biases will result. [INSERT TABLE 4]

Assessment of associations between foods

The analysis of food associations resulted in a list of foods that are commonly eaten together. An example of the ten association rules with the greatest support for full fat milk is given in Table 5. The first rule in Table 9 shows that sugar implies milk at breakfast. The rule is valid because 38.3% of the sample in NNS95 reported sugar consumption at breakfast and of those individuals who consumed sugar, 61.9% also consumed milk. Using professional judgement, this list was assessed by a number of dieticians and it was seen to be similar to food combinations reported in a traditional diet history interview. [INSERT TABLE 5]

DISCUSSION

The final structure of the database was a two tier hierarchy containing 433 groups. The formation of the groups was informed by the statistical analysis described here, but also involved some professional judgement on the part of trained dieticians to ensure that food groups were clearly identifiable to untrained respondents. In a

subsequent step, food names were adjusted to ensure usability of the food lists and face validity testing was performed by employing trained professionals unrelated to the project and asking them to construct food lists that they believe comprise the Australian diet and may be addressed during a traditional diet history interview. The resultant documents were used to cross-check the hierarchical framework and ensure common food items were not missing and the most common names had been used.

Following completion of the hierarchical framework further reduction of the food groups occurred during the development of the user interface. Following advice and expertise of web and graphic designers some food groups needed to be separated further to allow friendly on-screen display. This data has been reported elsewhere (Probst and Tapsell, 2007a).

The greatest advantage of the DietAdvice food database is that it combines a large number of food items into a reasonable number of foods groups to include in a computerised diet history interview. The foods are agglomerated in such a way that the most commonly eaten foods have the greatest influence on the macronutrient composition of the group and in this way the error due to grouping, which we have called database error, is reduced. The food groups successfully identify the main foods consumed and the foods responsible for the greatest variation in intake of the macronutrients of interest.

One significant disadvantage of the technique is that the nutrient values assigned to each food group are based on the NNS95. Due to its age, this data is potentially out-of-date which may affect interview results. However, considering the presence of other errors such as under-reporting and recall errors in this type of interview significant errors are not anticipated. In particular, the nutrient values for each food item are the latest available, it is only the relative weight placed on each

food within each group that may be out of date. These issues can only be ameliorated by undertaking another large nutrition survey in Australia.

Although some database error was introduced by aggregating food items into groups, the magnitude of the errors was reasonable considering other error sources.A final benefit to the process is an understanding of the level of error introduced into an individual's diet by replacing individual foods with aggregated groups. Collection of an individual's diet history and measurement of energy and macro-nutrient intake using computerised technologies will improve the efficiency of dietary assessment techniques and also all for further work in this area to elucidate the level of error associated with reporting of different groups of foods.

ACKNOWLEDGEMENTS

The authors would like to thank Pieta Autenzo, Marijka Batterham, Marian Bare, Linda Blackmore, Rachel Cavanagh, Andrew Dalley, David Elsner, Chester Goodsell, Barry Harper, Lori Lockyer, Owen McKerrow, Karl Mutimer, David Skoumbourdis and Therese O'Sullivan for their contributions to this project.

FUNDING DISCLOSURE

The project was funded under the ARC Linkage Grants scheme and is a joint project between the University of Wollongong, Illawarra Division of General Practice and Xyris Software Pty Ltd.

REFERENCES

AKBAY, A., ELHAN, A., OZCAN, C. & DEMIRTAS, S. (2000) Hierarchical cluster analysis as an approach for systematic grouping of diet constituents on the basis of fatty acid, energy and cholesterol content: application on consumable lamb products. *Medical Hypotheses,* 55**,** 147-154.

AUSTRALIAN BUREAU OF STATISTICS (1995) National Nutrition Survey: Nutrient Intakes and Physical Measurements. Australian Bureau of Statistics.

BURKE, B. A. (1947) The dietary history as a tool in research. *Journal of the American Dietetic Association,* 23**,** 1041-1047.

JOHNSON, R. A. & WICHERN, D. W. (2002) *Applied multivariate statistical analysis,* Upper Saddle River, New Jersey, Prentice Hall.

KNAPP, H. & KIRK, S. A. (2003) Using pencil and paper, Internet and touch-tone phones for self-administered surveys: does methodology matter? *Computers in Human Behavior,* 19**,** 117-134.

MONNIER, L., COLETTE, C., PERCHERON, C., PHAM, T. C., SAUVANET, J. P., LEDEVEHAT, C., et al. (2001) [Dietary assessment in current clinical practice: how to conciliate rapidity, simplicity and reliability?]. *Diabetes & Metabolism.,* 27**,** 388-395.

OCKE, M., BUENO-DE-MESQUITA, H., GODDIJN, H., JANSEN, A., POLS, M., VAN STAVEREN, W., et al. (1997) The Dutch EPIC food frequency questionnaire. I. Description of the questionnaire, and relative validity and reproducibility for food groups. *Int J Epidemiol,* 26**,** S37-48.

PROBST, Y. & TAPSELL, L. (2007a) What to Ask in a Self-Administered Dietary Assessment Website: The Role of Professional Judgement. *Journal of Food Composition & Analysis,* 20**,** 696-703.

PROBST, Y. C., MCKERROW, O., LOCKYER, L., STEEL, D. & TAPSELL, L. C. (2007) First stage development of a dietary assessment website for use in general practice. *International Journal of Learning Technology* 3**,** 32-50.

PROBST, Y. C. & TAPSELL, L. (2007b) What to ask in self-administered diet history interviews: The role of professional judgement. *Journal of Food Composition and Analysis,* 20**,** 696-703.

SHAI, I., SHAHAR, D., VARDI, H. & DRORA, F. (2004) Selection of food items for inclusion in a newly developed food-frequency questionnaire. *Public Health Nutrition,* 7**,** 745-749.

TAPSELL, L. C., DANIELS, S., MARTIN, G. S., KNIGHTS, S. K. & MOSES, R. G. (2002) Performance of a research diet history for use in clinical studies involving pregnant women with and without gestational diabetes mellitus in the Illawarra region, New South Wales. *Nutrition & Dietetics: The Journal of the Dieticians Association of Australia,* 59**,** 127-135.

WEBER, B., SCHNEIDER, B., FRITZE, J., GILLE, B., HORNUNG, S., KUEHNER, T., et al. (2003) Acceptance of computerized compared to paper-and-pencil assessment in psychiatric inpatients. *Computers in Human Behavior,* 19**,** 81-93.

WINDHAM, C. T., WINDHAM, M. P., WYSE, B. W. & HANSEN, R. G. (1985) Cluster analysis to improve food classification within commodity groups. . *Journal of the American Dietetic Association,* 85**,** 1306-1314.

WIRFALT, E., MATTISSON, I., GULLBERG, B. & BERGLUND, G. (2000) Food patterns defined by cluster analysis and their utility as dietary exposure variables: a report from the Malmo Diet and Cancer Study. *Public health nutrition,* 3**,** 159-173.

WITTEN, I. H. & FRANK, E. (2000) *Data Mining: practical machine learning tools and techniques with Java implementations,* San Diego, Academic Press.