Public Health Nutrition: 4(2B), 689–692 DOI: 10.1079/PHN2001155

Correlation analyses as a step to identify foods that are sources of inter-individual variability in nutrients; their use for the development of food based dietary guidelines

Catherine Leclercq* and Davide Arcella
National Institute for Food and Nutrition Research (INRAN), Via Ardeatina, 546, 00178 Rome, Italy

Abstract

Food Based Dietary Guidelines (FBDG), in order to be attainable, should take into account the cultural context of the population for which they are developed. They need to focus on foods that actually determine the intake of nutrients for which desirable changes have been identified. Inter-individual variability of nutrient intakes – a measure of the heterogeneity of intakes – is determined by the variability in the consumption of its food sources among the population. The foods that determine a high proportion of inter-individual variability in nutrient intake can be identified on the basis of data banks that describe food and nutrient intakes in the population. The experience accumulated to design Food Frequency Questionnaire through selection of key foods suggests that high quantitative contributors to a specific nutrient may not be important determinants of its intake. On the other hand, the Pearson correlation between the intake of each food source and the total nutrient intake allows quantification of the percentage of variability explained by each item and takes into account the possibility of correlations between different food sources. Once a key food is identified, several strategies are available to modify its intake in the population: through changes in the percentage of consumers/in the mean portion size/in the frequency of intake. The anticipated level of change can be predicted according to the strategy adopted.

Keywords
Nutrient intake
Food sources
Patterns of food consumption
Inter-individual variability
Pearson correlation
Food based dietary guidelines

Introduction

It is largely recognised that Food Based Dietary Guidelines (FBDG), in order to be attainable, should take into account the cultural context of the population for which they are developed[1]. Study of the relationship between health and diet has led indications in terms of desirable nutrient intakes. The intake of a specific nutrient may more easily be increased or decreased if the desired level of intake is already attained in some segments of the population; consumers should be oriented towards dietary patterns that are both healthy and in line with some existing patterns. If the food sources of a specific nutrient are consumed with very homogeneous consumption patterns in the population, it will be more difficult to alter its intake. For example, most Italian children who consume milk do so with a rather standard pattern in terms of quantity (one cup) and frequency (once a day, at breakfast). Even if milk is a good source of calcium it will be difficult to increase calcium intake through milk since in this population there is no existing pattern of consumption that lead to higher intakes of milk.

Food based dietary guidelines will therefore be more attainable if they focus on foods that actually ‘make the difference’ in terms of nutrients. These foods obviously vary according to country[2,3]. An in depth analysis of the situation is necessary for each of the nutrients considered. In most European countries, databases are available that describe the diet of representative samples of the population in terms of food intakes. The appropriate combination of such data with that of food composition databases enables the quantification at individual level of the nutrient intake from each food group[4]. Several methodological approaches can be used to analyse these data in order to provide useful information for the development of Food Based Dietary Guidelines (FBDG). Quantiles and discriminant analysis allows an assessment of the heterogeneity of intakes for the nutrient under study and to identify contrasting food consumption patterns and their determinants[5]. In this paper, the authors describe a simple tool aimed at identifying the foods that determine a high proportion of inter-individual variability in nutrient intake.

*Corresponding author: Email leclercq@inn.ingrm.it

Preliminary remarks on the quality of databanks

Before proceeding, it is necessary to stress that no result of statistical analyses can be interpreted before a scrupulous analysis of the quality of the data that describe food and nutrient intakes in the population. The main potential sources of bias are (1) the underestimation of food and nutrient intakes due to selective underreporting of certain foods by specific segment of the population[6], (2) the underestimation of the percentage of consumers and overestimation of the inter-individual variability of intakes due to the use of short duration surveys[7], (3) the comparison of incompatible data collected through different survey technique[8], (4) the use of unreliable food composition data[9].

If these sources of error are not taken into account biased Food Based Dietary Guidelines or other nutrition policy actions may be planned.

Methodology for the selection of foods to be included in a Food Frequency Questionnaire (FFQ)

A methodology to assess the proportion of inter-individual variability explained by different foods or food groups has already been developed in a different area from that of Food Based Dietary Guidelines: for the design of data-based quantitative Food Frequency Questionnaires (FFQ). These two research areas have some similarities in their scope and the experience accumulated to design FFQ's can in part be useful for the development of Food Based Dietary Guidelines. In order to develop a FFQ, a limited number of food items that are determinants of the intake of specific nutrients under study must be selected on the basis of sound databases of food and nutrient intake in the target population. The identification of the major quantitative food sources of a nutrient appeared to provide useful but insufficient information: some foods may be high quantitative contributors to a specific nutrient without being important determinants of its intake. Thus, in a study conducted on a large sample of Italian elderly population, eggs were found to be responsible for 32% of the intake of cholesterol but no correlation ($r^2 = 0.00$) was found between the consumption of eggs and cholesterol intake[2]. This could be explained by the fact that eggs are eaten in standard quantities among this population. Having a high or a low intake of cholesterol was mainly due to the consumption of other sources of cholesterol such as cakes and cookies: this group of foods contributed only about 4% of the overall intake but explained 29% of the variance in cholesterol intake. On the other hand, in the case of saturated fats, olive oil was found to be the highest contributor (15% of the overall intake) but was not a determinant of the overall intake ($R^2 = 0.00$). These results suggest that the frequency of intake of cake and cookies is more useful than that of eggs to classify elderly subjects according to their cholesterol intake and that the frequency of intake of olive oil is of no use to classify subjects according to their saturated fat intake.

The foods contributing to a large proportion of the variability in nutrient intake may be identified by stepwise multiple regression where the intake of the nutrient of interest is the dependent variable and intakes of foods are the independent variables. In 1993, Corrao *et al.*[10] performed such analysis to develop a FFQ specific for Italian hospitalised subjects. The scope of the authors was to develop a FFQ with a reduced number of food items with respect to the complete list of 152 items, while maintaining the greatest possible part of the informative contents of the original questionnaire. Results showed that 30 items were necessary to assess 77% of the overall protein intake and 96% of the overall vitamin A intake. A more limited number of items, 5 for vitamin A and 18 for proteins, were sufficient to explain 95% of the variability in these nutrients. These were the foods to be included in the reduced FFQ.

Pearson correlation as a tool to identify key foods

Also in the case of Food Based Dietary Guidelines, foods that explain a large part of the inter-individual variability in nutrient intake are more important than foods that explain a large proportion of nutrient intake. However, the stepwise analysis described above does not provide a complete picture of the foods that 'make the difference' in terms of nutrient intake. If the intakes of two food items which are both determinants of the nutrient under study are correlated, the regression of the weaker will appear to be far lower and its importance underestimated. The Pearson correlation (r) between intake of each food source and the total nutrient intake therefore appear to be a useful complement to identify all those foods that "make the difference" in terms of nutrients. The square of this coefficient (R^2) is a measure of the portion of total variation explained by the intake of each single food, which is not weakened by possible correlation with other food sources. An example of its application may be given by using data from a consumer survey carried out in Italy[11]. A total of 193 teenagers recorded all foods and beverages ingested during 14 consecutive days. Food products were categorised into 14 groups: pulses, cereals, vegetables, tubers, fruit, meat, fish and sea foods, milk and dairy products, eggs, added fats, sweets, non alcoholic drinks, alcoholic drinks and other products. Data were analysed in order to identify the foods that determine fat intake. Results are presented in Table 1. The multiple regression analysis was performed with total fat intake as the dependent variable and the consumption of all food groups as independent variables. Added fats contributed on average 43% of the overall fat intake and were found to be responsible for 96% of the variability in fat intake. Vegetables intake contributed on average less

Table 1 Identification of the food groups that determine total fat intake in a sample of Italian teenagers through regression analysis and correlation analysis

	Food groups (independent variables)	Partial R^2 (dependent variable: total fat intake)	P	Correlation with total fat intake	P	Contribution to total fat intake (g)	%
1	Added fats	0.9647	0.0001	0.831	0.0001	42.9	42.5
2	Sweets	0.0149	0.0001	0.568	0.0001	9.4	9.3
3	Meat	0.0061	0.0001	0.714	0.0001	17.7	17.5
4	Cereals	0.0042	0.0001	0.769	0.0001	6.6	6.5
5	Milk and dairy products	0.0020	0.0001	0.397	0.0001	15.6	15.5
6	Other products	0.0015	0.0001	0.313	0.0001	1.6	1.6
7	Eggs	0.0001	0.0653	0.515	0.0001	2.5	2.5
8	Alcoholic drinks	0.0001	0.0766	0.327	0.0001	0.0	0.0
9	Pulses*	–	–	0.174	0.0154	0.1	0.1
10	Vegetables*	–	–	0.482	0.0001	0.7	0.7
11	Tubers*	–	–	0.3205	0.0001	0.62	0.6
12	Fruit*	–	–	0.2213	0.0020	1.15	1.1
13	Fish and sea foods*	–	–	0.1621	0.0243	0.89	0.9
14	Non alcoholic drinks*	–	–	0.0865	0.2317	1.16	1.2

* These variables did not meet the 0.15 significance level for entry into the model.
n = 193.
Source: unpublished data derived from a consumer survey[11].

than 1% of the overall fat intake and did not meet the 0.15 significance level for entry into the model. However, a correlation was found between vegetables intake and fat intake ($r = 0.48$; $p = 0.0001$), suggesting that vegetables might be a determinant of fat intake. The further step performed to investigate this association was to perform between-food correlation analyses between vegetables and each of the nine food groups that were included in the regression model. The highest correlation was found with added fats intake: $r = 0.59$, $p = 0.0001$. The reason for this association was found by a further analysis of the association between vegetables and added fats within eating occasions (meals) and within mixed dishes in the food intake databank. It appeared that added fats were present in 96% of the eating occasions that included vegetables and in 88% of the mixed dishes that included vegetables as an ingredient. These results suggest that in the cultural context of this study sample, vegetables represent a very small contributor but an important determinant of fat intake: vegetables are almost always seasoned with fats. An increase in vegetable consumption may therefore cause an increase in fat intake if the seasoning habits are not concurrently modified.

Identification of strategies to modify the intake of key foods

Once a food is found to be an important determinant of the intake of a nutrient for which a change is desirable, different strategies are available to try and modify its level of consumption.

The first option to explore is that of increasing or decreasing the percentage of consumers. A low percentage of food consumers may be the main reason for a high correlation between the intake of this food and the total nutrient intake. In the elderly study mentioned above, liver consumption was responsible for only 14% of the mean intake of vitamin A but explained 78% of its variability[3]. Each of the other foods explained less than 6% of the variability in vitamin A. Liver was the main contributor to total vitamin A intake in spite of low frequency of consumption, only 10% of those sampled ate it at least once a week. In this case, the average intake of vitamin A could be increased by increasing the percentage of liver consumers through promotion of this item. On the contrary, vitamin A deriving from tomatoes could not be increased by increasing the percentage of tomatoes eaters: it was already 97%.

If a key food is already consumed by a large proportion of the population, two other options may be considered to increase its consumption: higher eating frequency and/or larger serving sizes. In fact, the daily or weekly consumption of a food is obtained by multiplying the daily or weekly frequency of intake by the average serving size. The theoretical effect on fibre intake of an increase in the frequency of fruit consumption from 1 serving per day to 2, 3 and 4 servings per day was calculated by Gibney[12] on the basis of consumption data of a sample of Irish women. The impact on the average fibre intake was found to be in the order of 11% for each level of change. Another option proposed in the same paper was to increase the serving size for pulses.

Some simple calculations using the above mentioned data on Italian teenagers[11] are proposed here. In this sample, pulses were found to be an important determinant of fibre intake. The theoretical effect of the different strategies that could be used to increase fibre intake through pulses were compared through analysis of the databank. Since pulse consumers already represented 87% of the sample and their mean fibre intake was 15.4 grams per day – only 0.2 grams more than that of non consumers – the incremental change in mean fibre intake

https://doi.org/10.1079/PHN2001155 Published online by Cambridge University Press

Table 2 Identification of three different options leading to the same increase of fibre intake in a sample of Italian teenagers

	Present situation	Options		
		1	2	3
Frequency of pulses consumption among pulses consumers	1 every 5 days	1 every 5 days	**1 every 3 days**	**1 every 4 days**
Average pulses serving size (grams)	80	**131**	80	**105**
Average fibre intake (grams per day)	15.2	16.0	16.0	16.0

Source: unpublished data derived from a consumer survey[11].
Bold character is used to evidence changes.

that would be obtained by increasing the percentage of pulses consumers would be quite small. Increasing the frequency of intake and/or the serving size might be more efficient. Table 2 examines three different options that would theoretically increase the average fibre intake from the actual level of 15.2 to an intermediary target of 16.0 grams per day by modifying the frequency of intake (option 1), the serving size (option 2) or both (option 3). In order to reach the target of 16 g of fibre per day with the option 1, the portion size of pulses should almost double (from 80 to 131 grams), which appears difficult. A change in the frequency of pulses consumption (option 2) from 1 every 5 days to 1 every three days appears more feasible. Option 3 is a convenient hypothesis, it shows that an intake of fibre equal to 16.0 grams per day could theoretically be obtained with a mean frequency of pulses consumption of 1 every 4 days and average serving size increased of one third.

The choice of the option could be based on further statistical analysis by calculating the portion of inter-individual variability in nutrient intakes respectively explained by the frequency of intake and by the serving size using the above described R2. Focus-group attitudinal research could also be carried out to determine which strategy would be most popular.

Conclusion

An in depth analysis of prevailing food habits is a necessary pre-requisite for the development of Food Based Dietary Guidelines. Stepwise regression analyses can usefully be completed by Pearson correlation in order to identify key foods for Food Based Dietary Guidelines. Once a key food is identified, several strategies are available to modify its intake in the population: through changes in the percentage of consumers/in the mean portion size/in the frequency of intake. The anticipated level of change according to the strategy adopted can be predicted through further analysis of the food intake data banks. This should allow to propose feasible and culturally acceptable dietary advices.

References

1 FAO/WHO. *Preparation and use of food-based dietary guidelines*. In: Report of a Joint FAO/WHO consultation; 1996; Niscosia, Cyprus. Geneva: WHO, 1996.
2 Freudenheim JL, Krogh V, D'Amicis A, Scaccini C, Sette S, Ferro-Luzzi A, Trevisan M. Food Sources of Nutrients in the Diet of Elderly Italians: I. Macronutrients and Lipids. *Int. J. Epidemiol.* 1993; **22**: 855–68.
3 Krogh V, Freudenheim JL, D'Amicis A, Scaccini C, Sette S, Ferro-Luzzi A, Trevisan M. Food Sources of Nutrients in the Diet of Elderly Italians: II Micronutrients. *Int. J. Epidemiol.* 1993; **22**: 869–77.
4 Greenfield H, Southgate DAT. *Food Composition Data – Production, Management and Use*. London: Elsevier Applied Science, 1992.
5 Löwik MRH, Welten DC. Quantiles and Discriminant analysis. *Pub. Health Nutr.* 2000, (this issue).
6 Becker W, Foley S, Shelley E, Gibney M. Under-reporting in dietary surveys – implications for development of food-based dietary guidelines. *Brit. J. Nutr.* 1999; **81**(S2): S139–42.
7 Lambe J, Kearney J. The influence of survey duration on estimates of food intake- relevance for food-based dietary guidelines. *Brit. J. Nutr.* 1999; **81**(S2): S139–42.
8 Serra-Majem L. National, household and individual food data. *Pub. Health Nutr.* 2000, (this issue).
9 Leclercq C, Valsta L, Turrini A. Food composition issues – implications for the development of food-based dietary guidelines. *Pub. Health Nutr.*, 2000 (this issue).
10 Corrao G, Busellu G, Torchio P, Floridi S, D'Amicis A, Recchia C, Di Placido R, Giardini P, Mandorino B, Di Orio F. Reduction modality of a dietary questionnaire. *Epidemiol. Prev.* 1993; **17**: 209–18.
11 Leclercq C, Berardi D, Sorbillo MR, Lambe J. Intake of saccharin, aspartame, acesulfame K and cyclamate in Italian teenagers. Present levels and projections. *Food Add. Cont.* 1999; **16**: 99–109.
12 Gibney MJ. Development of food-based dietary guidelines: a case-study of fibre intake in Irish women. *Brit. J. Nutr.* 1999; **81**(S2): S151–2.