Public Health Nutrition: 4(5B), 1159–1165 DOI: 10.1079/PHN2001237

A preliminary assessment of the use of household budget survey data for the prediction of individual food consumption

A Naska[1], VGS Vasdekis[2] and A Trichopoulou[1,*]
[1]Department of Hygiene and Epidemiology, Medical School, University of Athens, 75 Mikras Asias Street, Athens 115 27, Greece: [2]Department of Statistics, Athens University of Economics and Business, 76 Patission Street, Athens 104 34, Greece

Abstract

Objective: To compare individualised household budget survey (HBS) data with food consumption values derived from individual nutrition surveys (INSs).
Setting: Four European countries: Belgium, 1980–85 and 1987–88; Greece, 1993–94; Norway, 1992–94; and the United Kingdom, 1985–88.
Design: Household budget survey data were individualised with non-parametric models. Individual nutrition survey data were converted into 'HBS-like' estimates, with the application of recipe-based calculations and yield factors for weight changes during food preparation. To correct for over- and underreporting or recording in both surveys, quantities (in g day^{-1}) of 14 principal food groups were expressed as fractions of total food quantity (in g day^{-1}). For each food group, INS and HBS-derived mean values were calculated for 24 research units, jointly defined by country (four countries involved), gender (male, female) and age (younger, middle-aged and older). Pearson correlation coefficients were calculated and correlation diagrams were drawn.
Conclusion: The results of this preliminary analysis show that there is value in the nutritional information derived from HBSs. However, additional and more sophisticated work is required, in order to derive reliable point and interval estimates of individual food consumption based on HBS data.

Keywords
DAFNE
Household budget surveys
Food consumption
Nutrition monitoring

In recent decades, Europeans have experienced pronounced changes in their eating behaviour. The development of new production methods in the crop and livestock sectors of agriculture and the advancement of food science have increased the quantity and variety of food available. This increase has coincided with a shift in disease patterns towards nutrition-related conditions, such as cardiovascular disease (CVD), cancer, diabetes, obesity, osteoporosis and hypertension. CVD is the main cause of death in the European Union (EU), accounting for 42% of all deaths before the age of 74 years. Cancer accounts for 29% of all deaths in men and 22% of female deaths[1].

Against this background, monitoring food consumption patterns across Europe becomes essential, since it could form the basis of an early warning system for the formulation of nutrition policies. In the modern world of rapid change, nutrition surveillance and intervention programmes should make use of dietary surveys that have built-in mechanisms of continuity over time and extensive coverage.

Small *ad hoc* dietary surveys are undertaken in most European countries. National nutrition surveillance programmes, however, should rely on data collected in the context of nation-representative and regularly conducted individual nutrition surveys (INSs). Being expensive and labour-intensive, these surveys are undertaken only in a limited number of countries, usually those with robust economies and years of experience in the field of dietary monitoring. Furthermore variable dietary assessment methods are used, making difficult to accomplish comparability at the international level[2].

Countries with no routine information on the food consumption of their population and those interested in comparing their national dietary patterns with those of other populations have traditionally used the Food Balance Sheets (FBS) assembled by the Food and Agriculture Organisation (FAO). Since 1949, FBSs have been regularly collected on a world-wide basis and, in spite their limitations, they are often used to follow over time trends in the availability of food commodities at the population level.

Comparable between-countries information on food availability can also be provided by data collected in Household Budget Surveys (HBSs). The HBS can be thought of as occupying a position between the FBS and

*Corresponding author: Email antonia@nut.uoa.gr

the INS. Like food balance sheets, the HBSs allow between-country comparisons at a regular basis but, in moving from total population to household level, the HBS can provide a more detailed description of the dietary choices of the population, as well as of population sub-groups[3].

In Europe, there is a need for a dietary assessment tool that would provide a continuous and comparable flow of information. Urged by this need, a series of projects has been implemented aimed at the development of a cost-efficient way of using food and related data already collected in household budget surveys. These data would provide valuable complementary information to that derived from individual nutrition surveys.

The exploitation of HBS-derived data for nutritional purposes has been evaluated in the context of the DAFNE (DAta Food NEtworking) initiative. The DAFNE project has demonstrated that comparisons at the international level[4], using food and socio-economic data from national HBSs, are feasible. However, investigation of the HBS food data through comparisons with INS-generated information is required, in order to use the HBS data confidently for food monitoring trends and inter-country comparisons.

An EU project, entitled 'Compatibility of household budget and individual nutrition surveys and disparities in food habits',* aims at comparing individualised HBS data with food consumption values derived from INSs. The present paper briefly describes the methodology for rendering the two datasets comparable and presents a preliminary evaluation of their compatibility.

Material and methods

Food availability data from four European countries (Belgium, Greece, Norway and the United Kingdom) were retrieved from the DAFNE databank in order to be compared with INS data collected in these countries in corresponding time periods.

In Belgium, data from the 1987–88 HBS were compared with data collected in the Belgian Interuniversity Research on Nutrition and Health (BIRNH), conducted from 1980 to 1985 on a representative sample of the Belgian population ($n = 11\,302$) aged from 25 to 74 years[5]. In Greece, HBS data collected in 1993–94 in the greater Athens area were compared with data on food intake of 5478 residents of the Athens area, aged 27–82 years. The INS data were collected around 1994 in the context of the Greek component of the European Prospective Investigation on Cancer and Nutrition (EPIC)[6]. In Norway, HBS data of 1992, 1993 and 1994 were retrieved from the DAFNE databank and compared with the food data collected in the nation-wide NORKOST study, conducted in 1993–94 on a representative sample of 5008 adult Norwegians[7]. Finally, in the United Kingdom, data from four HBSs carried out in 1985, 1986, 1987 and 1988 were compared with the 1987–88 National Dietary and Nutritional Survey (NDNS)[8] of British adults ($n = 2197$).

Due to the diverse nature of the HBS and INS data, direct comparisons would be difficult to interpret. Modifications were therefore introduced, taking into account the characteristics of the different datasets. The HBS-generated data refer to foods available at the household level and no indication is given on what each member of the household consumes. Apart from limited and usually aggregated information on ready-to-eat meals, the food data collected in HBSs concern raw food items at the commodity level. The specially designed INSs, on the other hand, record food intake at the individual level by collecting information on the food items, mixed and recipe dishes consumed.

For the individualisation of HBS data, a non-parametric modelling approach was used. The model is based on the assumption that household food availability, during the recording period, is the sum of the food quantities available to all household members, characterised only on the basis of their age and gender. A discretisation argument transforms the model into an ordinary regression one, for which the model coefficients represent the mean individual availability, according to age and gender. The estimation of model parameters is accomplished through penalised least squares[9,10].

To bypass the inherent uncertainties in the process of rendering the two datasets comparable at the level of the dietary information collected, the INS-derived food consumption was converted into 'HBS-like' food availability. The latter data were then compared with the original HBS values. The methodology applied for converting the INS data to 'HBS-like' data is presented elsewhere[11] and proceeds with the application of:

1. yield factors, to allow for weight changes during cooking;
2. recipe calculations, to disaggregate mixed dishes and recipes into their raw ingredients;
3. edible proportion factors, to estimate the weight of the purchased food from which the consumed item was derived; and
4. a 10% reduction factor, applied in the original HBS data to allow for food wasted, spoiled or given to pets.

To prevail over differences in the classification system of the various surveys and to facilitate the between-countries interpretation of findings, all food items were classified according to the DAFNE classification system. The system allows grouping food data, expressed at commodities level, under 45 categories, which can be further classified under 14 main food groups[12]. Deviation from the DAFNE grouping system was accepted only

*The FAIR-97-3096 (Tasks 1–3) project.

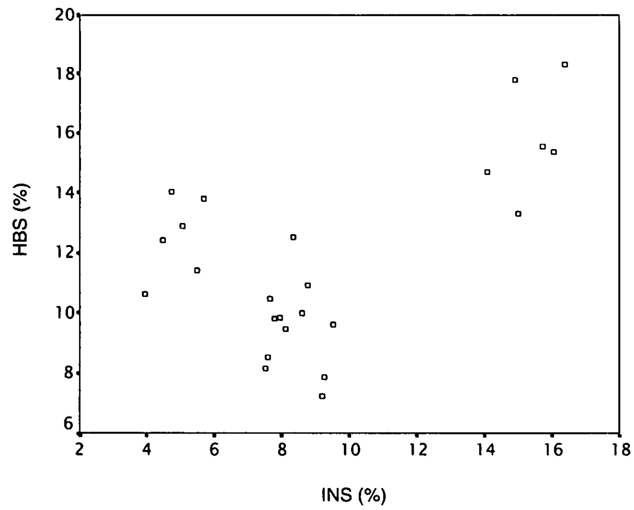

Fig. 1 Correlation diagram of the mean daily individual intake of cereals expressed as percentage of total daily dietary intake on the basis of household budget (HBS) and individual nutrition surveys (INS)

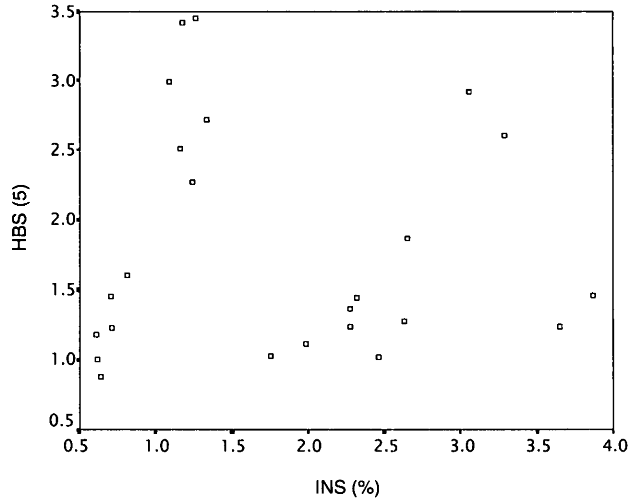

Fig. 3 Correlation diagram of the mean daily individual intake of fish and seafood expressed as percentage of total daily dietary intake on the basis of household budget (HBS) and individual nutrition surveys (INS)

when the reclassification of INS codes was impossible due to their aggregated nature.

Finally, it is common practice in interpreting food consumption data collected in the context of INSs to exclude underreporters. Underreporting can arise either by a conscious or sub-conscious failure to report everything eaten, or by a modification of usual eating habits for various reasons (e.g. feeling unwell, dieting to lose weight, attempting to report 'healthy' dietary habits). In the present analysis, misreporters, unwell persons and self-declared dieters were not excluded from the INS datasets. The expectation is that their attitudes and food choices would also be reflected in their purchases and the protocol of HBS does not allow for the identification of such bias.

Statistical procedures

This preliminary analysis relies on a few simple tasks.

- HBS data were individualised, following the procedure previously indicated.
- Because no individuals could be identified for which both HBS and INS-derived estimates would be available, correlations had to rely on research units identified by country, age and gender. Since there are four countries involved, two genders and three main

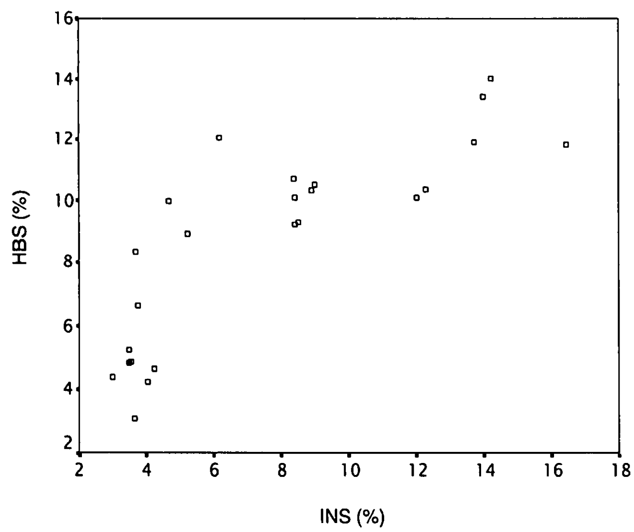

Fig. 2 Correlation diagram of the mean daily individual intake of meat and meat products expressed as percentage of total daily dietary intake on the basis of household budget (HBS) and individual nutrition surveys (INS)

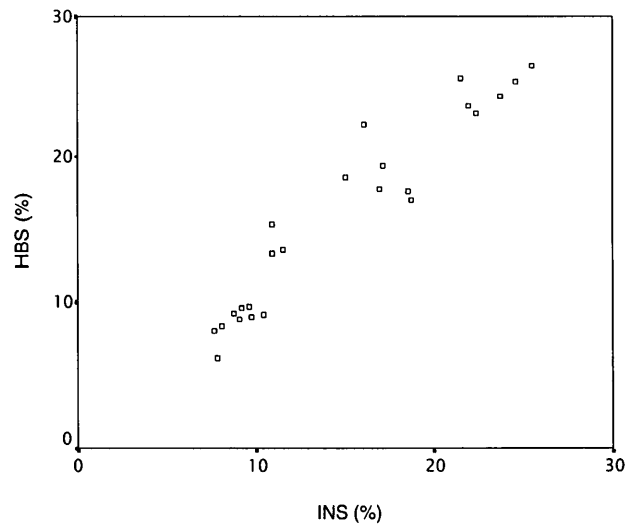

Fig. 4 Correlation diagram of the mean daily individual intake of milk and milk products expressed as percentage of total daily dietary intake on the basis of household budget (HBS) and individual nutrition surveys (INS)

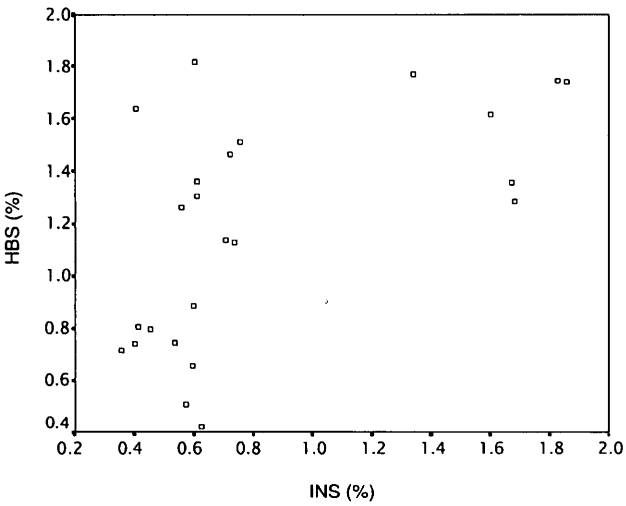

Fig. 5 Correlation diagram of the mean daily individual intake of eggs expressed as percentage of total daily dietary intake on the basis of household budget (HBS) and individual nutrition surveys (INS)

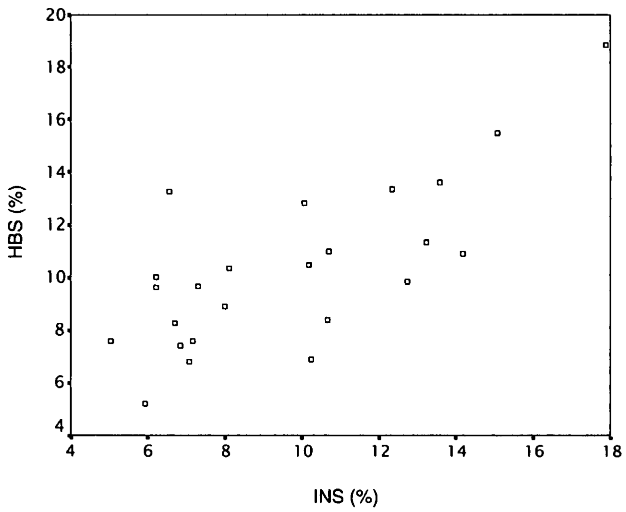

Fig. 7 Correlation diagram of the mean daily individual intake of starchy roots (potatoes) expressed as percentage of total daily dietary intake on the basis of household budget (HBS) and individual nutrition surveys (INS)

age categories (up to 45, 46–65 and 66+), 24 research units can be identified for which INS and HBS-derived mean values can be calculated. In order to consider these 24 research units as replicate measurements, it must be assumed that correlations are not differentially modified by age, gender and country, an assumption that is unlikely to hold true but also unlikely to be seriously violated.

- Nutritional validation exercises are either explicitly or implicitly energy-adjusted[13,14]. For example, it is common practice to express consumption of a particular macronutrient as a percentage of total energy intake. This process not only addresses the fact that most nutrients are associated with total energy intake, but also helps to correct for systematic across-the-board over- or underreporting and over- or underrecording[15]. When total energy intake cannot be calculated – as, for example, in some of the HBS data or in some food-frequency questionnaires limited to only a few foods – correction for over- and underreporting or recording can be attempted by expressing quantities of foods (in g day^{-1}) as fractions (percentages) of total quantity of food (in g day^{-1})[16]. Although misreporters were not excluded for reasons stated before, the above process was used by necessity in the preliminary analysis to allow for possible misreporting or recording.
- Pearson correlation coefficients can then be calculated

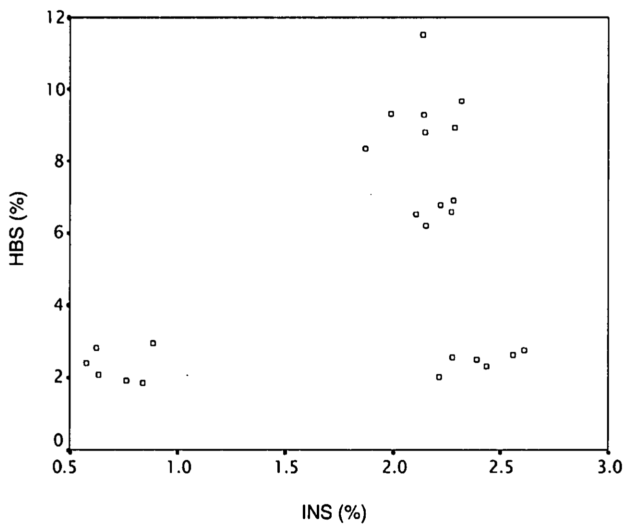

Fig. 6 Correlation diagram of the mean daily individual intake of total added lipids expressed as percentage of total daily dietary intake on the basis of household budget (HBS) and individual nutrition surveys (INS)

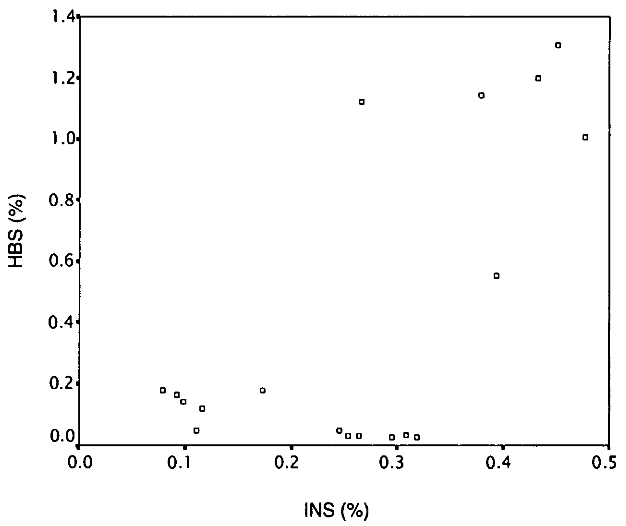

Fig. 8 Correlation diagram of the mean daily individual intake of pulses expressed as percentage of total daily dietary intake on the basis of household budget (HBS) and individual nutrition surveys (INS)

Table 1 Correlation coefficients (Pearson) of the mean daily individual values (in g) of the principal food groups as evaluated through household budget surveys and individual nutrition surveys and expressed as percentage of total daily dietary intake (in g)

Food group	Correlation coefficient*	*P*-value (two-tailed)
Cereals	0.57	0.004
Meat and meat products	0.82	$<10^{-3}$
Fish and seafood	−0.04	0.86
Milk and milk products	0.96	$<10^{-3}$
Eggs	0.58	0.003
Total added lipids	0.42	0.04
Starchy roots (potatoes)	0.74	$<10^{-3}$
Pulses	0.68	0.002
Vegetables	0.91	$<10^{-3}$
Nuts	0.44	0.03
Fruit	0.93	$<10^{-3}$
Sugar and sugar products	0.74	$<10^{-3}$
Alcoholic beverages	0.74	$<10^{-3}$
Non-alcoholic beverages	0.94	$<10^{-3}$

* Correlation coefficients for each food group derived from 24 points representing categories jointly defined by country (Belgium, Greece, Norway and the United Kingdom), gender (male, female) and age (younger, middle-aged and older).

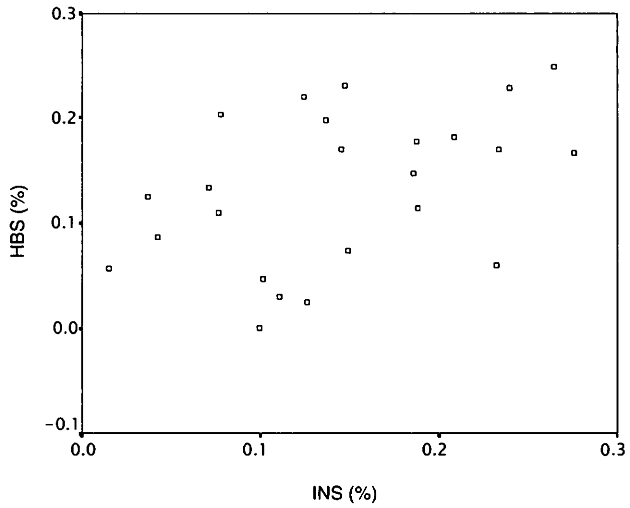

Fig. 10 Correlation diagram of the mean daily individual intake of nuts expressed as percentage of total daily dietary intake on the basis of household budget (HBS) and individual nutrition surveys (INS)

in the usual way and correlation diagrams can be drawn.

- It is obvious that prediction of individual consumption on the basis of individualised HBS data requires calculation of the fraction of any particular food group, before regression-based prediction, and then back calculation to absolute values. However, these steps are simple and obvious numerical calculations.

Results

Table 1 shows Pearson correlation coefficients for the 14 main food groups on the basis of 24 points defined by country, gender and age. With the exception of fish and seafood, correlations are good or very good and they are all statistically significant. Figures 1–14 depict the correlation diagrams for the 14 main food groups.

Discussion

These results should be considered as preliminary, but they are presented in order to demonstrate that there is value in the nutritional information derived from HBSs. The correlation coefficients may or may not be completely unbiased, but simulation procedures are required in order to identify the possible magnitude and direction of this bias. It is recognised, for instance, that large

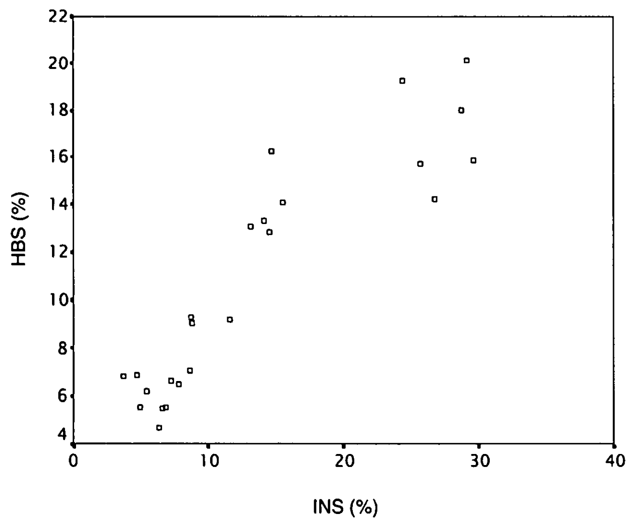

Fig. 9 Correlation diagram of the mean daily individual intake of vegetables expressed as percentage of total daily dietary intake on the basis of household budget (HBS) and individual nutrition surveys (INS)

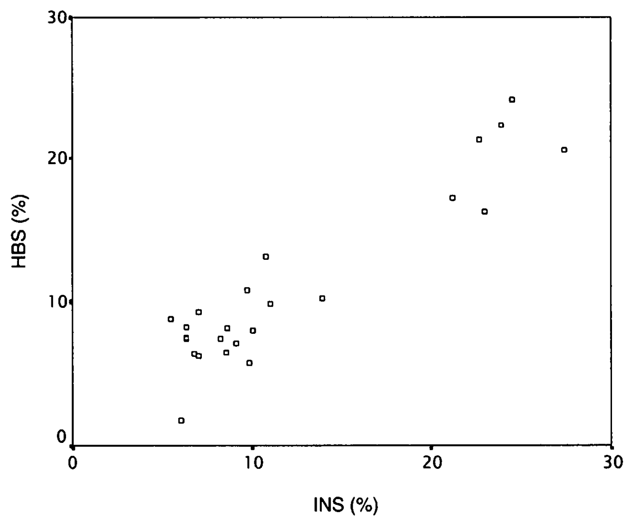

Fig. 11 Correlation diagram of the mean daily individual intake of fruit expressed as percentage of total daily dietary intake on the basis of household budget (HBS) and individual nutrition surveys (INS)

https://doi.org/10.1079/PHN2001237 Published online by Cambridge University Press

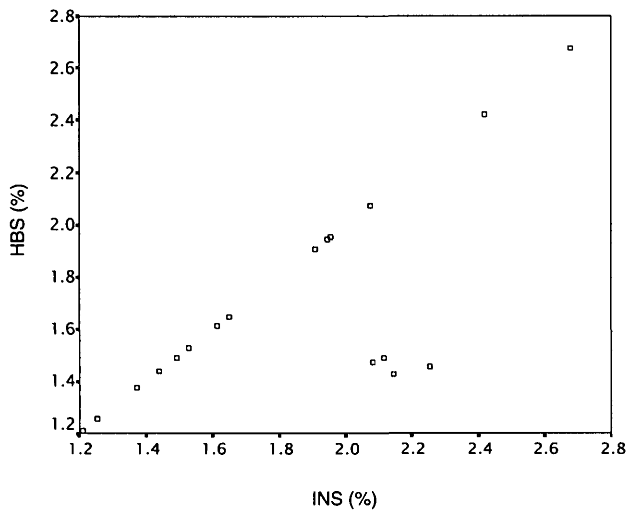

Fig. 12 Correlation diagram of the mean daily individual intake of sugar and sugar products expressed as percentage of total daily dietary intake on the basis of household budget (HBS) and individual nutrition surveys (INS)

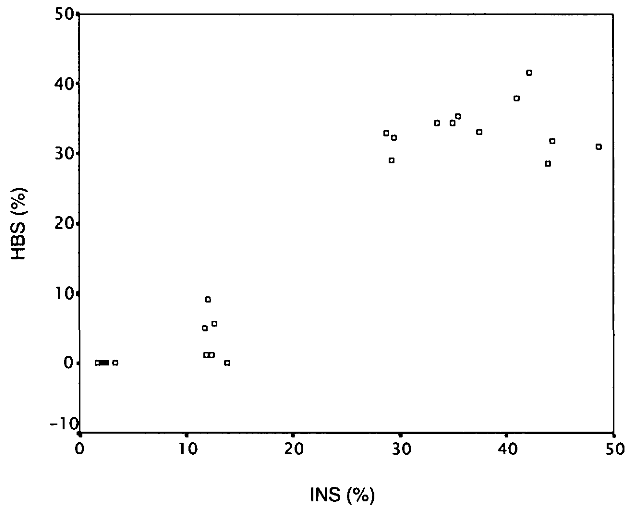

Fig. 14 Correlation diagram of the mean daily individual intake of non-alcoholic beverages expressed as percentage of total daily dietary intake on the basis of household budget (HBS) and individual nutrition surveys (INS)

systematic differences in consumption patterns of particular food groups between countries could increase the correlation coefficients, but it is not immediately obvious that this is a shortcoming of the approach. Indeed, discrimination of sharply different consumption patterns should be the primary objective, when individualised HBS data are used to predict individual nutrition consumption that is generally considered as the gold standard.

It is tempting to try to compare correlation coefficients between various food groups (for example, over 0.90 for fruits and vegetables but less than 0.70 for pulses or cereals). However, confidence intervals for correlation coefficients relying on 24 points are generally large and significance levels refer only to differences from the null value of no correlation. Instead, the emphasis should be on the general overall satisfactory pattern of positive correlation between HBS and INS-derived estimates. Additional and more sophisticated work is required, however, in order to derive reliable individual consumption point and interval estimates on the basis of HBS data.

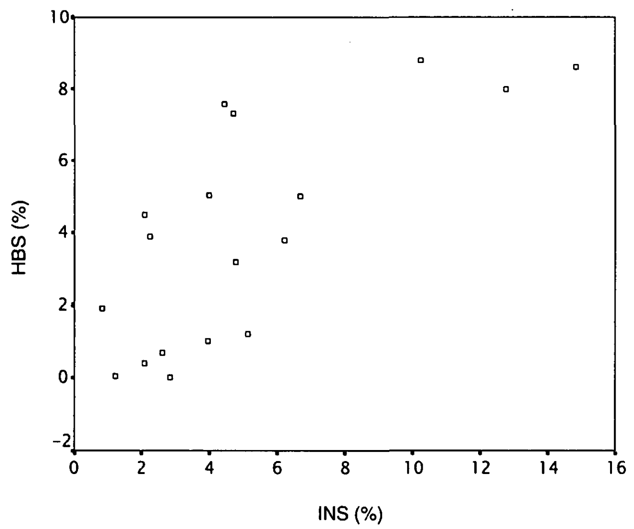

Fig. 13 Correlation diagram of the mean daily individual intake of alcoholic beverages expressed as percentage of total daily dietary intake on the basis of household budget (HBS) and individual nutrition surveys (INS)

The individual-based food consumption surveys, when undertaken as adequately as possible, could represent the optimal assessment of the eating habits of a population. However, the cost for contemplating a truly international system to monitor dietary intakes at the individual level seems unrealistic. Preliminary results of the present analysis show that there is considerable scope in using HBS data to achieve an average estimate of the populations' food habits and to run international comparisons.

Notification

The British HBS data are Crown copyright. They were made available by the Office for National Statistics (ONS) through the Data Archive, based in the University of Essex. Neither the ONS nor the Data Archive bears any responsibility for the analysis or interpretation of the data reported here.

Material from the Dietary and Nutritional Survey of British Adults, made available through the Office of Population Censuses and Surveys, the Ministry of Agriculture, Fisheries and Food, the Department of Health and the ESRC Data Archive, has been used by permission of the Controller of Her Majesty's Stationery Office. Those who carried out the original analysis and collection of the data bear no responsibility for the further analysis or interpretation of it.

Acknowledgement

The work presented in this paper was undertaken in the context of the FAIR-97-3096 project (Tasks 1–3) of the European Union.

References

1 Lang T. Food and nutrition. In: Weil O, McKee M, Brodin M, Oberlé D, eds. *Priorities for Public Health Action in the European Union*. European Commission, Employment and Social Affairs, 1999; 138–56.

2 Friedenreich CM. Methodologic issues for pooling dietary data. *Am. J. Clin. Nutr.* 1994; **59**(Suppl. 1): 251S–2S.

3 Trichopoulou A. Monitoring food intake in Europe: a food data bank based on household budget surveys. *Eur. J. Clin. Nutr.* 1992; **46**(Suppl. 5): S3–8.

4 Byrd-Bredbenner C, Lagiou P, Trichopoulou A. A comparison of household food availability in 11 countries. *J. Hum. Nutr. Diet.* 2000; **13**: 197–204.

5 De Henauw S, De Backer G. Nutrient and food intake in selected subgroups of Belgian adults. *Br. J. Nutr.* 1999; **81**(Suppl.): S37–42.

6 Riboli E. Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann. Oncol.* 1992; **3**: 783–91.

7 Johansson L, Solvoll K, Bjorneboe GE, Drevon CA. Dietary habits among Norwegian men and women. *Scand. J. Nutr.* 1997; **41**: 63–70.

8 Gregory J, *et al. Dietary and Nutritional Survey of British Adults*. London: HMSO, 1990.

9 Chesher A. Diet revealed? Semiparametric estimation of nutrient intake–age relationships. *J. Roy. Statistic. Soc. A* 1997; **160**: 389–428.

10 Vasdekis VGS, Stylianou S, Naska A. Estimation of age- and gender-specific food availability from household budget survey data. *Public Health Nutr.* 2001; **4**: 1147–9.

11 Naska A, Paterakis S, Eeckman H, Remaut AM, Trygg K. Methodology for rendering household budget and individual nutrition surveys comparable, at the level of the dietary information collected. *Public Health Nutr.* 2001; **4**: 1151–6.

12 Lagiou P, Trichopoulou A and the DAFNE contributors. The DAFNE initiative: the methodology for assessing dietary patterns across Europe using household budget survey data. *Public Health Nutr.* 2001; **4**: 1133–9.

13 Willett WC. *Nutritional Epidemiology*. New York: Oxford University Press, 1998.

14 Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens H, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am. J. Epidemiol.* 1985; **122**: 51–65.

15 MacMahon B, Trichopoulos D. *Epidemiology. Principle and Methods,* 2nd ed. Boston, MA: Little, Brown and Company, 1996; pp. 242

16 Petridou E, Malamou H, Doxiadis S, Pantelakis S, Kanellopoulou G, Toupadaki N, Trichopoulou A, Flytzani V, Trichopoulos D. Blood lipids in Greek adolescents and their relation to diet, obesity, and socioeconomic factors. *Ann. Epidemiol.* 1995; **5**: 286–91.