A Data Entry System for Dietary Surveys Based on Visual Basic for Applications Programming

CrossMark

COMPUTER-BASED INTAKE collection and nutrient analysis systems have been used to manage dietary intake information obtained from large groups of individuals.[1-4] Some can automatically classify foods into food groups.[3,4] However, these diet analysis systems were largely designed to collect short-term intake data by food record and 24-hour dietary recall. When used with a standardized diet assessment protocol, bias and incomplete information can be minimized.[1,5]

Few systems used for collecting information on food intake contain an extensive list of cultural and local food from around the world, nor do they allow documentation of unusual eating habits. In the United States, the Nutrition Data System for Research (NDSR) is a dietary data collection software program developed primarily for research by the Nutrition Coordinating Center, University of Minnesota. It includes an extensive list of foods, recipes, and nutrients, and it can accommodate recipes specific to individuals.[6,7] NDSR uses the standardized dietary data collection protocol, the automated multi-pass

method, which guides the process of the 24-hour recall with five sequential steps to reduce incomplete information. The steps are as follows: 1) asking the participants to list all of the food items consumed in the previous day; 2) asking the participants for potentially forgotten food items by listing the nine most frequently neglected foods; 3) probing for the time and occasion in which each food item was consumed; 4) probing for detailed information on each food item; and 5) probing for potentially forgotten details of each food item.[1,2,5] In Europe, the EPIC-SOFT software was developed by the EPIC Software Corporation for the European Prospective Investigation on Cancer (EPIC) study to collect and analyze food intake, recipes, nutrient intake, and food groups assessed with the 24-hour dietary recall or the food record.[3] In Brazil, a dietary data entry system and nutrient composition tables were developed for the Brazilian Household Budget Survey (HBS), a large periodic survey starting in 2000.[8] In 2009, the evaluation of 7 days of individual dietary intake was initiated by using the food record method, in addition to the HBS method used in previous surveys.[8] The HBS method documents household expenses for food and beverages in a given period, but it does not collect information on how foods were prepared and consumed, what foods were eaten out of the home, and which foods were consumed by each family member.[9]

Each aforementioned system has built-in nutrient composition tables and serving portions; furthermore, some systems allow for insertions of customized recipes that consider specific geographical areas, dietary practices, and foods. In countries in which the diet analysis software does not have a comprehensive list of local food items, recipes, and serving sizes, users of these systems must rely on nutrient composition and recipe

information from other regions or countries, assuming that the dietary information from other countries is similar to the local information. Therefore, the creation of dietary data collection and analysis software that can be used and adapted easily by users from any region in the world would facilitate collection and analysis of dietary data in many parts of the world. For a borderless dietary data system to be practical to use in situations with few resources, it needs to be a quick, versatile, and low-cost system.

This article describes the development of a dietary data entry system for research that is user-friendly, multipurpose, and adaptable to different lists of food items, recipes, and serving portions for any region of the world, based on the Visual Basic for Application (VBA) in Excel (version 2010, Microsoft).

RATIONALE AND DESIGN

In situations in which available commercial software is not suitable for specific population settings and the nutrient composition of local foods must be manually entered to include local food information, an alternative system for data entry for dietary surveys is needed. Without an alternative system, dietary data entry of population surveys involves repetitive entry of food items and recipes, time-consuming calculations of nutrient and serving size amounts, and matching dietary with nondietary data.

The DietSys system is a surrogate application developed by Sinara L. Rossato, MSc, PhD, a dietitian from Brazil, who has experienced the manual dietary data entry as part of her studies for masters and doctoral degrees.[10-12] The DietSys was created to facilitate incoming dietary data entry and is designed for entering intake information collected with 24-hour recall or food record methods. It also can collect nondietary information along with dietary information. DietSys also enables the calculation of the

This article was written by **Sinara L. Rossato,** MSc, PhD, a research fellow, Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA, and a research fellow, School of Medicine/Department of Nursing, Universidade do Estado de Sao Paulo Julio Mesquita Filho, UNESP, Botucatu, Brazil; **Teresa T. Fung,** ScD, RD, adjunct professor of nutrition, Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA, and associate professor of nutrition, Department of Nutrition, Simmons College, Boston, MA; and **Marcela P. Rodrigues,** MSc, PhD student, Universidade Federal do Rio Grande do Sul—UFRGS, Porto Alegre, Brazil.

Table. Computational procedures involved in a manual data-entry and nutrient intake assessment when specific software are not suitable for a diet survey; number of files created for a diet survey data entry; number of spreadsheet in the file; and time consumed using the DietSys and the manual system[a]

Procedure	File	Spreadsheet	Time in Months DietSys	Manual system
Enter nondietary information	A	1	0.4	0.5
Enter individual dietary information	B	1	2.0	1.4
Review records for accuracy	B	1	0.7[b]	1.3[b]
Simplify the multiple food items spellings	B	1	0[c]	0.5
Encode food item for servings table	B	1	0[d]	1.8
Convert serving portions into weight and volume	B	2	0[d]	0[e]
Check for inconsistency by revising outlier information	B	1	0[d]	1.1[b]
Break down recipes into ingredients	C	1	0[d]	2.0
Revise every new recipe and check for inconsistency	C	1	0[d]	0.7[b]
Insert recipes in the original file	B+C	1	0[d]	1.9
Classify food items into food groups (optional)	B	1	0[c]	0.5
Estimate nutrient composition by encoding every food item	B	3	0[d]	2.0
Insert lookup function to search nutrient information in other spreadsheets	B	1	0[d]	0[e]
Check for inconsistency in nutrient contents	B	1	0.9[b]	0.9[b]
Create a pivot table to arrange information in a nested scheme	B	4	0[e]	0[e]
Merge nondietary and dietary information	B+A	5	0[d]	0.1
Export to the statistical software	D	6	0[e]	0[e]
Total time consumed in months			4.0	14.6

[a]The time was estimated based on previous data collection/entry in surveys of 130 to 160 individuals who had the diet assessed by applying four 24-hour dietary recalls,[10-12] considering that one investigator worked around 20 days/mo, 5 h/day in data entry and nutrient intake estimation.

[b]Final procedure consists of revising the data entered and comparing the database with the original questionnaires for capture errors and inconsistencies as a quality control procedure.

[c]The system had the food items and the food groups standardized beforehand.

[d]The system leads to the completion of this step automatically during the data entering.

[e]Less than 1 hour is required to execute these procedures.

nutrient composition of diet surveys, using local nutrient composition, serving size, and recipe tables. Using Microsoft Excel resources, the DietSys was programmed to merge both the dietary and nondietary information at the time of data entry, and the data bank was arranged to facilitate the calculations and statistical analysis.

DietSys may be an alternative approach for registered dietitian nutritionists involved in data collection and analysis. In previous studies on populations of healthy adults and adolescents in Brazil,[10,11] data were manually entered from the 24-hour recall into Excel, using the steps described in the Table.

DietSys reduces the number of steps in the Table from 17 to the following five:

1. Enter data.
2. Review inconsistency and revise the data entered.
3. Insert the LOOKUP function to bring from the database the information of interest, such as grams or milliliters, and energy or nutrients.
4. Create a pivot table to rearrange data for calculation, such as organize the data per participant, per day, or per meal.
5. Export food and nutrient data to statistical software for analysis.

VBA was used to address three issues involved in the rationale of this system:

1. to manage repetitive data entry, such as to insert one food group or item for several participants using different options for serving sizes, and to identify erroneous calculation of recipe ingredients by allowing the user to review the recipe before recording in the databank;
2. to compel users to interact with different forms and arrangements of foods, recipes, and serving sizes from the respective tables; and
3. to execute procedures that are not available in Excel, such as recording and formatting the data in the data bank, converting serving sizes to amounts of food items or ingredients, revising ingredient proportions from recipes before depositing the data in the data bank, and filling out empty data cells in the final database after data were recorded if information was missing.

The VBA programming is a built-in resource for Microsoft Office software that consists of added codes that execute specific procedures not originally built into the Excel software. To enable the user to activate the procedures, one may insert new buttons linked to sequences of codes. For the DietSys system, we added a few new buttons to trigger specific tasks. For example, the button "Record data in
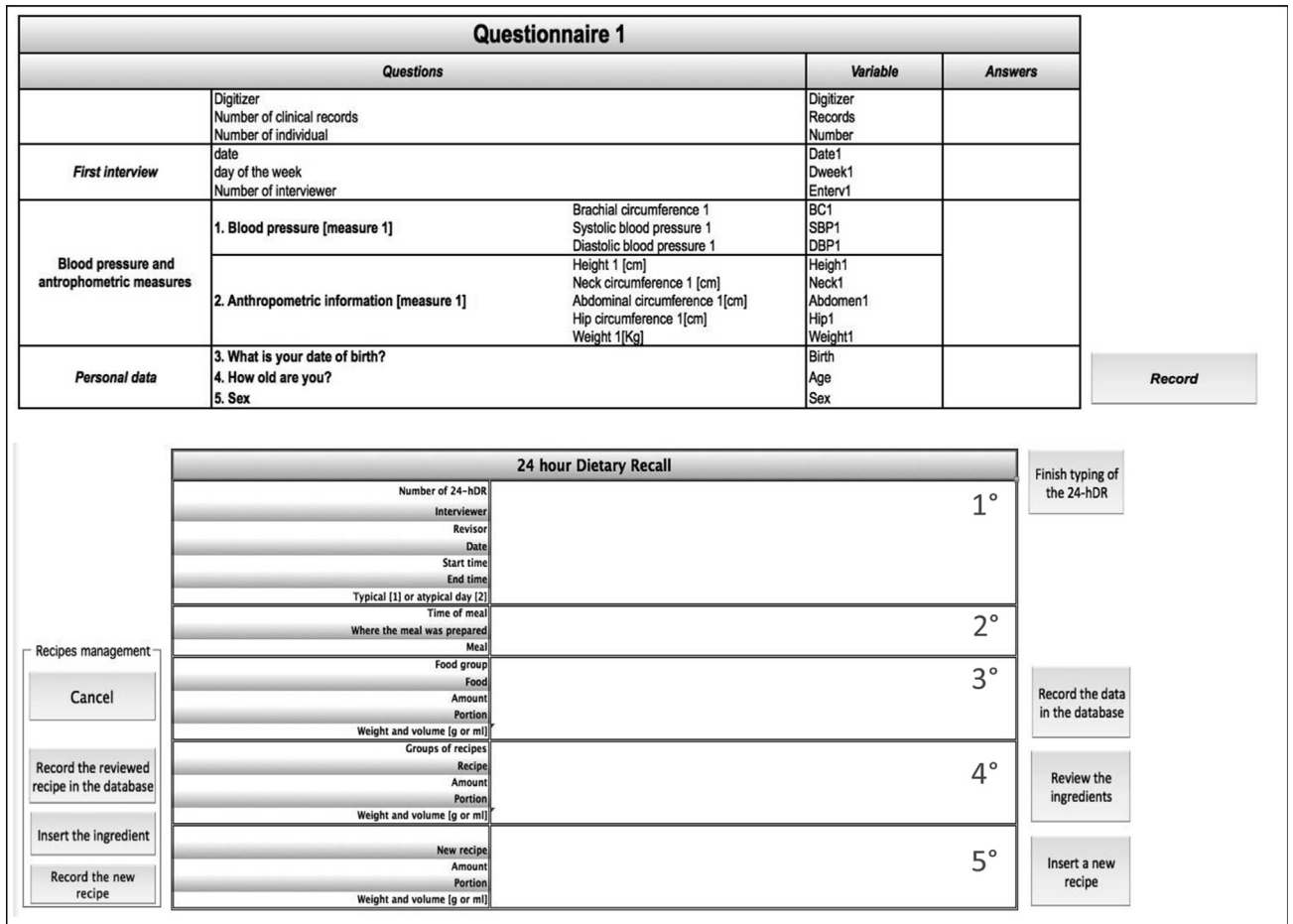
Figure 1. Spreadsheet illustrating the questionnaire one, out of three, for entering of nondietary information, containing examples of variables.

the databank" is intended to be used after the user selected the food group, food item, and portion size, to record the data in the data bank. Another example is the "Review the recipe" button, which allows the user to verify the proportions of each ingredient and decide whether the recipe represents the participant's report or a new recipe must be included.

The pivot table is a built-in tool in Excel that contains four areas, which are displayed in a task pane: 1) filter area, 2) column area, 3) rows area, and 4) values area. The pivot table can be used to filter, calculate, and summarize large amounts of information. In DietSys, when the data entry is finished, the user can use the pivot table to calculate the total amount of food groups by dragging the participant identity number and the food group variable to the column area and the amount in grams to the rows area. Sequentially, the user

will select which calculation function will be of interest. For example, the "count" function will generate the frequency of consumption of each food group, the "mean" function will compute the average daily amount (in grams) of each food group consumed; and the "sum" function will calculate the total intake of each food group. The same procedure can be used to calculate total energy and nutrients intake or total intake of specific food items.

The data validation, another built-in Excel feature, was used to control what can be entered in a worksheet. It creates a drop-down list of items in a cell that restricts entries, such as a date range or whole numbers only, and creates custom rules for what can be entered. When the restrictions or rules are not followed, a pop-up box with an error message was programmed to appear. We used these features for quality control purposes, informing the

data entry personnel when errors were made (eg, when an inconsistent amount of portions was entered, an inconsistent date or hour was typed, or when information was typed in cells used exclusively for calculation purposes).

SYSTEM STRUCTURE
To illustrate the system, we used a simple research scenario that involves using three questionnaires to collect data from each study participant and one 24-hour dietary recall form. Each interview includes one nondietary information questionnaire and one 24-hour dietary recall. In total, three interviews were conducted, resulting in three nondietary questionnaires and three 24-hour dietary recalls. All three questionnaires (Q1, Q2, and Q3) were created only for illustrating the system operation; therefore, they are not related to the objective of the dietary

intake analysis (Figure 1). For instance, the first questionnaire (Q1) included information related to the interviews, participant record numbers, personal information, and biological and anthropometric measures, such as blood pressure and waist, hip, and neck circumferences. The researcher can include different variables according to the aims of the research project. Based on this structure, completion of the questionnaires follows this sequence: administer questionnaire 1 → record the data → 24-hour dietary recall → finish the data entry for interview 1 → administer questionnaire 2 → record the data → 24-hour dietary recall → finish the data entry for interview 2, and so forth.

The DietSys system was created to match dietary and nondietary information in a long-form structure, to include ingredients from recipes, and to allow for insertion of local food information such as the Brazilian Food Composition Table,[11] serving sizes and portions, and recipes. In this regard, DietSys can be adapted for different populations. For example, if a diet survey is conducted in Russia, the research team can substitute the Brazilian composition table, serving sizes, and recipes with the Russian food composition table, serving sizes, and recipes to make the analysis more accurate for the Russian population. To adapt the system for using a specific database, the researcher can add

information in three hidden spreadsheets: the first for the food composition table, the second for portion sizes, and the third for recipes.

To facilitate the data entry and nutrient content analysis, our system used the design described in the following sections, based on the system's structure illustrated in Figure 2.

ADVANTAGES OF DietSys

DietSys has two particular advantages in comparison with other data entry systems: 1) the databases used to calculate nutrient composition, recipes, and serving sizes are adaptable for different countries and regions; and 2) DietSys also merges nondietary with
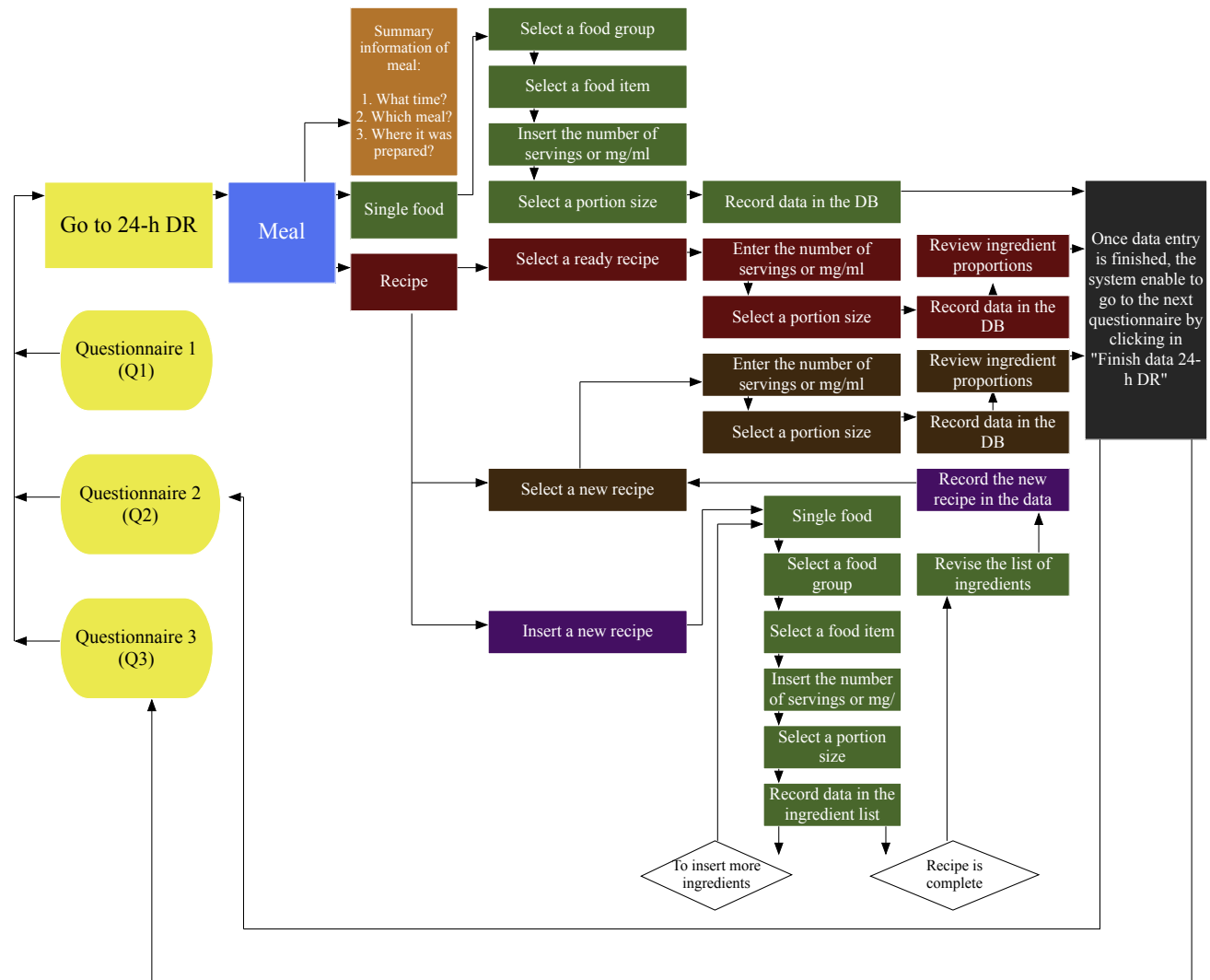
Figure 2. Spreadsheet containing the 24-hour dietary recall form for entering dietary information. 24h-DR=24-hour dietary recall; DB=data bank.

dietary information as the data are entered, which eliminates the need to merge data after they have been entered in different data banks, thus optimizing data quality and time.

Moreover, because DietSys lists foods, recipes, and servings for data entry, the system increased data quality by avoiding typographical errors and erroneous entries. DietSys allows for adding new recipes, local food items, and serving size information.

When compared with previous studies in which data were manually entered,[10-12] DietSys was much faster. In a study involving 120 to 150 individuals and three[10-12] to six[10] 24-hour dietary recalls per person, the manual data-entry method required an average of 14 months in 17 steps (Table). When data entry was done with DietSys, the process took less than 4 months and only required six steps. The nutrient composition table in DietSys could be customized by adding the nutrient data of foods that are not already in the database. In particular, specific cultural practices and locally produced foods could be recorded to facilitate research needs. As with many available diet analysis systems, the system was developed specifically for particular countries and regions and therefore could not be easily used outside those regions.

The current version of DietSys contains almost 1,300 foods and beverages and 577 recipes involving local Brazilian food items. It could easily be expanded to reach a similar number of food items as that of other databases.[3,5] From this perspective, DietSys can be expanded into an international dietary data system. By compiling information from different geographical areas, DietSys provides the opportunity for multicenter studies with greater precision of local information.

DietSys also contains functions that are available in established diet analysis software. Similar to Epic-Soft, in which recipe ingredients are presented as percentages of the overall weight of the recipe in grams,[13] DietSys also presents recipe ingredients as a percentage of the overall weight of the recipe, with the additional advantage that DietSys enables the insertion of local recipes.

Although nutrient analysis software from different countries often contains foods with the same name, preparation and content of nutrients and ingredients may differ significantly. For example, according to the US Department of Agriculture table, 100 g turkey breast has 90.0 kilocalories, 19.1 g protein, 0.7 g lipids, 1.9 g carbohydrates, 40.0 mg cholesterol, and 1,177.0 mg sodium.[14] In the Brazilian food composition table, 100 g turkey breast was listed with 128.0 kilocalories, 14.4 g protein, 6.8 g lipids, 1.4 g carbohydrates, 40.0 mg cholesterol, and 1,021 mg sodium.[13] Unsalted cooked black beans, a typical Brazilian dish, also varied in nutritional composition between the two tables. In the Brazilian nutrient composition table, 100 g cooked black beans were composed of 77 kilocalories, 4.5 g protein, 0.5 g lipids, 14.0 g carbohydrates, 8.4 g fiber, and 2.0 mg sodium,[14] whereas in the US Department of Agriculture table, 100 g unsalted cooked black beans had 132 kilocalories, 8.86 g protein, 0.54 g lipids, 23.71 g carbohydrates, 8.7 g fiber, and 1.0 mg sodium.[14] The recipes calculated in the current version of DietSys used a Brazilian table of recipes and household measurements as reference.[10]

The same functionalities were employed in other systems; for instance, EPIC-SOFT calculated predefined standard values of foods that were added in small quantities (such as grated cheese and nuts), and it also automatically calculated the quantities of fat used during cooking through standardized percentages.[13] FINDIET applied local recipe books as a reference for calculating recipes, thereby standardizing the quantity of salt and other reported seasonings,[15] which can be performed by DietSys, because the local recipes can be inserted as a database.

Validated and standardized data collection and data entry procedures ensure the accuracy of information. In NDSR, this is accomplished by using the automated multi-pass method for data collection.[5] DietSys also has built-in mechanisms to avoid unintended errors. The first mechanism involves boxes of information regarding the data validation function that appear during data entry. The second enables the selection of food groups, food items, recipes, and portion sizes from a list, preventing the user from making spelling and typographical errors. This can become an issue when entering several food items with similar terminology (eg, sugared coffee, coffee

sugared, black coffee with sugar, sugared black coffee), which eventually would demand further standardization to reduce the list of food items.

Another strength of DietSys in comparison with other software is the flexibility to modify the system structure to adapt to different research purposes. However, modifying the DietSys structure requires expertise from the researcher in two areas: 1) inserting food item information and breaking down recipes, which involves the hidden spreadsheets; and 2) knowledge of local foods and recipes in the absence of detailed information in data collection. Even though the creation of DietSys does not require advanced computing expertise, an experienced VBA programmer is preferred to make additional improvements in the structure, set of commands, and graphic layout.

CONCLUSION

DietSys may be a feasible software program for dietary data entry when suitable software is not available for a specific population. Using DietSys reduced the amount of time needed for data entry. It can maintain data quality by preventing typographical errors and merging dietary with nondietary questionnaires in one data bank at the time of data entry.

URL link for download: https://www.dropbox.com/s/blh6gimjsnu1s68/DietaryDataSet-2014.xlsm?dl=0

Steps for downloading and opening the file: Click on the link click on the 'download' or 'open' button at the right corner of the window → open the Excel file → enable edition → enable content.

References

1. Baranowski T. 24-Hour recall and diet record methods. In: Willett WC, ed. *Nutritional Epidemiology*. 3rd ed. Oxford University Press; 2013:50.
2. Blanton CA, Moshfegh AJ, Baer DJ. The USDA multi-pass method accurately estimates group total energy and nutrient intake. *J Nutr*. 2006;136(10):2594-2599.
3. Slimani N, Deharveng G, Charrondiere RU, et al. Structure of the standardized computerized 24-h diet recall interview used as reference method in the 22 centers participating in the EPIC project. European Prospective Investigation into Cancer and Nutrition. *Comput Methods Prog Biomed*. 1999;58:251-266.
4. Feskanich D, Sielaff B, Chong K, Bartsch G. Computerized collection and analysis of dietary intake information. *Comput Methods Prog Biomed*. 1989;30:47-57.
5. NDSR—Nutrition Data System for Research. http://www.ncc.umn.edu. Accessed June 15, 2014.
6. Schakel SF, Sievert YA, Buzzard IM. Sources of data for developing and maintaining a nutrient database. *J Am Diet Assoc*. 1988;88(10):1268-1271.
7. Dennis B, Ernst N, Hjortland M, Tillotson J, Grambsch V. The NHLBI nutrition data system. *J Am Diet Assoc*. 1980;77(6):641-647.
8. Sichiery R, Pereira RA, Martins A, Vasconcellos ABPA, Trichopoulou A. Rationale, design, and analysis of combined Brazilian household budget survey and food intake individual data. *BMC Public Health*. 2008;8:89-94.
9. Moubarach JC, Parra DC, Cannon G, Monteiro CA. Food classification systems based on food processing: Significance and implications for policies and actions: A systematic literature review and assessment. *Curr Obes Rep*. 2014;3:256-272.
10. Rossato SL, Olinto MTA, Henn RL, Anjos LA, Bressan AW, Wahrlich V. Seasonal effect on nutrient intake in adults living in Southern Brazil. *Cad Saúde Pública*. 2010;26(11):2177-2187.
11. Rodrigues MP. Avaliação da associação de consumo de feijão com arroz e pressão arterial em indivíduos hipertensos em tratamento. Master's thesis UFRGS—Universidade Federal do Rio grande do Sul; 2014.
12. Rossato SL, Olinto MTA, Henn RL, et al. Seasonal variation in food intake and the interaction effects of sex and age among adults in southern Brazil. *Eur J Clin Nutr*. 2015;1:1-8.
13. Reinivuo H, Valsta LM, Laatikainen T, Tuomilehto J, Pietinen P. Sodium in the Finnish diet: II trends in dietary sodium intake and comparison between intake and 24h excretion of sodium. *Eur J Clin Nutr*. 2006;60:1160-1167.
14. Hakala P, Knuts L-D, Vuorinen A, Hammar N, Becker W. Comparison of nutrient intake data calculated on the basis of two different databases: Results and experiences from a Swedish-Finnish study. *Eur J Clin Nutr*. 2003;57:1035-1044.
15. Reinivuo H, Hirvonen T, Ovaskainen ML, Korhonen T, Valsta LM. Dietary survey methodology of FINDIET 2007 with a risk assessment perspective. *Public Health Nutr*. 2010;13(6A):915-919.