Use of the South African Food Composition Database System (SAFOODS) and its products in assessing dietary intake data: Part II

[a] Wolmarans P, [b] Kunneke E, [c] Laubscher R

[a] Nutritional Intervention Research Unit, Medical Research Council [b] Division Dietetics, University of the Western Cape [c] Biostatistics Unit, Medical Research Council

Correspondence to: Dr Petro Wolmarans, e-mail: petro.wolmarans@mrc.ac.za

Keywords: South African Food Composition Database

Abstract

A number of dietary research methodologies are available for the collection of quantitative dietary intake data. The methods most often used in South Africa include the 24-hour dietary recall, the quantitative food frequency questionnaire, and the dietary intake record. To quantify dietary intake, the information typically required includes the energy and nutrient composition of foods commonly consumed in South Africa. The South African Food Composition Database System (SAFOODS) and its products, e.g. the printed tables and software program FoodFinder3, provide the researcher with the tools to convert food intake data into energy and nutrient(s) intake. FoodFinder3 can be used for the nutrient analysis of the data. It also enables the user to export the data to MS Excel for further analysis and for importing the data into other statistical packages. Coding for the type and quantity of food consumed is required however before the data can be electronically analysed. The Food Quantities Manual of the Medical Research Council provides the necessary information for the conversion of food intake data recorded in household measures into grams of food. To ensure that the quality of the dataset is high, several steps have to be undertaken before statistical analysis and reporting of the data can take place. Appropriate statistical methods are required for the analysis of the data as nutrient intake data are often skewed. Using a standardised protocol, validated questionnaires and the South African Food Composition Database (SAFOOD) for the analysis of dietary intake data could make the pooling of data from small scale or regional studies possible. This may impart an impression of energy and nutrient intake at the national level, and could, at least in part, compensate for the absence of regular national surveys.

S Afr J Clin Nutr 2009;22(2):59-67

Introduction

The collection and analysis of dietary intake data in a country is essential for establishing dietary inadequacies/excesses as well as the role of diet in the prevalence and development of diet-related health risks, which in turn impact on policy formulation for the country. To assess the dietary profile of a community appropriate methodologies need to be applied and dietary tools need to be used for the nutrient analysis of the diet. In Part I of this review series on the South African Food Composition Database System (SAFOODS) the characteristics of the food composition database were described.[1] In this second part of the review, the different methodologies that can be used for the collection of dietary intake data will be discussed, and how the products, e.g. the South African Food Composition Tables (SAFCT) and the software program FoodFinder3, which includes the South African Food Composition Database, can be used for the analysis of the dietary intake data.[2] Reference will also be made to selected appropriate statistical methods for the analysis of dietary intake data.

Collection of dietary intake data

Dietary research methodology

The aim of the study and the study design determine the methodology most suitable for the collection of the dietary intake data in a given setting. A number of different methodologies can be used for the collection of dietary intake data.[3] The advantages and disadvantages of the different methods are discussed in Gibson,[3] and summarised by Wolmarans and Wentzel-Viljoen.[4] The methodologies most often used in South Africa include the 24-hour recall, the qualitative (non-quantified) or quantitative (quantified) food frequency questionnaire and the weighed or estimated dietary intake record. These dietary assessment methodologies can also be used to determine other food constituents, such as food contaminants or polyphenols, or in a clinical setting to determine the dietary intake of clients. In this paper we focus only on the use of the different dietary assessment methodologies used for the collection of quantitative data in order to report on energy and nutrient intakes.

The 24-hour recall method

If a study requires quantitative data on nutrient intakes but time constraints are an issue and large numbers of the population are studied then the 24-hour recall methodology is often the method of choice. In South Africa, the 24-hour recall method has been used for the collection of dietary intake data in several large epidemiological studies to determine dietary intakes of the population.[5–7] For instance, in 1999 the National Food Consumption Survey (NFCS) used the 24-hour recall as one of the methods for the collection of dietary intake data.[8] In such studies the sample sizes are usually large, about 1 000 or more participants. It is crucial to remember that such studies should include dietary information for the weekdays as well as weekend days. An even distribution of all the days of the week is the ideal, but this is not easy to achieve since it is not always possible to collect data for Fridays and Saturdays. These days are often very important days of the week, as people tend to change their dietary intakes over the weekend. The training and standardisation of the fieldworkers in the use of the 24-hour dietary recall forms an integral part of the success of the collection of the data.

An example of a partly completed 24-hour recall sheet (Table I) could serve as a guideline for researchers. The headings of the columns of the form may have to be changed, as appropriate, to meet specific requirements. Moreover, a manual compiled by the researchers for a specific study is crucial in the training of the fieldworkers and it can also be used as a reference by fieldworkers during the collection of the dietary intake data. Another tool for the collection of dietary intake data is the Dietary Assessment Kit, compiled by Steyn and Senekal.[9] This tool includes photographs of commonly consumed foods, information on the food quantities, guidelines for the collection of dietary intake data, as well as a sample questionnaire.

Once the data have been collected, software programs containing energy and nutrient composition data of the food(s) eaten can be used for the analysis of the dietary intake data.

Table I: An example of a completed 24-hour dietary recall sheet for the collection of dietary intake data

Time that food was eaten*	What did you eat/ drink?	How was it prepared?	How much did you eat/drink?
08:00	Tea, Ceylon		180 mL
08:00	Milk, full cream		50 mL
08:00	Sugar, white		1 heaped teaspoon
08:00	Bread, brown	Shop sliced	2 slices
08:00	Margarine, hard brick	Brand name	2 level teaspoons
10:00	Apple, with skin		1 (7 cm diameter x 5 cm height)
12:00	Chocolate	100 g slab (brand name)	Five blocks
13:00	Chicken, breast without skin	Fried in sunflower oil	10 cm X 6 cm X 2 cm
13:00	Rice, white	Steamed	Half a cup
13:00	Chicken gravy	With fat	2 tablespoons

*Continue with the food intake recall for the rest of the day and cross-check at the end for quality control purposes.

Food frequency questionnaire

Dietary intake data can either be collected by means of a qualitative (Ql) or quantitative (Qn) food frequency questionnaire (FFQ). If the aim of the study requires that information on nutrient intakes should be reported in absolute values, then information has to be collected not only on the type of food consumed, but also on the quantity consumed. In the Transition, Health and Urbanization in South Africa Study (THUSA), as well as in the NFCS, a QnFFQ was used.[8,10] The collection of dietary intake data by means of a QnFFQ is a much more time consuming undertaking than the collection of dietary intake data by means of a 24-hour recall. An advantage of the QnFFQ, however, is that information is collected on the frequency of the consumption of food items and it provides an indication of habitual dietary intake. The interviewer who collects information by means of a QnFFQ needs to be well trained and experienced in the collection of food intake data. QnFFQs normally have a heavy respondent and interviewer burden. The quality of the data collected depends on the memory of the respondent, as data is normally collected to give an indication of the dietary intake during the past month. This methodology is therefore not recommended in cases where the memory of the respondent could be an issue, e.g. in the elderly and in young children. The QnFFQ tends to overestimate the intake of the population, especially for food items that are eaten regularly, but not on a daily basis. Overestimation of foods that are often consumed on a daily basis, such as bread, is less frequently experienced.

Dietary intake records

The method of choice for the collection of dietary intake data in intervention studies/clinical trials/experimental research is the weighed or estimated dietary intake record method. This method is usually also used as the reference method when other questionnaires, e.g. the QnFFQ, need to be validated. To keep an appropriate dietary intake record, the participants are requested to document everything they eat and/or drink for a number of days. The number of days can vary, e.g. three or seven days (to represent one week), or even a longer period of time, but the longer the period the dietary intake records are kept the higher is the risk that respondents may change their dietary habits. Furthermore, in order to ensure the proper collection of dietary intake data, different aids are used. For instance, a dietary instruction booklet explaining how the dietary record should be kept and training of the participants in the recording of their dietary intake are essential. It is also essential that an electronic scale, which can measure up to 1 g accurately, be used. This is however not always possible as these scales are expensive, and when a large number of participants are studied finances are not always available for such purchases. Other weighing scales such as household scales, weighing accurately up to 5 g, can sometimes be used provided these scales are standardised against an electronic scale and participants are trained on how to use the scales.

When appropriate scales for the weighing of food are not available then use of estimated dietary records is usually the method of choice. To ensure that the quantification of food portion sizes is correctly calculated, different types of food models and household measures, e.g. cups and spoons, are used. A summary of the tools that can be used for the quantification of food portion sizes has been published elsewhere.[4] A ruler is one of the important tools in the quantification kit as this can be used to determine the dimensions of the food items consumed, e.g. slice of cake, apple, cube of cheese, and it can even be used to determine the size of a mutton chop. The latter, however, poses some challenges, and overestimation of portion sizes is often a problem.

One of the golden rules for keeping dietary intake records is to train the participants in the importance of recording food(s) eaten, immediately after it was eaten, or as soon as possible thereafter. If this is not done then the collection of data is that of a dietary recall rather than a dietary record.

Coding of the dietary intake data

Once food intake data have been collected the information should be converted into a format that enables the determination of the energy and nutrient content of the diet. In order to analyse dietary intake data electronically, information on the food item code number and the quantity of food eaten is required. There are presently two methods available for the coding of food intake data for analysis purposes.

Electronic coding system

In FoodFinder3 an option is available for entering the food item (type) and the quantity of food eaten during the day.[2] The Meal Option, a data entry option, is chosen for the execution of this function. The program allows entering food intake data for the individual for a single day or for a number of days. It is not necessary to know the food item code for coding purposes as food item codes are registered automatically when the food item name is entered. The registration of wrong food codes can occur when the coder does not choose the appropriate food item to represent the food item eaten. To enable the coder to make the correct choices fieldworkers have to record the food item eaten in as much detail as possible, e.g. with fat or without fat, how was it prepared, what type of fat was used in preparation, brand name if possible, and related details. Therefore one of the most important steps in the collection of dietary intake data is to introduce quality control measures during the collection of dietary intake data and to cross-check information at the end of the interview to ensure that all the relevant information has been recorded. Another source of error in the registration of food codes is when wrong assumptions are made by the coder and the wrong food item is chosen to represent a food item eaten. To prevent this, coders should have a good knowledge about food preparation methods as well as the composition of food(s), e.g. there are different types of ice cream with various fat contents on the market. If only ice cream is recorded the coder has to make a choice and choose an appropriate food code for coding purposes. Not all the foods consumed, especially within a Western type of dietary pattern, are available in SAFOODS. Coders therefore have to make assumptions and choose an appropriate food item code. There are no general rules for making assumptions as each study has its own characteristics. Recording the assumptions made, however, is important in order to refer back to these assumptions for explaining and investigating discrepancies in the results. If the quantity of food eaten was reported in terms of household measures then the household measure can be entered into FoodFinder3 and the data will automatically be converted into grams of food eaten. This is possible because the information in the Medical Research Council (MRC) Food Quantities Manual (FQM) is included in the database of FoodFinder3.[11] This function enables the user to enter a half (0.5) cup of rice and the program translates this information into grams of food eaten. The option to enter the information on the weight of the food in grams is also available, e.g. when the quantity of food eaten was recorded in grams instead of household measures or the food eaten is of known weight, e.g. one small packet (25 g) of potato crisps. The only way to ensure that data is entered correctly is to double-check the information coded or to re-code about 10% of a randomly selected sample of the dietary records. It is preferable that the same coder re-code the data. See also the section on 'Quality Control of the Data'.

Manual coding

Another method available for the coding of the dietary intake data is manual coding. This is normally done when large datasets are coded and access to an electronic format of the SAFOOD is available through the MRC. Data has to be coded manually on specially designed coding sheets. The data is coded for the type (food item identity number) and the quantity in grams, of the food consumed. The MRC FQM is required for the conversion of the quantities of food reported in household measures into the quantities in grams.[11] Looking up the food item code numbers and the food quantities in the MRC FQM can be a time consuming exercise when coding the information manually.

Challenges of coding dietary intake data

Food items not available in SAFOODS

Presently there is information available on 1472 food items in SAFOODS.[12] Taking into account that there are thousands of foods available on the South African market it is not possible for food database compilers to include all the food items consumed in a country in the national food composition database.[1] There will often be a situation where alternative food items will have to be coded because a specific food item is not available in the SAFOOD. In such cases the coder should then search for a food item in the food composition database which best fits the food item consumed. Should no similar food item be available in the SAFOOD, other options for obtaining the nutrient information for a specific food item would include consulting other food composition databases such as the United States Department of Agriculture Database (http://www.ars.usda.gov/main/site_main.htm?modecode=12-35-45-00). Nutrient information from food labels such as health bars or special products, or from food manufacturers, are also alternative options for obtaining the necessary information on food items not available in the SAFOOD. This information can then be added to the existing information in the food composition database in FoodFinder3, and used for coding and analysis purposes. Guidelines for adding this information to the database are available in the FoodFinder3 user manual and the Help function of its software.

The SAFOOD includes nutrient information for many combined dishes, e.g. stews, puddings, cakes, tarts and salads. The possibility exists however that nutrient information on special dishes consumed may not be available in the SAFOOD, in which case the ingredients of such special dishes need to be coded separately. Examples of such special dishes include pizzas, mixed salads, and other recipe-based dishes. A recipe calculation is also recommended to determine the nutrient composition of combined special dishes such as stews, baked puddings and country-specific recipes that are not available in the SAFCT or FoodFinder3. A thorough understanding of food preparation techniques is required to execute recipe calculations successfully. The respective food items, available in the food composition database of FoodFinder3, should be selected as the ingredients of the recipe and the relevant quantities used in the recipe entered for recipe calculation purposes. The nutrient composition is then calculated and the recipe of the combined special dish is added to the food composition database of FoodFinder3 for coding and analysis purposes. Preparation methods also have to be taken into account with recipe calculations. It should always be born in mind that if raw ingredients are chosen as part of the recipe then retention factors have to be applied to compensate for moisture and nutrient losses due to the application of heat. FoodFinder3 is not programmed to compensate for the effect of moisture losses of a food item due to preparation methods (e.g. baking, boiling, roasting, frying, and other such food preparation techniques). In recipe calculations this has to be taken into account and the necessary adaptations have to be made manually before the food item is entered as a recipe ingredient, e.g. if raw meat is part of the ingredient list provision should be made for moisture losses, which could vary, depending on the meat cut and preparation method. The relevant weight of the cooked food item should be coded for analysis purposes. Information on retention factors is available in the literature and should be applied for the conversion of the nutrient content of raw food to cooked food.[13] In calculating the nutrient composition of a food item that has been marinated, the nutrient content should only be calculated for the marinade absorbed as some of the marinade could be lost during, for instance, grilling of the food item.[14] In order to calculate the nutrient composition of a given food item, those who calculate the nutrient composition of a recipe need to make assumptions about the losses. Weighing the food item before it is marinated and then afterwards could, for instance, provide some information for estimating losses. Losses of fibre, for example due to straining of fruit or vegetable purees, also have to be taken into account when recipe calculations are made.[14] After the ingredients of the recipe have been added to FoodFinder3 it is only necessary to enter the quantity of the special dish (recipe calculation) consumed. If calculations for a recipe are done manually the name of the recipe and the relevant nutrients per 100 g of the recipe can be added to the food composition database of FoodFinder3 for later use. It should be born in mind however that the nutrient values generated for a food item by recipe calculation are always estimated values.[14]

Food quantities not available

The MRC FQM is a useful tool for the conversion of food intake data reported in household measures to the actual weight in grams of the food consumed.[11] Information in the MRC FQM represents information on the weights of food items measured with different household measures or in different dimensions. This conversion from household measures to weight in grams is essential in order to analyse dietary intake data electronically. Not all the food items in SAFOODS have information on the weights of household measures. In such cases the researchers have to measure and weigh the foods themselves. This is also applicable to food items that are not available in SAFOODS and for which data are needed on the weight of the food. Users sometimes interpret the information on household measures, available in the MRC FQM, incorrectly as an indication of standard portion sizes. Standard portion sizes are often used for nutrition education purposes or in pre-coded dietary intake research questionnaires. The definition of a standard portion size for the same food item may differ, because the aim of using the concept of a standard portion size could differ. Researchers sometimes make use of standard portion sizes in pre-coded dietary questionnaires as they want to simplify the research process, or the aim of the study may require less exact data. In such cases it is then less time consuming to collect and code the dietary intake data if portion sizes are defined before the study is undertaken and the weight of these standardised portion sizes are pre-coded in the dietary intake questionnaire.

Converting food intake data to nutrient intake data

Irrespective of the method used for the coding of the dietary intake data the coded food intake data have to be linked to the energy and nutrient content information in the food composition database. The information on the type and quantity of food that was eaten and entered into FoodFinder3 is automatically linked to the nutrient information of the SAFOOD. In the database, information on the energy and nutrient composition of the food is captured per 100 g edible portion and calculations are therefore required to reflect this information for the relevant quantities of food consumed. FoodFinder3 has been programmed to convert food intake data to energy and nutrient intake data. This information on energy and nutrient composition of the diet can then be exported to MS Excel in different formats. Four options are available for exporting of the data to an MS Excel sheet. *Option 1:* All the information recorded, e.g. the respondent identification, the code number of the food item, the name of the food item, the quantity consumed and all the relevant nutrient content information for the specific food item; *Option 2:* The energy and nutrient intake of the individual per day; *Option 3:* The mean intake of the individual for the number of days over which the individual's dietary intake was recorded, and *Option 4:* The mean energy and nutrient intake of a group of individuals.

If data are coded manually, the information has to be computerised and recorded on a spreadsheet, e.g. an MS Excel sheet, for linking it to the electronic MRC copy of the SAFOOD. Biostatisticians familiar with data processing develop specialised programs in statistical packages, such as SAS, for the analysis of the dietary intake data that have been coded manually.

Analysis of dietary intake data

Dietary intake data are usually analysed to reflect the nutrient intake for one day in order to compare the data with reference values such as the Dietary Recommended Intakes (DRIs) or Acceptable Macronutrient Distribution Ranges (AMDRs). When the dietary intake data represent the intake for more than one day, e.g. a week, a month, or a longer period of time, the information has to be translated to represent the average intake for one day.

Manual analysis

Dietary intake data can be analysed manually using the information in the SAFCT, but this is a labour intensive activity. This option for analysis is usually only practical if food intake data for only a few days has to be analysed. It is recommended, however, that students in nutrition and recently qualified nutrition professionals who have no or limited experience with food composition data should first gain experience in the manual analysis of dietary intake data before they use software programs for such analyses. This will ensure a better understanding of the characteristics of the SAFOOD and the MRC FQM. Analysing data manually also makes the users more aware of such aspects as the missing data for specific nutrients, the meaning of trace and zero values, and the reasons why total fat is not the sum of the different fatty acids.[1] All of these factors are discussed in the introduction of the SAFCT and they all impact on the correct interpretation of the final results.

Analysis using software programs

Analysing dietary intake data by means of software programs is a much easier task and (in any case) essential for the analysis of large datasets. The disadvantage of using software programs, however, for the analysis of dietary intake data is often the lack of knowledge about the food composition database used in the software program. It is thus crucially important to read the manual provided with the software program to be used in order to prevent inaccurate reporting of the data. In this second part of the review, only the use of FoodFinder3 for the analysis of dietary intake data is discussed as the program includes the SAFOOD discussed in Part I of this review. Globally, many dietary analysis software programs are available using different food composition databases and should readers need more information they are advised to consult the Internet and search for the software programs containing internationally recognised food composition databases. The International Network of Food Data Systems (INFOODS) website can be consulted for information on international food composition databases (http://www.fao.org/infoods/index_en.stm).

Most of the questions about energy and nutrient values in the SAFOOD are answered in the frequently asked questions section of the hard copy of the SAFCT.[13,15] FoodFinder3 enables students and researchers in nutrition and those undertaking studies with small numbers of subjects to analyse their own data. The copy of the manual provided on the FoodFinder3 CD, as well as the Help Function, explain how to enter food intake data for an individual or a group of individuals.

The total daily intake or the mean intake of a number of days for an individual or a group of individuals can be calculated with FoodFinder3. It should, however, always be remembered that FoodFinder3 is mainly a data capturing program and not a statistical software package. To enable the researchers to carry out descriptive statistics and further statistical analysis the dietary intake data should be exported to MS Excel. This information can then be transferred into any other statistical package that the user is familiar with, or imported into statistical packages that allow one to import an MS Excel file. Typical statistical packages such as Statistical Analysis System (SAS), Statistical Package for the Social Sciences (SPSS) or Stata can be used for further data analysis.

If Option 1 is chosen for exporting the data to MS Excel, the exported data also enables the researcher to analyse the data to provide information on the main food items consumed, the mean portion sizes consumed and also calculate the contribution that the 16 different food groups in the database make to nutrient intakes. Each of the food items in the SAFOOD has, in addition to its four-digit code number, a two-number code to identify in which food group the food item belongs, thus making the calculation of the contribution of the different food groups to energy and nutrient intakes possible.

Quality control of the data

Quality control of dietary data is crucially important before the data is statistically analysed and presented. Preliminary analysis is often performed to assist with the 'cleaning' of the data.

There are several stages during the research process that may affect the quality of the dietary data that has been collected:

1. Choosing the correct dietary methodology would be the first step to ensure that the collected data is valid and reliable. For instance, if potential respondents cannot read and/or write they should not be requested to keep a dietary intake record. In addition, recall methods should not be used for the collection of dietary intake data on children younger than seven years of age.[16] Making use of a methodology such as the QnFFQ to collect data over a long period of time, e.g. one month, could be problematic in the elderly, where poor memory may impact adversely on the quality of the data.
2. The involvement of an experienced professional with relevant experience in coding dietary intake data is essential to ensure data of high quality. Choosing the wrong food item code when data is coded has a major adverse effect on the results. Furthermore, researchers may also not be familiar with the 'jargon' used in a specific community, and if the interviewer who collects the data does not properly document the food details it is difficult for the coding personnel to select the appropriate food item from the food composition database. An example of this is the case when chicken-feet and chicken-heads were consumed and the data collectors only recorded 'walky-talky', the popular name used by the community for these food items. Assumptions have to be made in a situation like this and if detailed descriptions are not available the wrong food codes could be used, which will impact negatively on the final results.
3. Overestimation or underestimation of portion sizes is also often a problem with the recall methods, e.g. the 24-hour recall and the QnFFQ. If appropriate food models are not available for quantification purposes then estimation of portion sizes could be problematic, especially for food items such as meat cuts. More information on the type and use of different food models for accurate quantification purposes is summarised elsewhere.[4]
4. The infeasibility of including in a food composition database all foods available in a country poses special challenges for the coding personnel of food intake data, especially when the population studied follows a Western type diet containing a large variety of food items. In such cases an alternative food item(s) has to be chosen with the attendant limitations. For example, if there is no information on ice cream cones in the database then something like puffed rice would be chosen. This will of course not give the exact same nutrients as one would have expected had chemically analysed data on the nutrient composition of the cone been available. Choosing a food item with more or less the same nutrient profile is nevertheless better than not coding the food item at all. The FoodFinder3 software program makes provision for a specific food item to be added to the database, if the researcher has nutrient analysis information on a given food item.
5. Another error often made when dietary intake data is coded manually is when the wrong quantity of food is coded, e.g. 5000 g of sugar instead of 0005 g. This will of course have major implications when results on the mean intake of, for example, added sugar are reported.
6. Researchers should never accept dietary intake data at face value; they should always ask relevant questions. Moreover, and after the preliminary analysis of the data for means, standard deviations and frequency distributions, researchers should always identify any outliers in the data set and establish the reason(s) for such outliers:
    i. If values for individual nutrients seem to be very high or very low (outliers) the original data should be evaluated for the quantities and codes used for a given food item. The translation of estimated food intake data into grams of food consumed is

in itself a challenge and sometimes the dimensions reported are not realistic, and also need to be re-evaluated.

ii. Specific nutrients, e.g. vitamin B_{12}, cholesterol and vitamin A may often be identified as outliers in the dataset. This is often due to the consumption of food items that are major sources of these nutrients. It is then important to ensure that there are legitimate reasons for the nutrient value identified as an outlier.

iii. Nutrient values identified as outliers could also be due to errors in the food composition database, e.g. there may be a mistake with the decimal points. Should this be the reason for an outlier value the compilers of the food composition database should be informed in order to correct the error in the database.

Once the researchers have resolved all the queries in the dataset and are satisfied that the data are clean they can proceed with the final analysis of the data.

Advanced statistical analysis

The study design(s) (Table II) and the dietary methodology used for the analysis of the data play a crucial role in the selection of the appropriate statistical tests for analysing dietary intake data.

Table II: Study designs most often used in dietary intake studies

Study design	Characteristics
Cross-sectional studies (also called surveys, epidemiological studies and prevalence studies)[17]	Describe the collection of data at a specific time Dietary intake data and data on other variables are collected at one point in time.
Cohort studies	Involve a group of people who are studied over a period of time Dietary intake data are collected at the start of the study and the group is monitored to evaluate, for example, whether a high intake of saturated fatty acids increases the risk of having a heart attack.
Case-control studies	Study of people with a specific/rare health outcome, e.g. oesophageal cancer (case), and selection of a subject (control) with the same characteristics, but without the disease Dietary intake data are collected retrospectively to determine which nutrients could possibly be responsible for the development of the disease.
Experimental studies[17] (including clinical trials [normally used for experimental studies in medicine where humans are involved]), controlled trials [that include a control group]	These studies normally test the effect of different dietary variables on biological outcomes in a randomised (open or blind) control trial Dietary intake data are collected at the baseline in two or more groups and then again during and/or at the end of the dietary intervention period.

The selected dietary methodology for the collection of the dietary intake data also impacts on how the data can be statistically analysed. The advantage of the single 24-hour recall, for instance, is that it can be used to describe the mean intake of a given population. To ensure that the sample size is large enough, Gibson provides guidelines and a formula for determining the number of subjects required to reflect the group mean for a specific nutrient.[3] The calculation should be based on all the nutrients of interest and the worst case scenario should be assumed for determining the required sample size.[3] One of the disadvantages, however, is that a single 24-hour recall should not be used for the classification of respondents into high or low intake groups, e.g. what percentage of the population consumed more or less than the reference value for a specific nutrient, as it does not reflect usual dietary intake.[3] Another limitation of a single 24-hour recall is that it only reflects the intake of one day, whereas the biochemical nutrient status data is a reflection of dietary intake and metabolism over a longer period. Repeated 24-hour recalls will therefore be needed to correlate the dietary intake data with the biochemical measurements or to categorise respondents.[3] Within-subject and between-subject variation influence the number of 24-hour recalls needed to meet the aim of the study. Researchers should therefore consult dietary research methodology handbooks for calculating the number of repeated 24-hour recalls needed for their particular study.[3] If only one 24-hour recall has been collected other variables such as anthropometry and biochemical measurements can be used for the categorisation of the population into different groups, e.g. those with high or low values, and the nutrient intakes of the two groups can then be tested for statistical differences. In contrast to the use of a single 24-hour recall, information from the QnFFQ and the dietary record can be used for categorising people based on their nutrient intakes or to compare the intakes with reference values.

Researchers are usually familiar with programs such as MS Excel, which enable them to carry out basic processing of the dietary intake data, e.g. to calculate means, standard deviations, medians, and frequency distributions of the data. The former are always important, to enable the researchers to familiarise themselves with the characteristics of the dataset, in order to, as mentioned previously, ensure quality control of the dataset, and to make the necessary corrections when/where required before advanced statistical analyses are undertaken.

Examples are now given in order to provide the reader with some guidance on the statistical test(s) that could be applied when dietary intake data are analysed. The aim is not to discuss different statistical tests in detail, but merely to give some guidelines to the researcher who needs to analyse dietary intake data. It is however important to note that the researcher should consult a statistician already in the planning phases of a study, especially if advanced statistical tests are required for data analysis.

Example 1

Study design: Cross-sectional study

Aim of the study: To determine whether the saturated fatty acid intake of participants with a high plasma cholesterol concentration is higher than those with a normal plasma cholesterol concentration.

Research setup: Collect dietary intake data by means of a 24-hour dietary recall and collect blood samples for the analysis of the lipid profile of 200 participants.

Diet related research question: Is there a significant difference between the saturated fatty acid intakes of two groups of subjects: Group 1 (N = 120), with a high plasma total cholesterol concentration

(> 5.0 mmol/L); and Group 2 (N = 80), with a normal plasma cholesterol concentration (≤ 5.0 mmol/L)?

Factors influencing statistical analysis: There are several factors that have to be taken into account when dietary intake data are analysed. Parametric tests can only be used if the data are normally distributed. Dietary intake data are usually not normally distributed, and in order to compare the saturated fatty acid intake between the two groups it is necessary to transform the data mathematically in an attempt to normalise the distribution, for use of a parametric test.[18] Sometimes it is not possible to get a normal distribution with transformation of the data and a non-parametric test has to be used for the analysis of the data. The advantages of non-parametric tests are that they do not depend on the normal distribution of the data or the mean, standard deviation or variance.[18] When data is not normally distributed it is meaningless to only report the mean intake and standard deviations. Median values and the range or percentiles should rather be reported. The median value is the least affected by extreme values.

Recommended statistical analysis: The two samples studied are independent and the sample size is rather small. The data for saturated fatty acids are not normally distributed in both groups and a non-parametric test is therefore recommended. In this instance the Mann-Whitney U test (also called the Wilcoxon rank-sum test) should be used.[18] Had the data been normally distributed and the sample sizes large, a two-sample t-test should have been used. The Wilcoxon test is still valid for data from any distribution, whether normal or not, and is much less sensitive to outliers than the two-sample t-test. When the assumptions of the two-sample t-test hold, the Wilcoxon test is somewhat less likely to detect a location shift than is the two-sample t-test. However, the losses in this regard are usually quite small.

Example 2

Study design: Cross-sectional study

Aim of study: To determine whether the intake of dietary iron of children 4–8 years of age meets the Dietary Reference Intake (DRI) values for iron requirements.

Research setup: Collect 24-hour dietary recall data from 100 children living in a rural area in South Africa. Analyse the dietary intake data for iron intakes making use of the SAFOOOD.

Diet related research question: Does the iron intake of children 4–8 years of age differ significantly from the Adequate Intake (AI) values for iron?

Factors influencing statistical analysis: The intake data of iron are probably not normally distributed. Thus, one will have to do a suitable transformation to make the distribution normal in order to do a parametric analysis. Otherwise, a non-parametric approach can be followed to test whether the mean value of the sample differs significantly from 10 mg/day which is the AI for iron for children 4–8 years of age.

Recommended statistical analysis: Obtain the difference between the actual intake measurement and the AI of 10 mg/day. Then, for the parametric approach, do a paired t-test, and for the non-parametric approach, the Wilcoxon's signed rank test.

Example 3

Study design: A dietary experiment/clinical trial

Aim of study: To determine the effect of eating two portions (100 g each) of fatty fish per week, on the omega-3 fatty acids, eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), content of the red blood cell membrane phospholipids.

Research setup: Two groups of participants, each with a sample size of 50 subjects, who followed an experimental diet (habitual diet plus two portions of fatty fish per week, but no other fish) and a control diet (habitual diet and no fish per week), for a period of eight weeks. Blood samples were collected for the analysis of the fatty acid composition of the red blood cell membrane phospholipids at the end of the baseline period and after the eight weeks experimental period. To control for the dietary intake of EPA and DHA, participants kept one five-day dietary record during the baseline period and three five-day dietary records during the eight-week intervention period.

Diet related research questions: (1) Are the EPA and DHA concentrations in the red blood cell membrane phospholipids of participants who ate fatty fish for eight weeks significantly higher than in those who did not eat fatty fish? (2) In those who ate fatty fish, were the EPA and DHA concentrations of the red blood cell membrane phospholipids higher at the end of the intervention period than at the start of the intervention period?

Factors influencing statistical analysis: In this case the intake of omega-3 fatty acids may be evenly distributed as the same quantity of fatty fish was prescribed. The omega-3 fatty acid content of the red blood cell membrane phospholipids may however not be normally distributed. The two quantitative data sets are not independent as we have "before" and "after" data for the same individuals in the experimental group and also for the same individuals in the control group.

Recommended statistical analysis: One will obtain the difference between the "after" and "before" blood values. Keeping in mind all the statistical shortcomings of the data, as discussed in the two previous examples, then, depending on the distribution of the data, one should do either a paired t-test or the Wilcoxon's signed rank test. One can also do a logistic regression in order to obtain the odds ratio, which is a measure of the effect size of the "change" in the blood values, and simultaneously correct for other confounder variables.

Should there have been another group in the above-mentioned study example, there would have been more than two independent groups in the study. To test for significant differences between multiple groups, the Kruskal-Wallis test is recommended should the data not be normally distributed. This test will indicate that there is a difference between the groups, but further non-parametric testing will be required to test for the difference between specific groups.[18] The parametric approach would be to do an analysis of variance (ANOVA), followed by a pair-wise test (Tukey, for example) to indicate between which groups the difference exists.

Example 4

In the study mentioned in Example 3 the researchers also wanted to investigate whether there is a correlation between the EPA and DHA content of the diet and the EPA and DHA content of the red blood cell membrane phospholipids.

Factors influencing statistical analysis: See Example 3 for the factors influencing the choice of an appropriate statistical method.

Recommended statistical analysis: Use a Spearman's correlation test when the data is not normally distributed and a Pearson's correlation when the data is normally distributed.

Reporting of dietary intake data

Presentation of data

The results on dietary intake can be presented in tables, figures or graphs. Dietary intake data is often reported as means and standard deviations. Although this format allows for the easy comparison of results between studies, dietary intake data is often skewed, with large standard deviations. The best way of reporting dietary intake data is therefore to also give the median with the inter-quartile range.

In the SAFOOD the energy and nutrient information represent average values and in the interpretation of dietary intake data this should be kept in mind. To report values for energy intakes in decimals points should therefore be avoided. The best way of reporting energy and nutrient intake data is to use, as a guideline, the decimal point system in the applicable food composition database used.

Missing nutrient values

In Part I of this series attention was drawn to the fact that there are missing values for some nutrients in the SAFOOD.[1] Although food composition database compilers do always try to limit the number of missing values in the database it is not always possible to generate values for some nutrients. The presence of a large number of missing nutrient values in the database can lead to the wrong conclusion that a large percentage of the population has an inadequate intake of a specific nutrient. In the analysis printout of FoodFinder3 the user is cautioned on the nutrients in the food composition database that have many missing values. In future, nutrients with a high percentage of missing values will probably no longer be available in FoodFinder3, in order to prevent the incorrect reporting of nutrient intake data.

Comparing results with reference values

In comparing nutrient intake data with reference values, e.g. the DRIs, the researcher should ensure that the unit of measure used in the food composition database is the same as the unit of measure used for the reference value. As an example, presently the unit of measure for vitamin A is retinol equivalents (RE) in the SAFOOD, while in the DRIs it is retinol activity equivalents (RAE). For preformed vitamin A in foods RAE and RE is the same (1 µg retinol = 1 µg RAE = 1 µg RE), but for dietary provitamin A (β-carotene), RAE is twofold greater than RE (12 µg β-carotene = 1 µg RAE = 0.5 µg RE).[19] In calculating RE from β-carotene, 6 µg β-carotene = 1 µg retinol has been used in the SAFOOD. The unit of measure for a specific nutrient in the food composition database could also be in micrograms while the reference value is in milligrams. In the reporting of the data the researcher should ensure that these possible differences in units of measure have been attended to when the data is analysed and reported.

Impact on results when updating the SAFOODS

The updating of SAFOODS is currently infrequent due to time and financial constraints. The chemical analysis of food for nutrient composition is an expensive undertaking and this limits the number of food items for which the nutrient composition can be generated and added, or updated, in the database. Individual food groups are updated from time to time and this has an impact on the information available in the food composition database. Not only are new food items added to the database, which enable the users of the database to make better choices, but the nutrient content of a specific food item can also change during the updating process of a food composition database. The reason for this is that the nutrient information for a specific food item is sometimes "borrowed" from other databases. The main reference for a food item is printed in the SAFCT and this enables the user to establish whether the nutrient data is of South African origin or "borrowed" from another food composition database. The percentage of "borrowed" values in SAFOODS is constantly changing as the database is updated and more values of South African origin are added to the database. When new data are generated the nutrient values of a specific food item could change because the information for the South African product could differ from the values borrowed from another country. The product could either contain more or less fat for instance than the borrowed value or the preparation methods used could differ between the borrowed and the locally analysed food item. The effect of changes in the food composition database is especially important when dietary intake data are collected as part of longitudinal studies. The same database should therefore be used throughout the study to prevent interpreting changes in dietary intake as being due to a dietary intervention, when the change is in actual fact due to the changes in the relevant database.

Reporting on the dietary intake of South Africans

In the past, several large studies were undertaken to collect data and report data on the dietary intake of the South African population.[5–7] Most of these studies were carried out 20 to 30 years ago, and hence the data is now outdated. A NFCS was done in 1999 on children 1–9 years of age, but no national dietary intake data is available for adults.[8] In the past, data from individual studies undertaken in the country were summarised in order to describe the dietary intake of adult South Africans. In these studies different methodologies were used for the collection of data and the data is also not representative of the population.[20] The high costs involved and other logistical challenges may make it impossible, or certainly infrequent, to undertake national food consumption surveys on a regular basis in future. National data on food, energy and nutrient consumption and dietary patterns are, however, essential to monitor changes in dietary intake over time, to relate this to nutrition related health outcomes, and to formulate nutrition related health policy. In the absence of a national survey, the less desirable interim alternative that might afford an impression of trends at the national level might be for researchers in the country to use standardised dietary assessment protocols, which will include *inter alia*: (1) a sampling procedure; (2) questionnaires validated for different age groups; (3) a standardised coding system; (4) use of the SAFOOD for analysis; and (5) use of standard statistical methods of data analysis. The implementation of such a protocol, whenever a dietary intake study is undertaken in South Africa, would enable the pooling of the data, to afford an impression of the dietary intake, and trends thereof, of South Africans on a more regular basis.

Conclusion

In order to collect dietary intake data from the South African population, researchers require access to food composition data, tools to analyse the data, and familiarity with the appropriate methodology for collecting and analysing such data. Information on the energy and nutrient composition of foods commonly consumed in South Africa is available in SAFOODS, as well as in related products from SAFOODS, e.g. printed tables and the software program FoodFinder3. Further, using a standardised methodology for the collection and statistical analysis of dietary intake data in South Africa could increase the pool of knowledge on the energy and nutrient intake of the population, albeit as a much less preferred alternative to collecting dietary intake data at the national survey level regularly.

References

1. Wolmarans P, Danster NA. Characteristics of the South African Food Composition Database, an essential tool for the nutrition fraternity in the country: Part I. S Afr J Clin Nutr 2008;21:308–13.
2. FoodFinder3. Dietary Analysis Software. Parow Valley, Cape Town: Medical Research Council, 2002. Available from http://foodfinder.mrc.ac.za (Accessed 24/04/2009).
3. Gibson RS. Principles of nutritional assessment. 2nd ed. New York: Oxford University Press, 2005.
4. Wolmarans P, Wentzel-Viljoen E. The South African Food Composition Database and tools for collecting and analyzing dietary intake data. In: Steyn NP, Temple N (eds), Community nutrition textbook for South Africa: a rights-based approach. Cape Town: South African Medical Research Council, 2008:795–824.
5. Wolmarans P, Langenhoven ML, Benadé AJS, Swanepoel ASP, Kotzé TJvW, Rossouw JE. Intake of macro-nutrients and their relationship with total cholesterol and high-density lipoprotein cholesterol: the coronary risk factor study, 1979. S Afr Med J 1988;73:12–15.
6. Langenhoven ML, Steyn K, Van Eck M, Gouws E. Nutrient intake in the coloured population of the Cape Peninsula. Ecol Food Nutr 1988;22:97–106.
7. Bourne LT, Langenhoven ML, Steyn K, Jooste PL, Laubscher JA, Van Der Vyver E. Nutrient intake in the urban African population of the Cape Peninsula, South Africa. The BRISK study. Cen Afr J Med 1993;39:238–47.
8. Labadarios D, Steyn N, Maunder E, MacIntyre U, Swart R, Gericke G, Huskisson J, Dannhauser A, Vorster HH, Nesamvuni AE. The National Food Consumption Survey (NFCA): children aged 1–9 years, South Africa, 1999. Stellenbosch: Department of Human Nutrition, 2000. Available from http://www.sahealthinfo.org/nutrition/foodconsumption.htm (Accessed 24/04/2009).
9. Steyn NP, Senekal M. A Guide for the use of the Dietary Assessment and Education Kit (DAEK). Parow Valley, Cape Town: South African Medical Research Council, 2004. ISBN: 1-920015-09-04.
10. MacIntyre UE, Kruger HS, Venter CS, Vorster HH. Dietary intakes of an African population in different stages of transition in the North West Province, South Africa: The THUSA study. Nutr Res 2002;22:239–56.
11. Langenhoven ML, Conradie PJ, Wolmarans P, Faber M. MRC food quantities manual. 2nd ed. Parow Valley, Cape Town: South African Medical Research Council, 1991.
12. Wolmarans P, Danster N, Dalton A, Schönfeldt H. Condensed food composition tables for South Africa. Parow Valley, Cape Town: South African Medical Research Council, 2009.
13. Kruger M, Sayed N, Langenhoven ML, Holing F. Composition of South African foods. Vegetables and fruit. Supplement to the MRC food composition tables 1991. Parow Valley, Cape Town: South African Medical Research Council, 1998.
14. Powers CH, Abbot Hess M, Kimbrough M. How accurate are your nutrient calculations? Why culinary expertise makes a difference. J Am Diet Assoc 2008;108:1418–22.
15. Sayed N, Frans Y, Schönfeldt HC. Composition of South African foods. Milk and milk products, eggs, meat and meat products. Supplement to the MRC food composition tables 1991. Parow Valley, Cape Town: South African Medical Research Council, 1999.
16. Cameron ME, Van Staveren WA. Manual on methodology for food consumption studies. Oxford: Oxford University Press, 1988.
17. Dawson B, Trapp RG. Basic and clinical biostatistics. 4th ed. New York: Lange Medical Books/McGraw-Hill, 2004:7–22.
18. Harris JE, Boushey C, Bruemmer B, Archer SL. Publishing nutrition research: a review of nonparametric methods, Part 3. Am J Diet Assoc 2008;108:1488–96.
19. Otten JJ, Hellwig JP, Meyers LD. Dietary reference intakes. The essential guide to nutrient requirements. Washington, DC: National Academic Press, 2008, p. 170.
20. Nel JH, Steyn NP. Report on South African food consumption studies undertaken amongst different population groups (1983–2000): average intakes of foods most commonly consumed. Pretoria: Department of Health, Directorate of Food Control, 2002. Available from: http://www.mrc.ac.za/chronic/foodstudies.htm (Accessed 24/04/09).