

Critical Reviews in Food Science and Nutrition, 46:329-336 (2006)

Copyright © Taylor and Francis Group, LLC

ISSN: 1040-8398

DOI: 10.1080/10408390600688156



Taylor & Francis
Taylor & Francis Group

Criteria for Optimizing Food Composition Tables in Relation to Studies of Habitual Food Intakes*

MARÍA JOYANES and LOURDES LEMA

Agencia Española de Seguridad Alimentaria, Ministerio de Sanidad y Consumo, Madrid, Spain

The purpose of this study is to increase the accuracy, reliability, and precision of food composition data and, in consequence, better approximate nutrient intake estimations and recommendations. To do this it is necessary to specify and taken into account factors that play an important role in the variation of composition in order to avoid excessively broad dispersions and irregularities in data distributions. This implies the presentation of representative and, as consequence, extrapolable data, with nutritionally grounded confidence intervals. This study suggests a methodology that better approaches the accuracy, reliability, and precision of food composition data.

Keywords representative samples, nutrients, non-nutrients, requirements, DRIs

OBJECTIVES

The main objective of this work is to propose a methodology that will obtain representative food composition data and, as a consequence, permit the extrapolation to similar foods, which fall in a preestablished range based on nutritional criteria. This work attempts to identify the uncertainties in the process of compiling food composition data in order to improve the actual results obtained of food composition.

INTRODUCTION

In the last decade, epidemiological studies in nutrition in Spain led to the somewhat unexpected conclusion that the majority of apparently healthy subjects were below the recommended daily nutrient intakes. DRIs.¹⁻⁷ (Dietary Reference Intakes) assumed that there has been some lack of precision in the former methods. The old intakes recommendations carried an overestimation due to the generalization process, among other potential factors of imprecision.^{5,8,9}

Current interest in the relation of diet and health has stimulated and increased the demand for food composition data. Many researchers made efforts to optimize the nutritional evaluation based on food composition tables (FCTs). In any quantitative study on human nutrition, information on the nutrient composition of foods is essential to estimate adequate nutrient and energy intakes to meet the requirements. Although, the big differences are not great, they can play an important role, but not necessarily a parallel role, in the nutritive value.¹⁰ It is also necessary to determine food consumption related to risk assessment, which means evaluating the role of dietary modifications. However, the studies are limited in assessing dietary risk and clarifying etiologic relationships because there is a lack of detailed information on both nutrient and nonnutrient content of many foods.¹¹ An important use of FCTs is that it could be the basis to develop supply policies.¹²

Foods are often complex chemical and physical entities in which physiologically active ingredients are not present in pure form and the physiologic effect may be due to synergy between several compounds. In addition to the problems created by a lack of biomarkers other challenges exist in developing appropriate study designs for foods.¹³ The optimization of methodology to contribute to a more recommended adequate energy and nutrient intake considers specific subject groups and symmetric distribution of requirements.¹⁴

Since 1945 the Food and Agriculture Organization (FAO) has worked with member countries to increase food production and food supplies and improve nutritional well-being.^{15,16} A 1983

*Oral presented in 4th International Food Data Conference. Bratislava, August, 42-26, 2001.

Address correspondence to María Joyanes, Agencia Española de Seguridad Alimentaria (in Instituto de Salud Carlos III), C/Carretera Majadahonda-Pozuelo, s/n, 28220-Majadahonda (Madrid), Spain. E-mail: gjoyanes@isci.es

conference organized by the United Nations University (UNU) proposed an international network of food data systems¹⁷ to address the need for and limitations of food composition databases. Rand (1987)¹⁸ considered that the majority of FCTs used in the United States came from only one source: the data published by the Secretary of Agriculture (They are made available to the public through Agriculture Handbook No. 8). Haytowitz (1995)¹⁹ gave information from USDA's Nutrient Data Bank containing a wealth of information on the composition of foods. However, Katch (1995)²⁰ cited a government report from the General Accounting Office which concluded that some of the Handbook No. 8 nutritional data had so little documentation on how the data were produced, that the possibility of some nutrient values not being reliable existed, and thus, suggesting to evaluate procedures to direct in a better way the generation of food composition data. This is only one aspect of this complicated task. International food data base design was created in 1992.²¹ One goal involves minimizing errors in food sampling, chemical analysis, the construction of food composition tables, nutrient data bases for food groups and the use of nutrient analysis systems which required more documentation²² for non-nutrients, for example the flavones in citric fruits.²³

On the other hand, over the last two decades there has been an attempt to develop new parameters relating to each nutrient. What is more, specific fiber, fat, etc., have been tabulated separately along with many aspects that were not considered before.^{24,25} Also, new discoveries have been taken into account regarding the properties of many food components which can interact with nutrients. Some of these substances, not previously considered as nutrients before, exert specific nutritional functions and they have been proposed to be considered as such in some cases.²⁶ However, the application to epidemiological studies cause such errors, that it has become necessary to try to develop a newer and more accurate approach.^{4,27-29}

Since the data about food composition are used for the evaluation and also for the planning of nutrient intakes both at individual and group levels, one of the more important uses is to calculate specific nutrient intakes from intake data. The reliability of such calculations depends not only on the accuracy of food-consumption estimates but also on the food-composition data on which these estimates are based. In addition we must take into account the difficulty of matching foods as they are eaten to entries in those databases.^{17,30} Moreover, they are used by a broad range of professionals for different scientific or financial objectives,¹⁸ so it is easy to understand the impossibility of obtaining a food composition databank that will be perfect for any use. Because of this, it is better to plan their content following a basic criterion. That is, to consider FCTs as an acceptable approximation to the real nutrient/non-nutrient content which, due to the different precision that their use needs, these FCTs should be taken as a correct approximation, using the methodological statistical tools to identify and select the adequate samples for analysis. This means that FCTs should be a scientific study of information which describes the food component content according to a previously established work plan and they should

not only be data collected from different sources. Southgate and Greenfield (1992)³¹ detail the principles for the preparation of nutritional data bases and food composition tables and confirm that "the true test of quality of a value is the accuracy with which it predicts the composition of another equal representative sample of the food." This argument is the general principle which any data collection system should comply with, and this paper is based on the endeavour to optimise any quantitative study of human nutrition.

CONSIDERATIONS RELATED TO THE AREAS OF COMPETENCE AND CONTENT OF FCTs

The current commercial interchanges do not facilitate the difficult task of delimitating the area of competence of FCT because, in spite of their content which includes fresh or prepared food, and simple or complex recipes which adjust to the needs of the area users, there are still some questions to resolve: (1) What is the food origin, although bought in the local market? (2) How are the data obtained for FCT? Directly from local food analysis, using bibliographic data or are both methods used at the same time? (3) When analysing an area product, is it taken into consideration that its composition could be significantly influenced by variations of local environment such as different kinds of soil, fertilizer, water or sunlight intensity in the case of vegetables; or differences in the types of feeding or breeding animals? If it is influenced, is this shown in the FCT? (4) If food industries from the area of competence use the data of FCT as a reference value, do these industries consider the origin of the product they are canning and confirm it is the same product?

The considerations regarding the content of FCTs can first be focused from two different perspectives: from the relation of foods that should be included, and from the range of nutrients and non-nutrients that should be specified. Another perspective is related to the form in which the product is analysed (raw, boiled, fried, canned, etc.) and its effects on the accuracy with which the results are presented.

The priorities concerning the foods and range of nutrients and other components to include have been very well described by Southgate and Greenfield (1992).³¹ Also Deharveng et al. (1999)³² have established some recommendations so that the nutritional values from different FCTs and data systems can be compared. On the other hand, although the inclusion of data from raw food is sometimes questioned because of not being consumed as such, it is fundamental to include it because of the following reasons. (1) It is in the raw product where the possible differences in the content of nutrients and no nutrients of foods dependent on seasonal and geographical variations^{10,33-40} or on any other environmental factor able to be identified ("controllable") should be determined, which differ from the residual variability inherent to any kind of food. (2) These are the reference points to evaluate the changes that the nutrients and other components of the food will suffer during storage and through the

elaboration process, including manipulation and heating.⁴¹⁻⁴⁴ (3) Therefore, it is concluded that all these data from raw food are essential for the calculation of adequate conversion factors, which will facilitate users of FCTs globally or in the stages of transformation⁴⁵ of fresh products to nutrient intakes. These will in fact contribute to the increase of the accuracy of data and as a consequence, the reliability of calculating specific nutrient intakes from the food intake data.

To consider food as living matter leads us to another perspective, which is to differentiate components intrinsic from others which are not.

In addition to the major components, foods have vitamins, minerals, and other minor components. Some of these minor components in many foods (but especially in vegetables), are biologically active and can interact with nutrients in order to change their bioavailability, or to maintain or improve their physiological functions and/or to prevent diseases.²⁶ Because they are biological materials and because there are not two identical individuals, there is a natural variation in their composition.

These intrinsic components suffer additional modifications in quantity, proportion, and organic distribution induced by environmental factors, including their own feeding. Also, other components can be incorporated to living matters by way of their food chain or by other means. These other components do not necessarily condition their composition, they simply join the organisms.

This third perspective leads us to two fundamental facts:

1. Foods not only present natural variations in their composition but also others, known or not, brought about by the media in which they are grown. It is essential to take them into account or to investigate them, if it proceeds, in order to obtain homogeneous and representative food samples.
2. The need to give serious consideration to the pertinence or not of including additives and contaminants in the FCTs and databank must be discussed with expert assessment.

CONCENTRATION VARIATIONS IN FOOD CONSTITUENTS

The analytical results of biological materials imply two sources of variability: that of the material itself and that of the method of measurement. Therefore, variability for food composition data include all variability attributable to sampling plans, sample handling, analytical method, and analytical quality control.⁴⁶

The aim of this paper is to focus particularly on the study of sampling plans, rather than on the other phases mentioned in the reference above, which have been stated explicitly with different perspectives by Southgate and Greenfield, 1992³¹; Feinberg et al., 1992⁴⁷; Holden, 1996⁴⁸; Deharveng et al., 1999.³¹ Though the phases sampling plans and sample handling are partly superposed, the latter constitutes mainly the required analytical protocols for the techniques used while sampling plans are the

appropriate sampling proceedings to obtain food samples representative of the population of origin (e.g. wide group of individuals from which a reduced but representative sample has been extracted to be analysed).

Because food is a biological material, another range of variability must be added to the inherent (residual) variability. This other variability that must be added is caused by potentially identifiable factors which influence the analytical results. These causes of variability mentioned before, could be generally named as "environmental factors that influence inner constitution of living matter, and are sometimes cited as possible causes of irregularities in the distribution of obtained data. This is not considered enough, and therefore these causes should be taken into account and their connection with the analytical results should be investigated.

The distribution of irregular data with excessive dispersion is, in principle, an indication of sample heterogeneity. It might be that the distribution of some nutritional relevant component is not apparently Gaussian, but it is impossible to know if it is really not Gaussian, without any further research. However, it is more reasonable to suppose that any quantitative variable could be studied according to the Gauss Law and try to find out, as far as possible, the cause of the heterogeneity.

It is necessary to collaborate with professionals from other areas such as agriculture, stockbreeding, fishing, fish-farming, etc., in order to obtain information about the original population of the food, that is, to obtain information on raw materials and breeding, cultivation or fishing conditions. An incomplete study of the population can lead to apparently correct conclusions, which are really unacceptable.⁴⁹

The lack of representative sampling can increase the degree of bias in the estimates of central tendency and cause errors in the estimates of variance. Bias or lack of precision in component estimates will contribute to erroneous conclusions in scientific investigations of the effects of dietary intake.⁴⁸

REQUIRED CONDITIONS OF REPRESENTATIVE SAMPLE

This objective is not new and many researchers have been making an effort to propose different ways to resolve the required conditions of the representative sample. Southgate⁵⁰ has compiled the data quality of sampling.

In order to have a representative population sample, the sample must be homogeneous, randomized and of adequate sample size. However the way in which the population is considered representative depends on the goal of the study. This means that if one makes a pool (proportional or not) of the whole sample which constitutes a universe of each food in which the global mean and standard deviation is comprehended as the totality of the universe of each food, it could be statistically correct. However, the use of these studies leads to a virtual situation, where the majority of the samples and the majority of the situations could be far from reality. This, also plays a role in the nutritional

studies. In addition, a greater number of sample sizes could be required among other broad imprecision, strictly from the point of view of food composition. The reason why the results must be presented as a pool, divided in groups or alone, depends on if they have similar or different amounts nutrients composition, respectively. Because of this, we consider that expressing the results as a mean and standard deviation is not enough. Before deciding whether to present the results together or separately, it is necessary to check the similarity or difference of the samples. In addition, it must be kept in mind that results should be presented as maximum and minimum values give the range information in which the data fall, but at the same time the maximum and minimum values are the two less probable possibilities. For example, one must decide if the mangos that have a minimum carotene content ($16 \mu\text{g}$) should be taken together with the other varieties of mangos with $420 \mu\text{g}$ β -carotenes content.²⁵ Independently and before considering the new conversion factor to vitamin A,⁵¹ that choose if the sample with must be take together with the sample without. Also one must decide if the mussels, which have a seasonal difference more than 20% of the protein content and also have a difference in fat content over 24%, should be included as a mean.¹¹

Specifically the proposal/goal of our present work assumes that the sample must be: (1) Homogeneous in all factors which are already known to have an influence on the experiment or if suspected that they could have, so that everything will coincide except what is being studied. In this way, excessive and irregular dispersions in data distribution will be avoided. (2) Randomized in order to eliminate the risk of systematic error or bias of the sample that might occur due to other possible factors of influence still unknown to the researcher. (3) An adequate size in order to obtain a range of variability (confidence interval to the arithmetic mean in the population), whose maximum range is prefixed when planning the work. With good homogeneous sample criteria, the size does not need to be excessive.

The statistics give different strategies of sampling that help in the theoretical design of the experimental work must be selected in function of the final objective. This work gives a simple application of the statistics to achieve the objective mentioned above.

PROPOSED METHODOLOGY TO OBTAIN REPRESENTATIVE SAMPLES

To elaborate an FCT it is necessary to differentiate two phases:

- (1) To decide which food must be included in the FCT.
- (2) To plan the best approach to obtain data nutritionally useful food composition data.

$$\bar{x} \pm \sigma/\sqrt{n-1} \quad (1)$$

The priorities when including a food in a FCT should be the marketing, consumer patterns, etc., many researchers agree on

this point.^{52,50,xxx} After this phase, this information related to the market share or volume sales, among others, of each food kind, should not be used as proportional factors to decide the homogeneity of the food sample because this information does not directly affect food composition. That is, when the aim is to elaborate a databank at a national level, and a food is needed to be sampled in order to determine its composition, the proportionality of sales and consumption patterns are essential to decide which variety/ies to include or in which location/s the sampling should be carried out. However, once this decision has been made, the different kinds of samples included must be equally considered. Because, if at any of the experiment's phases it is appreciated that the results among samples differ significantly, the foods are different in terms of composition and the results will be presented separately. But if they do not differ, what is the need of keeping proportionality if they are the same?

On the other hand these studies are also a prime source of orientative information of the factors that could play a role in the food composition, such as those referring to the geographic source of food, which is essential to evaluate the possible influence of local factors; the seasonality in which they are sold or used for the industry; post harvest, slaughter or collection handling, including maximal storage time and conditions. The date and sampling beginning (at the field, wholesale or retail level, which is not necessarily the same for the different food groups) will be based on this information too. In this phase also, it must be decided as to which foods it is possible to obtain representative samples and which not. It would be useful to establish a marked difference between a "single" or "mixed" (multi-ingredient) food.¹⁷ Because it is easier to obtain representative samples from raw materials and those derived products which, even whilst containing another ingredient apart from salt, have a less variable formula and are also used as raw materials for domestic use, than from multi-ingredient recipes. Thus, in case of precooked dishes and other commercial "mixed" recipes, the actual information regarding the constituents of the elaborated product can be included in the label. However, one possibility of including foodstuff as eaten in FCTs, could be to selected by means of correct sampling, some recipes and the elaboration conditions of some of the more commonly consumed dishes by the population, according to their area of competence, and then process them according to their habits but under controlled conditions and, citing the ingredients, their proportions and the elaboration conditions, showing the analytical results of the precooked dish in the FCT.

$$\sigma/\sqrt{n-1} \quad (2)$$

To achieve the second phase the aim must be the nutritional usefulness of the data. And our proposal to obtain representative samples of the food is explained and developed in the following four milestones which overlap in time, and they are shown in Figure 1.

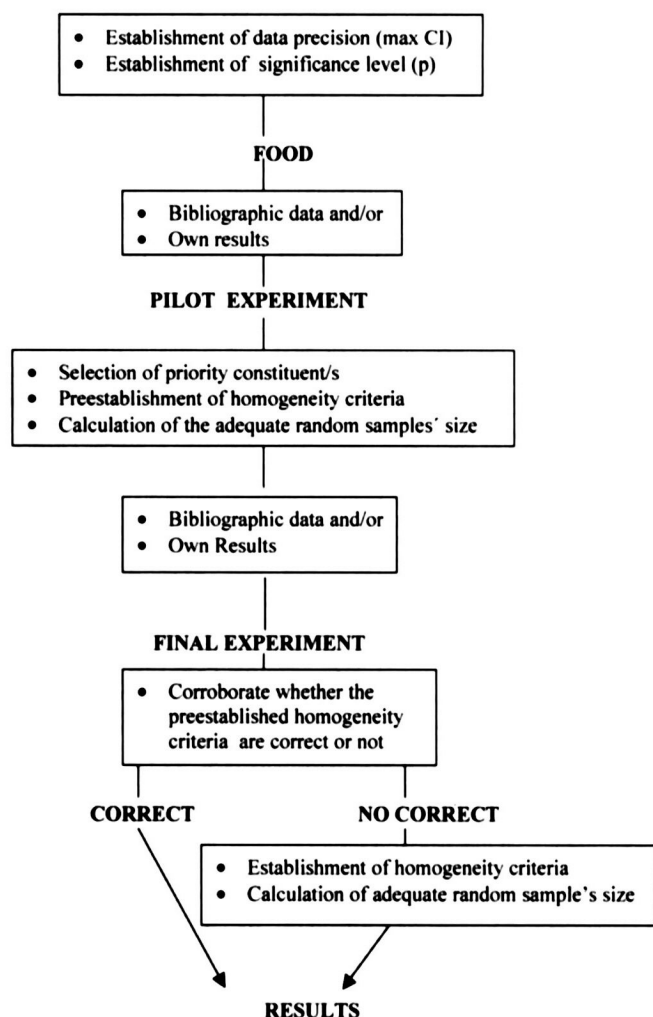


Figure 1 Proposal procedure to obtain representative samples.

To Formulate a General Vision of a Procedure to Establish Homogeneity Criteria of Sample.

If there is insufficient or no previous information on the composition of a food, a pilot experiment should be carried out on small samples. These first estimates help to establish the homogeneity criteria of the samples in relation to the factors of influence which are suspected to alter the concentration of the food constituents. For example, variety, cultivation in a greenhouse or in the open air, and the geographic origin (which involves the kind of soil, the type of water used for watering, etc); variety of breed confronted with intensive or extensive types of animal farming (which implicates a different feeding and the possibility, or not, of using growth enhancers), etc.

For the pilot experiment of this food, as many random samples (number of sample units $n \leq 10$) should be extracted as groups obtained by combining the factors of influence. If one food (a concrete species) is considered a population, these groups are subpopulations theoretically homogeneous in relation to the factors of influence, extracting one sample out of each of them.

Although provisional, due to the small sample size, the results of the analyses provide information about the population which constitutes that food, and based on the results it is decided whether to continue with the previously established subpopulations or regroup or subdivide them with other homogeneity criteria, or even not to subdivide the population because no important differences are appreciated.

Once the decision has been made, an arithmetic mean ($\bar{x}$) and a standard deviation (σ) will be obtained for each food's component of each apparently homogeneous sample (established on the basis of the results obtained in the pilot experiment), which will be used both to calculate the sample size suitable to carry out the definitive experiment. The definitive experiment will confirm whether the homogeneity criteria established with the results of the pilot experiment are correct or not. To carry it out, as many random samples as it has been decided in the pilot study must be extracted, but of size "n" (whose calculation will be later discussed).

Figure 2 shows a concrete case in which three varieties must be analysed: A, B, C, of a vegetable species X which is grown in both a greenhouse and in the open air, and whose major cultivars are located in two different areas of the country.

Because of the lack of sufficiently documented bibliographic information, a pilot experiment in which variety, type of cultivation and geographic source will be confronted, is proposed. From the results of the pilot experiment it follows that no difference seems to exist among varieties. Therefore they group together, but the differences between the type of cultivation and geographic source remain. Four subpopulations are thus constituted to carry out the definitive experiment.

During the course of this experiment it is appreciated that the small differences initially observed in the pilot experiment between the geographic source of both types of cultivation are not significantly important, but a meaningful and important difference among the analytical results of each kind of cultivation

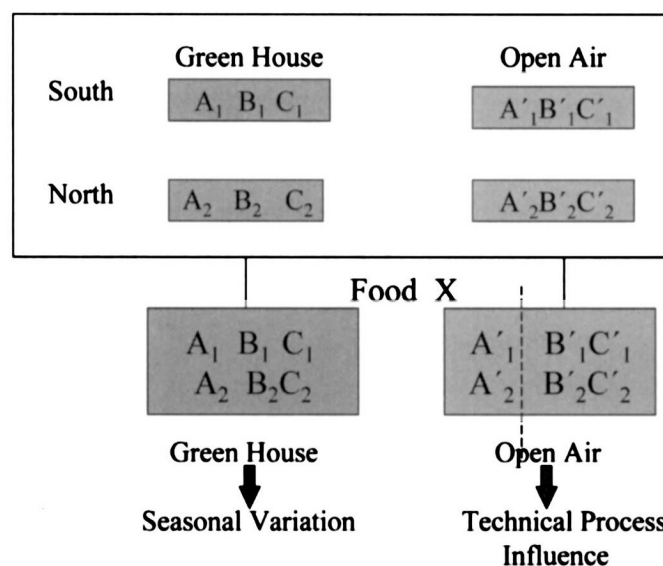


Figure 2 Sampling example to establish homogeneity criteria.

remains. It is decided to group together the samples of both geographic sources and the adequate sample sizes are again calculated using the $\bar{x}$ and σ obtained from the analysed units up to then from each kind of cultivation, which might be enough or not. When finishing the definitive experiment, if there is corroboration of significant differences between the results of both types of cultivation, this difference will be shown in the analytic results and presented in the FCTs.

Also, the varieties grown in the open air are seasonal, but those grown in greenhouses are cultivated all year long. Consequently, these last ones can be used to elaborate studies of seasonal variation. On the other hand, it is known that two of the varieties grown in the open air are used to make canned food of the product. Because of this it is possible to study the influence of the technological processes on them. The same procedure could be used to study the influence of storage time and conditions, manipulation, other heating processes, etc.

To Establish the Precision of Results and Calculate Adequate Sample Size

The desired precision of the results to be shown in the FCTs is decided at the beginning of the experiment; that is, the maximum width of the confidence interval (CI) which must accompany the arithmetic mean of each food constituent when the final results are shown in the FCTs, are established at the beginning of the experiment. The appropriate sample size (n) depends on the precision required.

To predetermine the appropriate "n" of the samples to carry out the definitive experiment, three known values are used the $\bar{x}$ and σ obtained from the pilot experiment, and the maximum limits of the CI established.

The CI for the means of the population is: (Formula 1), where (Formula 2) is the standard error; t is the Student corrector coefficient, it is tabulated and depends on the sample size and the established security level (for example, 95%). As " t " depends on " n " and " n " will be predetermined, $t = z = 2$ is established (This is an approximation to the normal distribution for 95% security).

Different Guidelines Used to Fix the Maximum Limits of the CI of Each Food Constituent:

- the first, which is based on general statistic criteria for biological distributions, establishes that it is usual to set the interval (Formula 1), so that the proportion represented by value (Formula 3) with respect to the population mean, which is estimated from the sample mean ($\bar{x}$), in no case exceeds 20% (relative error < 0.2).⁵²

$$t\sigma/\sqrt{n-1} \quad (3)$$

$$\frac{t\sigma/\sqrt{n-1}}{\bar{x}} < 0.2 \rightarrow t\sigma/\sqrt{n-1} < 0.2\bar{x} \Rightarrow \text{CI max.: } \bar{x} \pm 0.2\bar{x} \quad (4)$$

With the above mentioned approximation ($t = z = 2$), the following equality for the calculation of the adequate "n" of the

sample is established.

$$t\sigma/\sqrt{n-1} \approx 2\sigma/\sqrt{n-1} = 0.2\bar{x} \rightarrow n = \frac{2^2\sigma^2}{(0.2\bar{x})^2} + 1 \quad (5)$$

- The second guideline proposes that the maximum CI margins ($\bar{x} \pm Y$) are fixed according to nutritional criteria. It would be sufficient to establish the equality (Formula 6) and take out "n" in order to predetermine the appropriate sample size.⁴⁹

$$Y = 2\sigma/\sqrt{n-1} \quad (6)$$

It is important to base the maximum width ($\pm Y$) of the CI on nutritional criteria, since these will dictate the later statistical significance of the differences. That means, the basis to establish the significance or not of the differences found in the concentrations of each food component in the definitive experiment, will be nutritional based criteria. The goal must be to establish these maximum widths in accordance with the ranges of corresponding human nutrient requirement estimates because, if one of the principal uses of the data is to value and plan nutrient intake, it seems logical that the maximum range of variability (CI) in which data oscillate should be that dictated by requirements, always when it is possible. Otherwise, how can good assessments or recommendations be established, or acceptable estimates of nutrient intakes be made?

The key is to obtain not only satisfactory estimates of food composition data, but at the same time as Beaton²⁷ considers, the development of knowledge of human nutrient requirements of nutrients.

Selection of Prime Constituents of Sample

When the homogeneity criteria are set in the pilot experiment it should be decided in relation to an which component or components they are established, since each analysed sample has shows several.

Different situations might occur which should be resolved according to reasonable nutritional criteria. One of these could be to give priority to those components from which a specific food is the main contributor to the diet. Let us suppose, for example, the situation in which the protein concentration of a vegetable with a low protein content has to determined It might occur that, although the protein concentration oscillates above 20% of its mean ($\pm 0.2\bar{x}$), it does not exceed the acceptable preestablished CI for the protein and therefore, the sample, although dispersed "per se," is homogeneous in relation to the protein requirement range. This means that the protein does not need to be submitted to homogeneity tests in that particular food. The same may happen with other components of which the food is not a principal contributor, and this is one of the reasons for proposing the establishment of the maximum CI of each nutrient according to the nutrient requirements rather than according to the maximum interval $\bar{x} \pm 0.2\bar{x}$.

A different case would be that of a food which is the main contributor to the diet for two of its components both of them suffering significant important variations in their concentration

due to the influence of one or more environmental factors. If these concentration fluctuations are so discordant that it is impossible to establish joint homogeneity criteria, the solution will be to establish independent homogeneous subpopulations for each main component in relation to their influence factors or factors, expressing this subdivision and the results obtained in the FCT.

A similar situation is found when calculating the appropriate sample size. Logically, the same component as that for establishing the homogeneity criteria will be selected and, if there were several of these, it would be reasonable to predetermine that what requires the largest sample size; that is, the more greatly dispersed (greater variation coefficient) or less secure (greater probability of error "p").⁴⁹

Another factor that has not been considered until now which can affect the homogeneity criteria, is that of water. Although the water content influences the organoleptic properties of a food, as it is a component of the same it would be absurd to establish an acceptable maximum CI for its concentration. In this case it is more useful, for signification reasons, to take the CI max.: $\bar{x} \pm 0,2\bar{x}$ according to statistical criteria.

It could be that following the analytical protocol for determination of the humidity content in different samples from the same food, the water content varies significantly from one samples to others. In these cases, and since FCT results refer to an edible portion of 100 g, it would be correct to give this information. It is essential that the food's water content is shown in FCTs in order to be able to correctly identify and compare the foods.³²

Proposal of Statistical Probes to Obtain Homogeneity Criteria and Proper Final Presentation of Results

In relation to the statistical tests used to establish homogeneity criteria, current computer programmes facilitate the application of different tests depending on how the data obtained adjust to normal distribution.

If representative samples are achieved, it is to be expected that the adjustment is acceptable even with samples that are less than thirty units ($n < 30$), and it would be sufficient to apply the Student corrector factor. However, if all attempts fail, the median and non-parametric tests can be used. These will permit two or more samples to be compared in block, although with less precision. On the other hand, these non-parametric tests can be useful at the pilot study when, having small samples, it is intended to contrast hypotheses in order to approximately establish the first homogeneity criteria.

For normal distributions, the computerized analysis of variance offers different possibilities for block evaluation of a sample group's homogeneity with respect to one or several influencing factors. This test will be complemented with others that particularize the difference between means minimizing the error, and with the covariance analysis which permits the detection of the influence of underlying factors, if this happens to be necessary.

It should be remembered that from each "random sample" (constituted by "n" units) as many "simple random samples" (constituted by "n" units) as components analysed are obtained,

and that confrontations between samples are carried out for each component separately, with the above mentioned priorities, milestone 3. Obviously, from each unit of simple random sample will be extracted so many aliquots needed such as the corresponding analytical protocol requires.

At the end of the definitive experiment, an arithmetic mean $\bar{x}$ and a standard deviation σ are obtained for each component. Since the data are useful as being attributable to other similar portions of the food which is being analysed, the results obtained from the samples should be generalised. This generalization is based on another concept (mentioned before): the standard error (Formula 7), which measures the imaginary dispersion which would show the successive means obtained given the hypothetical repetition of the experiment.⁵¹ Therefore, if the mean of the population is estimated from the sample mean, a dispersion range must accompany this, and this is the confidence interval.

$$(S_m = \sigma / \sqrt{n - 1}) \quad (7)$$

The CI of 95% ($p < 0,05$) for the mean of the population is $\bar{x} \pm 2S_m$ in the Gauss Curve, and $\bar{x} \pm tS_m$ for small samples ($n < 30$).

If all the homogeneity and appropriate sample size criteria previously mentioned have been followed, this confidence interval should not exceed in terms of width the maximum CI prefixed for each food component. From the above mentioned it can be deduced that the recommended final presentation of results are three data: the mean, the CI and the established security level.^{51,53} This last one could be indicated in the general form at the beginning of the work, and it is enough.

(Formula 8), along with their units of measurement.

$$\bar{x}(\text{CI } 95\%: \bar{x} - tS_m \text{ to } \bar{x} + tS_m) \quad (8)$$

CONCLUSIONS

It is recommended for data accuracy in relation to sampling, that the importance of studying and getting to know food populations be stressed before the presentation in FCTs or inclusion in databanks of their component concentrations. When necessary, specify any important nutritional differences encountered and their effect in the corresponding section for each food.

ACKNOWLEDGEMENTS

We want to express our sincere thanks to Mrs Elena Primo for her kind help in the documentation of the work.

REFERENCES

- Applications in dietary assessment: A report of the interpretation and uses of dietary reference intakes and upper reference levels of nutrients. Institute of Medicine. 2000. Washington DC: National Academy Press.
- Beaton, G.H., Interpretation of results from diet history studies. In: *The Diet History Method*. Ed. L. Kohlmeier. London: British Library Cataloging in Publication Data applied for, 1991.

- Butrum, R.R. 1992. Overview of international food composition data bases. *World Rev. Nutr. Diet.*, **68**:1-14.
- Buzzard, I.M., Pricek, S., and Feskanich, D. Preparation for dietary data analysis: How the structure of data base influences data collection and analysis and vice versa. In: *The Diet History Method*. Ed. L. Kohlmeier. London: British Library Cataloguing in Publication Data applied for, 1991.
- Carrasco, J.L. 1995. *El Método Estadístico en la Investigación Médica*. Sexta Edición. Madrid: Editorial Ciencia 3.
- Carriquiry A.L. 1999. Assessing the prevalence of nutrient inadequacy. *Public Health Nutr. Mar.*, **2**:23-33.
- Chamodiere, U.R., Vignat, J., Møller, A., Ireland, J., Becker, W., Church, S., and Farran, A., et al., 2002. European Nutrient Database (ENDB) for Nutritional Epidemiology. *J. Food Comp. Analysis*, **15**:435-451.
- Chatefield Charlotte., 1955. *Tablas de Composición de Alimentos* (Minerales y Vitaminas). FAO. Roma.
- Chatefield Charlotte., 1965 *Tablas de Composición de Alimentos* (Minerales y Vitaminas). FAO. Roma.
- Chu, N., Clydesdale, F.M. 1975. The effect of concentration, thermal processing and storage temperature on the interaction between alpha-ketoglutaric acid and tryptophan. *J. Milk Food. Technol.*, **38**:573-580.
- Clydesdale, F.M. 1997. A Proposal for the Establishment of Scientific Criteria for Health Claims for Functional Foods. *Nutr. Rev.*, **55**:413-422.
- Clydesdale, F.M., Ho, C.T., Lee, C.Y., Mondy, N.J., and Shewfelt, R.L. 1991. The effects of post harvest treatment and chemical interactions on the bioavailability of ascorbic acid, thiamine, vitamin A, carotenoids and minerals. *Crit. Rev. Food. Sci. Nutr.*, **30**:599-638.
- Combs Jr., G.F. 1996. Should Intakes with Beneficial Actions, Often Requiring-Supplementation, Be Considered for RDAs? *J. Nutr.*, 2373S-2376S.
- Deharveng, G., Charrondière, V.R., Slimani, N., Southgate, D.A.T., and Riboli, E. 1999. Comparison of nutrients in the food composition tables available in the nine European countries participating in Epic. *Eur. J. Clin. Nutr.*, **53**:60-79.
- Dietary Reference Intake for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride. Institute of Medicine. 1999. Washington DC: National Academy Press.
- Dietary Reference Intake for Thiamin, Riboflavin, Niacin, Vitamin B₆, Folate, Vitamin B₁₂, Pantothenic Acid, Biotin, and Choline. Institute of Medicine 1999. Washington DC: National Academy Press.
- Dietary Reference Intake for Vitamin A, Vitamin K, arsenic, boron, chromium, molybdenum, nickel, silicon, vanadium, and zinc. Institute of Medicine. 2001. Washington DC: National Academy Press.
- Dietary Reference Intake for Vitamin C, Vitamin E, Selenium, and Carotenoids. Institute of Medicine. 2000. Washington DC: National Academy Press.
- Dietary Reference Intakes for Energy, Carbohydrates, Fiber, Fat, Protein and Amino Acids (Macronutrients). Institute of Medicine. 2002. Washington DC: National Academy Press.
- Dietary Reference Intakes: Proposed Definition of Dietary Fiber. Institute of Medicine. 2001. Washington DC: National Academy Press.
- Doménech, J.M. 1996. *Métodos Estadísticos en Ciencias de la Salud*. Barcelona: Editorial Gráficas SIGNO.
- Dwyer, J.T. 1994. Future directions in Food Composition Studies. *J. Nutrition.*, **124**(9 Suppl):1783S-1788S.
- FAO. 1970. *Amino-acid content of Food and Biological data on proteins*. Roma.
- Feinberg, M., Ireland-Ripert, J., and Favier, J.C. 1992. Validated data banks on food composition: Concepts for modelling information. *World Rev. Nutr. Diet.*, **68**:49-93.
- Ferro-Luzzi, A., and James, W.P.T. 2000. European diet and public health: The continuing challenge. In: Kafatos, A., and Codrington, C.A. Eds. Heraklion, Crete, Greece: *Eurodiet Working Party Reports*, 18-20.
- Gonçalves Ferreira, F.A., and Da Silva Graça, M.E. 1961. *Tabela da Composição dos Alimentos Portugueses*. Direcção-Geral de Saúde. Instituto Superior de Higiene Dr. Ricardo Jorge. Lisboa-Porto.
- Haytowitz, D.B. 1995. Information from USDA's Nutrient Data Bank. *J. Nutr.*, **125**:1952-1955.
- Hendrick, S., Kwang-Won, L., Xia X., Huei-Ju, and Murphy, P.A. 1994. Defining Food Components as New Nutrients. *J. Nutrition.*, **124**(9 Suppl):1789S-1792S.
- Holden, J.M. 1996. Expert Systems for the Evaluation of Data Quality for Establishing the Recommended Dietary Allowances. *J. Nutr.*, **126**(Suppl):2329S-2336S.
- Holden, J.M. 1995. Food sampling strategies for energy intake estimates. *Am. J. Clin. Nutr.*, **62**(Suppl):1151S-1157S.
- Joyanes, M^a.G., 1979. Estudios sobre el valor nutritivo de la proteína demejillón y de su concentrado proteico. Madrid: *Fundación Juan March, Serie Universitaria n° 97*.
- Joyanes, M., González-Gross, M., and Marcos, A. 2002. The need to review the Spanish recommended dietary energy and nutrient intakes. *Eur. J. Clin. Nutr.*, **56**:899-905.
- Katch, F.I. 1995. U.S. government raises serious questions about reliability of U.S. Department of Agriculture's Food composition database. *Int. J. Sport. Nutr.*, **5**:62-67.
- Kübler W. 2001. Referenzwerte und die Auswertung von Verzehrerhebungen Zum Konzept der Dietary Reference Intakes (USA und Kanada). *Ernährungs-Umschau*, **48**:476-479.
- Lema, L., Mataix, F.J., and Varela G. 1984. Effect of cooking on the nutritional value of mussel protein (*Mytilus edulis*) with respect to seasonal variations. *Proc. MOCCA*, **1**:47-50.
- Lema, L., Mataix, F.J., and Varela, G. 1984. Effect of Processing in the Preparation of Preserves on the Nutritional Value of Mussel Protein (*Mytilus edulis* L.). *Proc. MOCCA*, **1**:43-46.
- Lema, M^a.L., Navarro, M^a.P., Mataix, F.J., and Varela, G. 1986. Influencia de los procesos de cocción y desecación a distinta temperatura sobre el valor nutritivo de la proteína del mejillón (*Mytilus edulis*). *Archiv. Latinoam. Nutr.*, **3**:495-504.
- Lupien, J.R. 1997. The Food and Agriculture Organization food-composition initiative. *Am. J. Clin. Nutr.*, **65**(suppl):1194S-1197S.
- Lupien, J.R. 1994. Dietary assessment issues of concern to policymakers: Statement from the Food and Agriculture Organization of the United Nations. *Am. J. Clin. Nutr.*, **59**(1 Suppl):269S-270S.
- Murphy, S.P. 2002. Dietay reference Intakes for the U. S. and Canada. Update on implications for Nutient Databases. *J. Food Comp. Analysis*, **15**:411-417.
- Paul, A.A., and Southgate, D.A.T. 1978. *McCance & Widdowson's The Composition of Foods*. London: Ministry of Agriculture, Fisheries and Food Medical Research Council. Elsevier/North-Holland Biomedical Press.
- Peterson, J., Beecher, G., Bhagwat, S., et al., 2001. Should we Aggregate or Disaggregate the Flavonone Data of Oranges, Mandarins, and Mandarin Hybrids? *4th International Food Data Conference. Bratislava, August*, 42-26. *Summary Books*, P-1-C-10.
- Rand, W.M. 1987. INFOODS y los datos de composición de alimentos. *Arch Latinoam Nutr.*, **37**:609-617.
- Scrimshaw, N.S. 1997. INFOODS: The international network of food data systems. *Am. J. Clin. Nutr.*, **65**(suppl):1190S-1193S.
- Smit L.E., Schönfeld, H.C., de Beer, W.H.J., Smith, M.F. 2000. The Effect of Locally and Season on the Composition of South African Whole Milk. *J. Food.*, **13**:345-367.
- Souci, S.W., Fachmann, W., and Kraut, H. 2000. *Food Composition Tables and Nutrition*. 5th Ed. Stuttgart: Medpharm, CRC Press, SVC P.
- Souci, S.W., Fachmann, W., and Kraut, H. 1994. *Food Composition Tables and Nutrition*. 5th Ed. Stuttgart: Medpharm, CRC Press, SVC P.
- Southgate D. 2002. Data Quality in Sampling, Analysis, and Composition. *J Food Comp. Analysis*, **15**:507-513.
- Southgate D.A.T., and Greenfield, H. 1992. Principles for the Preparation of Nutritional Data Bases and Food Composition Tables. In: International Food Data Bases and Informations Exchange. Eds. Simopoulos, A.P., and Butrum, R.R., Basel, Karger, *World Rev. Nutr. Diet.*, **68**.
- Srivastava, S., and Butrum, R. 1992. International food data base: Conceptual desing. *World Rev. Nutr. Diet.*, **68**:15-26.
- Tricopoulou, A. 1992. Monitoring food intake in Europe: A food data bank on household budget surveys. *Eur. J. Clin. Nutr.*, **62**(Suppl):1151S-1157S.
- Varela, G., Joyanes, M., and Torralba, A. 1977. Seasonal variations in the nutritive value of the mussel protein (*Mytilus edulis*). *Proc. Nutr. Soc.*, **36**:84.