INTEGRATING NUTRIENT DATA WITH FOOD FREQUENCY DATA

Elizabeth Randall, Ph.D., R.D.
Assistant Professor
University of Texas School of Public Health

It was approximately five years ago that I began attending the National Nutrient Data Base conferences. Actually my first exposure to this group was at the meeting held in Philadelphia. My memory is very accurate about that because it was truly a mind-opening experience. At that time I was developing a research program with a focus in dietary methodology and a nutrient data base was an important tool. Prior to that meeting, I was unaware of the enormous number of considerations to be made in developing a nutrient data base that bear directly on the validity of dietary measurements.

Since that time, the field of dietary methodology has shown a fascination with the food frequency measure, largely fueled by the needs of the nutritional epidemiologist. Nutrient data bases have ceased to be an important tool and have become an integral feature in the design of the food frequency measure.

Despite variation in settings in which the food frequency technique is now used, there are some common elements to the questionnaires. First, all food frequency measures have a list of foods. Secondly, respondents are queried about their frequency of use of these foods, over some specified period of time. The final characteristic is the quantification of portion sizes. Some frequency instruments disregard quantification and estimate merely how often foods are consumed. Others are described as semi-quantitative in that they specify a typical portion size and provide the respondent with limited options for adjusting this amount. Still other food frequency measures are fully quantified by collecting accurate data on usual portion sizes from the individual respondents.

For the analysis of data collected using a food frequency measure, many of the previous concerns about nutrient data bases remain - availability of accurate nutrient values in their metabolically active forms and recipe calculations - to name only two. Some problems diminish in relevance - coping with the enormity of an extensive nutrient data base and accuracy in food code assignment by data collectors, for example. There are new problems to face. Empirical data is desirable for the identification of foods to include on a food frequency form. Composite food items are commonly developed from the highly specific food items contained in the nutrient data base. Nutrient values need to be assigned to these composite foods and this need is often accompanied by a need for typical portion sizes.

The purpose of this presentation is to understand the methodological issues involved in developing a nutrient data base to analyze dietary intake data collected using the food frequency technique.

1. PURPOSE FOR DATA COLLECTION AFFECTS DATA BASE NEEDS

A. Epidemiological Studies

Needs of the nutritional epidemiologist have dominated the development of the food frequency technique (1-7). It is not surprising, therefore, that attention has focused on the identification of foods whose use supports the epidemiologist's need for categorization of individuals into levels of nutrient intake. Important food sources of nutrients within populations have been identified (8-9), as have foods that account for the variability in nutrient intakes in a population (10). Since the goal in data collection is to accurately classify individuals on the basis of their usual intake levels of nutrients, precise quantification of dietary intake has not been stressed (11).

B. Food Habits Identification

The food frequency instrument has been used by nutritionists for many years to describe patterns of food use over varying periods of time (12-15). Since it is known to provide qualitative data only (16), it commonly is used as a cross-check for more thorough methods of dietary intake assessment. It has also been used to develop profiles of food intake in order to distinguish individuals or groups on the basis of patterns of food use (17). The information obtained may be general (for instance identifying "meat and potato eaters" versus "vegetable consumers"); specific in terms of the foods whose use is described (such as the use of designated, highly processed foods for GRAS classifications) (18); or highly specific in terms of how foods are prepared, to determine differences in practices that could influence health.

C. Nutrition Intervention Programs

Some programs are designed with specific dietary changes, or goals, planned. The food frequency measure may be used to supply evidence of the change process occurring. For this purpose, the instrument would have foods specified as desirable to change, or may have form or preparation of food as the targeted change. Therefore the foods on the food frequency list may be highly specific, or very general when used in connection with an intervention program.

2. QUESTIONNAIRE CONSTRUCTION AFFECTS NUTRIENT DATA BASE NEEDS

Several approaches have been followed by investigators in selecting the foods to include on a food frequency questionnaire. One approach is to select foods from an existing data base, or table of food composition, that are dense in the nutrient(s) of concern (19). The individual foods identified would, because of their selection process, have accurate nutrient data available from existing nutrient data bases.

More recently investigators have reasoned that it is preferable to ask about the frequency of use of foods known to be consumed by the population. Important food sources of nutrients are defined as those making the greatest percent contribution to the intakes of targeted nutrients by the population (8,9). Some interest has been expressed in constructing a single food fre-

quency questionnaire for widespread applicability in nutrition research. Important food sources of most major nutrients are identified using national survey data (either USDA surveys of NHANES II). Individual foods that are highly similar in composition and role in the diet are grouped to form small food groupings, referred to as composite foods. The process results in a relatively short list of 25-125 foods that, in total, account for the population's intakes of the nutrients. This approach, however, requires information concerning appropriate nutrient values for these composite foods, values that must be calculated rather than extracted from an existing nutrient data base.

Some concern exists regarding the feasibility of one food frequency instrument designed for universal application. An alternative approach would be to identify the important food sources of nutrients within a population that is highly comparable to the study population (20). Such an approach is particularly attractive in an area, such as Texas, where distinctive regional or ethnic eating practices flourish. Investigators would need to acquire background data concerning foods actually consumed, and analyze them using conventional nutrient data bases. A food frequency questionnaire could then be constructed using the important food sources of nutrients for the specific population. A noteworthy advantage of this approach is that more detailed information is then available about the actual foods included in the composite food items, as well as their relative importance within the grouping.

3. CONSTRUCTING A NUTRIENT DATA BASE TO SUPPORT DATA ANALYSIS

These decisions, made at the time of questionnaire design, hold important ramifications for the construction of a nutrient data base to analyze the food frequency data. Questionnaire-specific nutrient data bases are advantageous, if not essential, to construct. The primary reason for believing so is that the food frequency technique makes use of a structured questionnaire format that defines the foods about which the respondent will be queried. There is no need, therefore, to assume the high costs and other inconveniences of a very large nutrient data base when nutrient values for only those foods on the questionnaire are relevant. Secondly, when a food list contains composite food items, there may be no valid nutrient values for these foods in existing nutrient data bases. Decisions to be made involve assigning nutrient values to the foods, and to issues of quantification of portion sizes.

A. Nutrient Values of Foods

1. Representative Foods

When the decision is to include foods that have been identified as used by, and important to the nutrient intakes of a specific population, the assignment of a food code is similar to this same activity when using a standard nutrient data base. You identify the most appropriate food code and accept the accompanying nutrient values.

Identifying the use of combination dishes or dishes made from special recipes, continues to defy resolution. Because the important food sources of nutrients are identified using standard techniques, such as the 24-hour recall or food record, the method used to code these special items will influence whether or

not they can be identified as unique food items. When a mixed dish is coded as such, it has the potential of being identified as an important food source of nutrients for a food frequency questionnaire. However, when it has been coded by individual ingredients in a recipe, unless special coding procedures are adopted, only its ingredients can be identified as important food sources. It then becomes a matter of subjective decision-making to develop a list of foods for a food frequency questionnaire that accurately reflects important food sources that can be asked of the respondents in a manner that allows them to supply accurate data. For example, tomatoes may be identified as an important food source of vitamin C. When recipes have been coded by ingredients, the interviewer may have to ultimately ask the question "How often do you consume tomatoes in any form?" Familiarity with the actual food data may permit the investigator to realize that the tomatoes refer to tomatoes included in casseroles. The interviewer could then be more directed with the question "How often do you consume casseroles containing tomatoes?". This second question provides more focus for the respondent, thereby supporting a more accurate estimate of intake. Unfortunately such focus needs to be supported by empirical evidence.

2. Composite Foods

The assignment of nutrient values to composite foods presents more complex issues. In the identification of important food sources of nutrients, individual food codes in the data base are rank ordered on the basis of their percent contribution to the population's intake of the nutrient. Extensive lists of foods are generated. The investigator then must collapse the highly differentiated food items appearing in large nutrient data bases into groupings of comparable foods.

This procedure can be relatively simple for some foods. For example, natural cheese could consist of cheddar, Swiss, and colby cheeses. A serving of cheese could be defined as 2 oz. and the nutrient values of the three types of cheeses averaged to yield nutrient values for the composite food.

Other composite foods generate more serious difficulties. The simple carrot may not be that simple. Carrots are identified in NHANES II data as an important food source of vitamin A in the U.S. population. There are many forms of carrot identified, including raw, frozen, canned. Peas and carrots are identified as a source of vitamin A. It seems reasonable to decide that peas and carrots should be grouped with carrots as a source of vitamin A. A small extension of the argument permits the inclusion of mixed vegetables, canned as well as frozen, with carrots. Beef stew is identified as an important food source of vitamin A as well. It is then a subjective decision as to whether or not beef stew would form a unique group or included as carrots with a question such as "How often do you eat carrots, including carrots in stew?"

This process for defining the food group "carrots" is repeated for many other groups. It is a point at which subjectivity is included in questionnaire design, and as such, is a point that warrants close scrutiny. At this time there is no way to form the composite foods with purely empirical evidence. It is advantageous to include foods in the form in which they actually are consumed, a goal that at least partially justifies the acceptance of this subjectivity.

Once the composite foods are formed, the need is to assign nutrient values to them. One technique would be simply to average the values for each of the items included in the composite food. In some instances, such as the example of natural cheese, this procedure would be justifiable. It would be justified when each food has similar nutrient values and is approximately equally commonly consumed within the population.

In other cases a more intricate procedure would be recommended. A weighting scheme would be preferable when individual food items differ in nutrient composition and/or in the proportion of respondents consuming them. When weighting is instituted, the nutrient values of the composite food would be computed by weighting the values for the individual items by the relative contribution of the item to the consumption of all items comprising the composite food.

The ability to compute valid weighted nutrient values depends on the effective sample size on which the weighting is based. Some food items are very dense in particular nutrients. Liver, for example, is a very dense source of vitamin A. Such foods will probably be identified as an important nutrient source if only one person consumes it. In this case the effective sample size would be one. The estimated contribution to the intake of vitamin A would be highly unstable since it would double if one more person in the population were to consume it in an equivalent serving size.

There is no accepted method for computing the weighted values, or to know the sample size needed to develop stable estimates of the important food sources and corresponding accurate nutrient values. Currently I have the thought that to circumvent some of these difficulties, it would be advantageous to identify the important food sources of nutrients using one of the numerous, smaller nutrient data bases. The effect of doing so would be to force the interviewer, who actually collected the dietary data, to code foods using a more restrictive number of codes. This procedure would eliminate some of the subsequent subjectivity involved in forming the composite foods.

3. Dietary Goals

There are occasions when a food frequency questionnaire could be constructed using dietary goals rather than actual food consumption patterns. A nutrition intervention program may have an objective such as the reduction in the use of high fat, high sodium foods, or simple carbohydrates, or they may encourage the consumption of high fiber foods or complex carbohydrates. The food frequency technique has been used to measure the success of such programs. The identification of foods for inclusion on the questionnaire would not come from nutrient data bases, but from targeted, feasible behavioral changes. Foods could be selected in general form, such as whole grain cereals or processed meats. Also, relevant behaviors could include food preparation techniques.

A nutrient data base for this type of questionnaire would have nutrient values developed for these general items. It would be desirable to have such values reflect market share information. The data base would also have to accommodate precision in recording how foods are prepared, since food preparation may be a major point of intervention. An additional problem that may not be immediately apparent is that of obtaining nutrient data for new food

products. Such programs frequently involve the adoption of dietary changes that the food industry is simultaneously supporting.

B. Quantification of Dietary Data

Food use data, when coupled with portion size data, can be used to estimate nutrient intake levels. Data analysis programs can be developed that compute nutrient intake values by assigning portion size values for each of the foods and automatically converting the frequency of use data to nutrient intake estimates. Some frequency questionnaires assign typical portion sizes, others obtain the quantification of food intakes from the respondents themselves.

1. Typical Portion Sizes

When either an unquantified or a semiquantitative food frequency instrument is used, level of nutrient intake can be estimated by applying typical portion sizes. These data, fortunately, are readily available in the U.S. from the national surveys utilizing national probability sampling, either USDA food consumption surveys or the NHANES surveys. These data have been published in several formats. The computer tapes are also available and can be used to compute typical portion sizes for many additional subgroups in the population.

A food frequency questionnaire can specify portion sizes, and if the acutal serving differs substantially from that, adjustments can be made in the reported frequency of use. NCI makes use of a semi-quantitative questionnaire that identifies a typical portion size and provides the respondent the opportunity to report if the the usual serving would be described as "small", "medium", or "large" relative to this. The computer analysis program would then have merely three options to use to convert the food use data to nutrient intake estimates.

Three options may not be sufficient in order to obtain accurate estimates. National survey data provides ample evidence that males and females differ significantly in their portion sizes, as do younger and older persons (21). There may also be differences attributable to race in portion sizes for specific foods. It would be a relatively simple extension of the established procedures to modify the data analysis programs to accommodate a larger range of typical portion sizes. Of course known portion sizes for specific populations, should that information be available, could also be used in constructing the data analysis program.

2. Calculated Portion Sizes

Other questionnaires may collect actual portion size estimates from the respondents, themselves. Expressed in another manner, a food frequency questionnaire can have less structure. A semi-structured questionnaire is able to collect more individualized data, but it does also restrict the amount of structure that can be built into the nutrient data base prior to analyses. Familiar problems to the seasoned nutrient data base user reappear. Respondents will vary in the units of measure they use in reporting portion size. A less structured data base would be needed that could convert several different units of measure to gram weights before calculating nutrient values. The more

structured data base would have had the units predetermined and dealt with by merely expressing portion size as some multiple of these.

At this point a serious problem emerges for questionnaires in which portion sizes are collected for composite foods. The individual items included in the composite food commonly are identified with the gram weight of their average portion size in the population. This gram weight is highly influenced by the moisture content of the particular form in which the food is identified. For example, the gram weight of a portion of a food that is drained prior to serving is substantially smaller than a portion of an equal amount of solids served without draining. The designer of a nutrient data base for food frequency data needs to be careful in the conceptualization phase if this problem is to be accommodated.

I would suggest that one possible solution would be to express the portion size in household measures rather than as gram weights. This procedure would not be totally satisfactory for those composite foods that differ strongly in form. For example, if a food item, "tomatoes", were to include tomatoes in all its forms, it could include the units of measure of slices, measuring cups, glasses. Clearly these measures cannot be averaged. Their gram weights could be averaged, but the result would be very difficult to interpret.

No adequate method has yet emerged to resolve these problems. Anyone working in this field needs to be careful in making decisions that will influence the validity of the nutrient data base. In terms of the structure, nutrient data bases designed to analyze food frequency data have a simplicity to be envied by investigators working with the more comprehensive ones. On the other hand, nutrient data bases for food frequency measures place rigid demands a clarity in conceptualization. Many of the questions to be answered in designing data bases for food frequency measures have not yet even been identified. The National Nutrient Data Base Meeting provides a good forum for their identification and resolution.

REFERENCES

1. Abramson, JH, C Slome, and C Kosovsky. Food frequency interview as an epidemiologic tool. Am. J. Pub. Health 53:1093, 1963.

2. Hankin, JH, RA Stallones and HB Messinger. A short dietary method for epidemiologic studies III. Development of questionnaire. Am. J. Epidemiol. 87:285, 1968.

3. Mettlin, CJ and S Graham. Methodological issues in etiologic studies of diet and colon cancer. Nutr. Cancer 1:46, 1978.

4. Keys, A. Dietary survey methods. In: Nutrition, Lipids, and Coronary Heart Disease, R Levy, B Rifkin, B Dennis and N Ernst (Eds). New York: Raven Press, 1979.

5. Beaton, GH. What do we think we are measuring? Proceedings of the Dietary Data Collection, Analysis and Significance. Univ. Mass. Res. Bull. 675:36, 1982.

6. Rasanen L and P Pietinen. A short questionnaire method for evaluation of diets. Prev. Med. 11:669, 1982.

7. Lee, J, LN Kolonel and JH Hankin. Cholesterol intake as measured by unquantified and quantified food frequency interviews: Implications for epidemiologic research. Int. J. Epidemiol. 14:249, 1985.

8. Block, G, CM Dresser, AM Hartman and MD Carroll. Nutrient sources in the American diet; quantitative data from the NHANES II survey. I vitamins and minerals. Am. J. Epidemiol. 122:13, 1985.

9. Block, G, CM Dresser, AM Hartman, and CD Carroll. Nutrient sources in the American diet; quantitative data from the NHANES II survey. II macronutrients and fat. Am. J. Epidemiol. 122:27, 1985.

10. Byers, T, J Marshall, R Fiedler, M Zielezny and S Graham. Assessing nutrient intake with an abbreviated dietary interview. Am. J. Epidemiol. 122:41, 1985.

11. Block, G. A review of validations of dietary assessment methods. Am. J. Clin. Nutr. 115:492, 1982.

12. Young, CM and MF Trulson. Methodology for dietary studies in epidemiological surveys. II Strengths and weaknesses of existing methods. Am. J. Pub. Health. 50:803, 1960.

13. Stefanik, PA and MF Trulson. Determining the frequency intakes of foods in large group studies. Am. J. Clin. Nutr. 11:335, 1962.

14. Axelson, JM and MM Csernus. Reliability and validity of a food frequency checklist. J. Am. Dietet. Assoc. 83:152, 1983.

15. Mullen, BJ, NJ Krantzler, LE Grivetti, HG Schultz and HL Meiselman. Validity of a food frequency questionnaire for the determination of individual intake. Am. J. Clin. Nutr. 39: 136, 1984.

16. Sanjur, D. Food consumption surveys: Issues concerning the process of data collection. Chap 7 in: Social and Cultural Perspectives in Nutrition. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1982.

17. Davis, MA, E Randall, RN Forthofer, ES Lee and S Margen. Living arrangements and dietary patterns of older adults in the United States. J. Gerontol. 40:434, 1985.

18. Department of National Health and Welfare (Canada). Nutrition: A national health priority. Ottawa: Information Canada, 1973.

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

20. RS McPherson, E Randall, MZ Nichaman and GR Newell. Ethnic differences in food sources of nutrients of Texas residents: Implication for use of food frequency techniques in epidemiologic research. A paper presented at the Annual Meeting of the American Public Health Association, Washington, D.C., November 20, 1985.

21. Randall, E., MZ Nichaman and WD Spears. Variability in important food sources energy, fat and vitamin A in the U.S. population. A paper presented at the XIII International Congress of Nutrition, Brighton, England, August 21, 1985.