# DEVELOPMENT OF A DATABASE FOR THE NUTRITION CRSP PROJECT

**Suzanne P. Murphy**
**Doris H. Calloway**

Department of Nutritional Sciences
University of California, Berkeley

Investigators conducting international nutrition studies, particularly those conducted in less developed countries, may have difficulty obtaining accurate nutrient values for the foods consumed. The Collaborative Research Support Program on Food Intake and Function (the Nutrition CRSP) involved studies in rural villages in Kenya, Egypt, and Mexico (Calloway et al., 1988). The purpose of the Nutrition CRSP was to investigate the effect of mild to moderate malnutrition on function. At each of the three locations, two days of weighed food intake per month for a year were gathered for up to 5 target individuals in each household as well as for the household as a whole. Data collection occurred primarily in 1984 and 1985. Thus, nutrient data were required for local foods in each of these locations. Last year at this meeting, Judith Ricci told us about the development of a nutrient data base for the Egypt food intake data (Ricci et al., 1989). Likewise, a data base was developed for Kenya foods (Murphy and Calloway, 1990) and an existing data base was modified and expanded for Mexico foods (Allen et al., 1987).

At the time of the initial analyses, focus was primarily on intakes of the macro-nutrients: energy, protein, fat, and carbohydrate, although the three country projects had additional nutrients on their data bases. Preliminary results indicated potential relationships between dietary quality, as well as quantity, and functional outcomes in these countries. Thus, interest grew in obtaining information on a much wider variety of nutrients in the various diets. As a result, we became interested in developing a single nutrient data base, with a large number of nutrients, which could be used in all three locations - with the potential to expand to other countries as the need arose.

Prior to beginning development of the new nutrient data base, we examined food intake data from each of the three countries, to determine the types of foods available in each country and the frequency with which each was consumed. In Egypt, staples included wheat and rice; a typical diet of a mother and her toddler included bread as well as a chicken dish (Table 1). In Mexico, corn tortillas are a staple, and a typical diet had less variety than those in Egypt (Table 2). In Kenya, diets were even less varied - a typical diet might have only three or four items, as shown in Table 3. On this day, the mother had only tea, porridge, and githeri, made of the staples, maize and beans. On the day shown, milk was the only animal-source food consumed by either the mother or toddler.

After compiling data on the frequency of consumption of all foods, we were ready to begin work on the nutrient data base itself. We used several guidelines when conceptualizing the data base. (1) A key to the ease of use and maintenance of the data base was that it contain a fairly small number of basic foods and ingredients. Consumption of mixed dishes is very common in these countries, so each country project collected detailed data on recipe ingredients. Thus, the nutrient data base did not need to contain values for many mixed dishes - only values for the ingredients that composed the mixtures. (2) Via substitutions and combinations, these basic foods would provide estimates of nutrient data for all foods reported in the three countries. (3) There would be no missing values on the data base - when analytic data were unavailable, a value would be imputed.

As a starting point, we used the UCB Minilist nutrient data base (Murphy and Calloway, 1986), which currently contains 48 nutrients and 234 food items. The Minilist was developed using these same general guidelines (Pennington, 1976), and has been used by students and for research projects for the past 15 years. It has been continuously updated and revised over that time, and was updated just before we undertook this project, using the revisions of USDA's Handbook No. 8 (USDA, 1976-1989). Importantly to the Nutrition CRSP effort, the

Minilist contains the nutrients which are of analytic interest - those which were potentially inadequate, such as vitamin A, folate, iron, and zinc, as well as two important food components that might alter bioavailability - phytate and dietary fiber. The Minilist also contains values for the essential amino acids, which is important for evaluating protein quality in these plant-based diets.

Obviously, however, foods that are appropriate for a United States data base often are not available in less developed countries, while staples in these other countries may be missing from a U.S. nutrient data base. Thus we began the process of modifying the UCB Minilist into the International Minilist (IML), via deletions, additions, and changes.

*Deletions* from the UCB Minilist were made to keep the data base to a manageable size. All mixed dishes were deleted, except those with a single major component; for example, chicken chow mein was deleted, but batter-fried chicken and cookies were retained. Foods that were specific to the U.S., such as instant breakfast, also were deleted. Finally, preparations that were not likely to be used in developing countries (e.g., frozen foods) were deleted. We also deleted two nutrients with insufficient international data: biotin and iodine.

*Additions* were more problematic, and required lengthy study of the foods commonly consumed and their nutrient profiles. Several types of additions were needed. First, staples that were not frequently consumed in the US were added: millet, sorghum, lentils, plaintain, and cassava. Next, meats and meat fats were added since those consumed in a developing country usually have a different composition from those in the U.S. - often they are higher in fat, so beef tallow and chicken fat were added to the data base, for use in combination with lean meats (i.e., a mixture could be specified that would match the fat composition of a typical meat as consumed).

Several fruits and vegetables that were unlike typical U.S. foods, and which might be significant sources of nutrients were added. The variation in the composition of these foods is a significant consideration, and sometimes required multiple entries (for example, four types of peppers). Also, the level of nutrients from soil contamination in all harvested foods is difficult to estimate, and probably highly variable. Better nutrient data on varieties of fruits and vegetables clearly are needed. There simply are not enough analytic data to have entries that are specific to location, season, etc., so nutrients which vary with the type of soil, season of the year, maturity, and/or contaminants must be regarded as estimates at best.
Cheeses are often lower in fat than typical U.S. cheeses, so ricotta, mozzarella, and 1% fat cottage cheese were added. Since mixed dishes are often made with copious spices, we added several of the more common (cinnamon, curry, and cayenne).

We used several sources for our new nutrient data. Much of what we needed was available in one of the USDA revised Handbook 8's (USDA, 1976-1989). We generally used these as the preferred source of data, since the methods and number of samples could usually be determined. These data were deemed preferable to more local, but often less current data, although we did make comparisons. In addition, we had three primary references, one from each of the three countries (FAO and USDA, 1982; USDHEW and FAO, 1968; INCAP and NIH, 1961). Several other publications proved very useful as well - both food composition tables (such as Paul and Southgate, 1978) and journal articles (such as one on the nutrient composition of tortillas by Ronhotra, 1985).

When published nutrient values could not be found for a nutrient in a food, values were imputed. Primary sources for imputed values included the USDA nutrient data bases for both the household and individual surveys (USDA, 1986a and 1988). These data bases have no missing values, so values have been imputed by the staff at HNIS when necessary. For some nutrients, it was necessary to do our own imputations. Sometimes we were able to impute values based on calculations - for example, a phytate value of tortillas based on the amount of cornmeal in a tortilla. Finally, we imputed values from similar foods as necessary. For example, amino acid profiles could be based on a similar food, and adjusted for the amount of protein in the food being added. Sometimes we imputed zeros, such as for vitamin D in plant foods.

*Changes* in the nutrient values for several food items were necessary. Fortification and enrichment is seldom practiced in these countries, so grain and dairy products were changed to their unenriched equivalents. Bioavailability is a major issue when studying diets in these countries, and we made several changes to facilitate estimation. First, all foods were flagged as being either plant or animal source, giving us the option of totalling two sets of nutrients (e.g., plant protein and animal protein, plant iron and animal iron). These flags will be extended in future versions of the data base (e.g., animal divided into meat, dairy, egg, etc.). Another change addressed the use of yeast in breads, and the effect of the phytase enzyme on the phytate content of grains. After an extensive literature search, we estimated the phytate destruction during fermentation of yeast breads, and included a negative value for phytate in yeast. Thus, breads with yeast, as in Egypt, will have a lower phytate content than breads without (Faridi, 1983).

Another type of change dealt with retention factors, to adjust for loss of nutrients during cooking. Although foods and ingredients are usually measured raw (e.g., flour), generally they are consumed after cooking (e.g., in bread). When nutrient data were available for cooked forms of a food, we generally chose that form (e.g., cooked lentils). If analytical values for a cooked product were not available, we applied retention factors (using data from USDA for the appropriate cooking method (USDA, 1986b)) to all foods that were likely to be cooked - primarily the various grains. The current IML has 234 food items (by coincidence we have added and deleted the same number of foods from the original UCB Minilist). There are relatively large numbers of fruits and vegetables (about 60) and meat/poultry/fish (36).

The process of developing the data base proceeded simultaneously with the task of deciding which foods on the IML would be substituted for each food item on the three country-specific nutrient data bases. The three country food coding schemes allowed for about 1500 food items (while there is occasional overlap, the total of distinct foods is still well over 1000). Each of these reported food items has been examined, and the appropriate IML food determined. This has been an interactive process - when no appropriate substitution existed, a food was added. A cross-reference directory was created, listing the original food code (the one used for the dietary data), and the IML code to be substituted. Two aids have been employed to make the substitutions more accurate:

> Each entry on the cross-reference list can have a multiplicative factor specified. For example, if a food has a higher or lower moisture content, the factor can be used to adjust the portion size, and thus all the nutrients proportionally. This option is frequently used when substituting cooked grains or legumes for the uncooked equivalents. For example, sukuma leaf in Kenya is another name for kale, so kale can be directly substituted (the adjustment factor is 1.0 since there is no change in the moisture). However, when a code for raw potato is used, we want to substitute nutrient values for cooked potatoes. If baked potato is used, an adjustment of 0.82 is necessary, since moisture is lost during cooking. In the other direction, when rice is cooked, moisture is gained, so the weight of raw rice must be multiplied by 2.78 to correspond to the nutrient value of cooked rice (one pound raw is equivalent to 2.78 pounds cooked).
>
> Mixtures also are frequently employed to combine IML foods to match the reported foods. This option is obviously used when creating combination dishes. For example, for mandazi, a fried dough consumed in Kenya, the ingredients are 70% wheat flour, 23% water, and 7% oil. In addition, mixtures can be used to adjust the fat content of meats or cheeses. For example, for Kenya "fatty beef", beef tallow is added to cooked beef chuck to give a beef that is higher in fat.

There are times when "orphan" foods are reported - meaning no (or very incomplete) nutrient data are available in any references, and no similar food exists on the IML. In this case, a best substitution is chosen and documented. Examples would be foods such as insects, wild rodents, and birds. In Kenya, we were able to obtain partial analytic data for termites, and used pork sausage as a substitution (since these insects are very high in fat). In Mexico, dove meat was reported, and was assumed to be similar to a high fat turkey meat.

The end product of the development of the three cross-reference lists will be a single food coding manual. It is hoped this manual will be expanded over time to be applicable to many developing countries. One issue that must be resolved is that of the nomenclature for foods. Sometimes the common names are confusing - we finally determined that Jew's mallow in Egypt was the same as jute (and suspect that Jew's mallow is a colloquial form of jute mallow). The same food will have different names in different countries (taro in Kenya is the same as arrowroot in the U.S.), while the same name may be used in different countries to refer to different foods (U.S. yellow yams are very different than international white yams). Thus, food names may need to be qualified by the country of origin, by a description, and usually by a botanic name.

We believe the International Minilist will be valuable for a number of purposes. Obviously, we will use it to estimate nutrient intakes by the populations of each country, and to identify those nutrients for which a population may be at risk. These estimates may need to be supplemented by more detailed (possibly chemical) analyses, but they provide a direction for further examination. We also will use nutrient intakes of individuals in multivariate analyses of the various Nutrition CRSP functional outcome measures (such as growth and cognitive performance) to determine potential relationships. Although a miniaturized data base cannot offer the specificity of a larger one, we believe it is not possible to be more specific because the current data for foods consumed in developing countries are limited and out-of-date. Furthermore, the variability of nutrient values by season and growing conditions often makes attempts to obtain specificity futile and possibly misleading. We believe that diet-function relationships of importance still will be detected in our large sample with data collected over a full year.

Perhaps the most important long-term use of the IML is for research design. For example, it would be possible to phrase research questions <u>before</u> data collection begins. If estimates of nutrients which are likely to be deficient in the local diet can be obtained from sample diets during a pilot study, investigators can better design the survey methods and instruments. Furthermore, compilation of these types of data bases provides a direction for much-needed further work on obtaining better data on the composition of foods that are important in developing countries. It is not until survey data are matched with existing composition data that the gaps become obvious and priorities for further analytic work can be set.

Supported by the Office of Nutrition, U.S.A.I.D., Cooperative Agreement No. DAN-1309-A-00-9090-00 and the University of California Agricultural Experiment Station.

We thank Birgit Liebl for her valuable assistance in updating and assembling the nutrient data.

**References**

1. ALLEN, L.H., PELTO, G.H., AND CHAVEZ, A. (1987). <u>The Collaborative Research and Support Program on Food Intake and Human Function, Mexico Project Final Report</u>. University of Connecticut, Storrs, and Instituto Nacional de la Nutricion Salvador Zubiran, Mexico, D.F.

2. CALLOWAY, D.H., MURPHY, S.P., AND BEATON, G.H. (1988). <u>Food Intake and Human Function: A Cross-project Perspective of the Collaborative Research Support Program in Egypt, Kenya and Mexico</u>. University of California, Berkeley.

3. FARIDI, H.A., FINNEY, P.L., AND RUBENTHALER, G.L. (1983). Effect of soda leavening on phytic acid content and physical characteristics of Middle Eastern breads. J. Food Science 48, 1654-1658.

4. FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS, AND U.S. DEPARTMENT OF AGRICULTURE. (1982). Food Composition Tables for the Near East. FAO, Rome.

5. INSTITUTE OF NUTRITION OF CENTRAL AMERICAN & PANAMA, AND NATIONAL INSTITUTES OF HEALTH. (1961). Food composition table for use in Latin America. NIH, Bethesda, MD.

6. PAUL, A.A., AND SOUTHGATE, D.A.T. (1978). McCance and Widdowson's The Composition of Foods. Her Majesty's Stationery Office, London.

7. PENNINGTON, J.A.T. (1976). Dietary Nutrient Guide. Avi, Westport CT.

8. MURPHY, S.P., AND CALLOWAY, D.H. (1986). Nutrient intakes of women in NHANES II, emphasizing trace minerals, fiber, and phytate. J. Am. Diet. Assoc. 86, 1366-1372.

9. MURPHY, S.P., WEINBERG-ANDERSSON, S.W., NEUMANN, C., MULLIGAN, K., AND CALLOWAY, D.H. (1990). Development of research nutrient databases: an example using foods consumed in rural Kenya. Submitted to J. Food Comp. and Analysis.

10. RANHOTRA, G.S. (1985). Nutritional profile of corn tortillas. Cereal Foods World 30, 703-704.

11. RICCI, J. JEROME, N.W., MOUSSA, W.A., GALAL, O., HARRISON, G.G., AND KIRKSEY, A. (1989). Developing an international food composition database: an illustration from the Nutrition CRSP. In: Proceedings of the Fourteenth National Nutrient Databank Conference. CBORD Group, Ithaca NY.

12. U.S. DEPARTMENT OF AGRICULTURE, HUMAN NUTRITION INFORMATION SERVICE. (1976-1989). Composition of Foods...Raw, Processed, Prepared. Handbooks 8-1 through 8-17, 8-20, and 1989 supplement. U.S. Government Printing Office, Washington DC.

13. U.S. DEPARTMENT OF AGRICULTURE, HUMAN NUTRITION INFORMATION SERVICE. (1986a). USDA Nutrient Data Base for Individual Food Intake Surveys. Release 2.1.

14. U.S. DEPARTMENT OF AGRICULTURE, HUMAN NUTRITION INFORMATION SERVICE. (1986b). Nutrient retention factors tape file. From: Data sets used to create Release 2 of USDA Nutrient Data Base for Individual Food Intake Surveys.

15. U.S. DEPARTMENT OF AGRICULTURE, HUMAN NUTRITION INFORMATION SERVICE. (1988). USDA Nutrient Data Base for Household Food Consumption Survey 1987-88. Preliminary version.

16. U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE, AND FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. (1968). Food Composition Table for Use In Africa. USDHEW, Bethesda, MD.

TABLE 1: SAMPLE DIET OF AN EGYPT MOTHER-TODDLER PAIR (GRAMS/DAY)

| Food | Mother | Toddler |
|---|---|---|
| Tea (with sugar) | - | 61 |
| Fenugreek tea (with sugar) | 130 | - |
| Buffalo milk | 50 | 50 |
| Aiysh balady (whole wheat flat bread) | 370 | 210 |
| Cooked egg | 80 | 80 |
| Karate (cheese "puffs", commercial product) | - | 10 |
| Cheese | 60 | 30 |
| Tomato (raw) | 300 | 250 |
| Orange (raw) | 250 | 180 |
| Firaak (chicken) with taro, tomato, water, onion, fat, and garlic | 648 | 542 |

TABLE 2: SAMPLE DIET OF A MEXICO MOTHER-TODDLER PAIR (GRAMS/DAY)

| Food | Mother | Toddler |
|---|---|---|
| Coffee with sugar | 240 | - |
| Atole d'harina made of rice flour, water, and sugar | - | 480 |
| Tortillas | 808 | 294 |
| Huevo (egg) with tomato, onions, oil, salt | 60 | 60 |
| Calabacitas (squash) with onions and oil | 90 | 60 |
| Carne guisada made of beef, tomatoes, onions, and salt | 190 | 80 |

TABLE 3: SAMPLE DIET OF A KENYA MOTHER-TODDLER PAIR (GRAMS/DAY)

| Food | Mother | Toddler |
|---|---|---|
| Tea with milk and sugar | 700 | 425 |
| Whole milk | - | 20 |
| Fermented porridge made of maize meal, water, and sugar | 575 | 260 |
| Gitwero made of potato, banana, water, oil, and salt | - | 275 |
| Githeri made of maize, beans, water, fat, and onions | 1070 | - |