Capturing Ethnic Diversity in the Database

Suzanne P. Murphy, Ph.D., R.D., Department of Nutritional Sciences, University of California, Berkeley

Food choices in the United States are expanding. This is partly because of new products introduced by U.S. growers and manufacturers, but it is also because of increased consumer interest in foods from other countries, combined with increased trade with other countries. Capturing ethnic diversity in a nutrient database is certainly an appropriate topic to consider for the next century; we are going to increasingly be living an a global community, and food exchange is no exception.

One indicator of the interest in foreign foods is the number of restaurants serving ethnic cuisine in a community. For example, the yellow pages for the Berkeley/Oakland area in 1993 list 183 restaurants by self-defined ethnic classification: there are 45 Chinese restaurants, 39 Mexican, 22 Japanese, but only 5 consider themselves American. I'm sure most of you share my dismay when a subject in a nutrition study reports that she ate dinner in a Chinese restaurant. Not only is she unlikely to remember, or even know, the ingredients in the dishes, but even if she did, we may not have those ingredients on our nutrient database. Much of the time we must compromise with a limited number of Chinese "mixed dishes", which we assume/hope will be close to the actual item.

Another source of ethnic foods in U.S. diets occurs as a result of the expanding interest in ethnic cooking at home. A recipe for risotto in a recent popular magazine, for example, calls for rapini (broccoli rabe), Chinese broccoli (gai laan), and arborio rice (medium grain). I couldn't find any of these items in USDA's Handbook No. 8 (1), so would probably use regular broccoli and regular white rice if a subject reported these foods. However, I'm uncomfortable doing that without knowing what kinds of errors may be incurred.

We also need ethnic foods, of course, when surveys include persons from various ethnic origins, especially those who have come to the United States recently. In California, we have an ongoing challenge to correctly identify ethnic food items from our large populations of persons from other countries: Mexico, China, Japan, Vietnam, Cambodia, and so forth.

Thus, nutrient data from other countries may be needed for imported food items, for foods from ethnic restaurants, and for foods prepared by immigrant population. The question then becomes, where can these values be found, so that ethnic diversity can be captured on the data base?

Certainly the biggest initial problem is that of obtaining accurate food descriptions, both of the food reported, and of foods on the various nutrient databases that are potential matches for the food reported. We've been very pleased with the concept of using Langual (2) to describe foods in a uniform manner, although for our purposes, a subset of the full Langual descriptors is adequate. In spite of several inquiries, however, we have not found any examples of forms that are used to collect food descriptors for ethnic foods. I'm in the process of developing a form of this sort, and would appreciate talking with any of you who have also gone thorough this process. Some of the descriptors that I think need to be included are: local food name, English food name, scientific name, color, maturity, part consumed, processing applied, fortification or enrichment, and method of preparation. These descriptors are subsets of those used by Langual, or by the faceted INFOODS description system (3). It is my hope to develop a form that is simple enough to be used by field personnel.

The next step is to find some actual nutrient data for the food item of interest. First, I'd like to assume the item is a "basic" food (that is, not a mixed dish). There are several standard sources of nutrient data for ethnic foods. Printed or published tables remain the main source of ethnic nutrient data; however, electronic databases are being used more and more frequently. Journal articles and written communications are also sources of useful data, of course, but tend to be used less often than printed or electronic tables.

To be most accurate, it's desirable to find data from the specific country of origin. How much variation might one expect? That depends, of course, on the food and nutrient of interest. For example, if we were interested in the vitamin A (RE of carotene) content of peppers, we would have a wide choice of values to use, depending on the country or region of origin. Table 1 summarizes values from four countries/regions: United States, Great Britain, China, and Latin America (1, 4-6). Even within a country, there would be a range of appropriate values, depending on the type of pepper (red vs. green; sweet vs. hot, cooked vs. raw, etc.). These are often difficult choices to make, especially if the food item is not completely described. At shown in Table 1, carotene values are generally higher for red peppers than for green, and for hot versus sweet peppers, but there is considerable variation within the categories.

One issue that is often debated is whether to use a U.S. value for foods from other countries. For example, are canned chili peppers imported from Mexico the same as chili peppers canned in the U.S.? It is certainly possible to argue that the U.S. values are more current, and use more modern methods, since many of the published analyses from Mexico are now 30 years old. However, the variety of chilies, and thus their nutrient values, may be very different in Mexico. In this particular case, we would choose to use the U.S. values for the green peppers, but probably the Mexican values for the red peppers. If more data bases carried confidence codes for the nutrient values (as was done in the recent publication by Mangels et al. (7)), these decisions would be easier.

This brings up some concerns when merging data from multiple nutrient databases; in particular, merging U.S. data and data from other countries. First, there are concerns about analytic methods. For example, older methods of analyzing carotenoids overestimated vitamin A activity by including non-vitamin A precursors. Nutrient definition also can vary across nutrient databases. Differences in factors used to convert carotene to retinol illustrate this concern—the Latin American database (6) uses a conversion factor that is three times higher than the one used by the U.S. tables. Finally, sampling can drastically affect the values, particularly if comparing values for a national sample (as is often the case in the U.S.) and samples from a specific location (as may be the case in smaller databases).

Better documentation on the sources of nutrient data would help users of ethnic data. This is a challenging task if done properly. I would suggest that complete documentation of sources, either on paper or (preferably) in electronic form, includes: (a) primary source of the data for a food item, (b) supplemental sources for specific nutrients, (c) sources consulted but not used, and (d) identification of imputed or calculated values, plus a description of the method used.

Some of the published tables that I have consulted frequently when looking for ethnic or international data are: Africa, 1968 (8); Latin America, 1961 (6); Near East, 1982 (9); East Asia, 1972 (10), and India (11); but the usefulness of some of these data is limited by out-of-date methods. I'd like to make an urgent request at this point—we desperately need better data on foods from other countries. I know USDA's Human Nutrition Information Service (HNIS) and the Nutrient Composition Laboratory are very aware of this issue, and are working on gathering data for common foods in the U.S. supply. However, in addition, we also need better data from the countries of origin.

The preceding discussion has addressed primarily issues when trying to match basic food items. The problems with mixed dishes are even greater, of course, since one needs information on the proportions as well as the identity of each ingredient. In some cases it may be possible to find published information on the nutrient content. Let's take salsa as an example. This is a traditional Mexican dish that has become very popular in the U.S. (I understand sales for salsa now exceed those of catsup). Table 2 shows two recipes for typical homemade salsas; these recipes were developed by HNIS for the Nutrient Database for Individual Intake Surveys (12). If the ingredients match those that my population uses for salsa, the chances are good that I can not only use their recipe, I can also use the nutrient totals on the survey data base. Of course, I need to check the proportions as well as the preparation method. I also should check that the moisture loss is about the same (7%) if I use the cooked salsa, and that nutrient retention factors have been applied (which is the case for the HNIS recipe). Table 3 shows the importance of choosing an appropriate recipe—the two illustrated here have very different nutrient profiles. But what are the chances that either of these salsas will match those prepared by a Mexican-American homemaker in central California? Perhaps not very high. What most users will need is a system that allows either new recipes to be entered, or better yet, some way to modify existing ones; e.g., I could take out the fat; change red peppers to green; alter the proportion of tomatoes, etc. These are not trivial undertakings, but systems like the one developed by the University of Texas (13) can prove very useful when trying to adopt typical U.S. recipes to those reported by ethnic communities.

In an effort to simplify a very complex problem, Dr. Doris Calloway and I are consolidating food composition data from multiple developing countries into a single data base. Three years ago at this conference, I talked about an international nutrient database, the International Minilist (IML), that contained foods for the three study sites of the Nutrition Collaborative Research Support Program (NCRSP) (14). The IML is a <u>single</u> nutrient database which can be used in different locations. When we calculated nutrient totals for local diets of village populations in the three countries (Egypt, Kenya, and Mexico) using the consolidated nutrient database, we found they were very similar to those calculated using country-specific nutrient databases. Although the IML was developed for assessing diets in other countries, we have since found that the same approach has been very useful for ethnic populations in California—in particular, we have used this concept for diets of Hispanics and of Native Americans. After working on these projects for several years, we concluded that it should be possible to index foods in any country or ethnic culture to a relatively small number of basic food commodities. Thus, with funding from USAID's Office of Nutrition, we are in the process of broadening this concept to three additional countries: India, Indonesia, and Senegal. The full system, including a dietary assessment program, in now called the WorldFood System. Again, we will evaluate our methods by comparing nutrient totals for local diets using both the IML and the country-specific nutrient databases. The guiding principles for this ethnic/international nutrient database are that a minimum number of foods on the IML is desirable for accuracy and maintainability; a valid substitute should exist on the IML for every country-specific food (and if it doesn't, then one should be added); and no nutrient values should be missing.

The core of the WorldFood System is the IML nutrient database with approximately 200 foods that represent basic foods consumed world-wide. Foods reported in diets are indexed in three ways to the IML foods: (a) directly; (b) using adjustment factors for differences in moisture content (e.g., dry rice may be indexed to cooked rice); (c) using recipes, which allow for multiple ingredients, differing fat levels, and nutrient fortification. Care is also taken to index cooked foods so that nutrient losses during cooking are considered. Thus, although the IML has only 200 foods, the list of country-specific foods is approaching 2000. With this indexing approach, the system can be expanded almost indefinitely with minimal addition of nutrient data.

Thus, we can estimate intakes of a wide variety of nutrients (with no missing values), while maintaining a relatively small nutrient database. This approach allows us to capture ethnic diversity, at least for initial estimates of likely nutrient adequacy for a population. However, users would wish to consider the precision of these estimates before undertaking specific intervention programs in an ethnic population. Chemical analysis of frequently consumed foods and of typical diets is often advisable.

The accuracy of our substitution scheme is greatly dependent on the availability and accuracy of published nutrient data for ethnic/international foods. As previously mentioned, these data are often limited. Nutrient databases from the U.S. Department of Agriculture contain the more common ethnic foods consumed in the United States, but it clearly is an almost insurmountable task to keep up with the expanding number of imported food items. Yet we need these data not only for local and regional research projects, but also for National Nutrition Monitoring efforts. Thus, for a variety of uses, we need to promote expanded analyses of international food items, and better ways of exchanging the results.

References

1. Human Nutrition Information Service, USDA. Agriculture Handbook No. 8, Composition of Foods...Raw, Processed, Prepared. Springfield VA: National Technical Information Service. 1976-1992.
2. Smith EC. Langual for database users. 16th National Nutrient Databank Conference Proceedings, SP Murphy, editor. Ithaca NY: The CBORD Group, 1991.
3. Truswell AS, Bateson DJ, Madafiglio KC, Pennington JAT, Rand WM, Klensin JC. INFOODS guidelines for describing foods: a systematic approach to describing foods to facilitate international exchange of food composition data. J Food Comp Anal 1991;4:18-38.
4. Holland B, Welch AA, Unwin ID, Buss DH, Paul AA, Southgate DAT. McCance and Widdowson's: The Composition of Foods, Fifth Edition. Letchworth, United Kingdom: The Royal Society of Chemistry Distribution Center, 1991.
5. Ershow AG, Wong-Chen K. Chinese food composition tables. J Food Comp Anal 1990;3:191-434.
6. Leung WW, Flores M. Food Composition Table for Use in Latin America. Guatemala City: The Institute of Nutrition of Central America and Panama, and Bethesda, MD: Interdepartmental Committee on Nutrition for National Defense, National Institutes of Health, 1961.
7. Mangels AR, Holden JM, Beecher GR, Forman MR, Lanza E. Carotenoid content of fruits and vegetables: An evaluation of analytic data. J Amer Diet Assoc. 1993;93: 284-296.
8. Leung WW, Busson F, Jardin C. Food Composition Table for Use in Africa. Rome: Food and Agricultural Organization of the United Nations, and Bethesda, Maryland: U.S. Dept. of Health, Education, and Welfare, 1968.
9. Polacci W, McHargue JS, Perloff BP. Food Composition Tables for the Near East. Rome: Food and Agricultural Organization of the United Nations, and Hyattsville, MD: Consumer Nutrition Center, U. S. Department of Agriculture, 1982.
10. Leung WW, Butrum RR, Chang FH. Food Composition Table for Use in East Asia. Rome: Food and Agricultural Organization of the United Nations, and Bethesda MD: U.S. Dept. of Health, Education, and Welfare, 1972.
11. Gopalan C, Rama Sastri BV, Balasubramanian SC, Narasinga Rao BS, Deosthale YG, Pant KC. Nutritive Value of Indian Foods. Hyderabad, India: Siddamsetty Press, 1989.

12. Human Nutrition Information Service, U.S. Department of Agriculture. Nutrient Database for Individual Intake Surveys, Release 5. Hyattsville MD: U.S. Department of Agriculture, 1993.

13. University of Texas Health Science Center. Food Intake Analysis System, Version 2. Houston TX: University of Texas, 1992.

14. Murphy SP, Calloway DH. Development of a database for the Nutrition CRSP project. Proceedings of the Fifteenth National Nutrient Databank Conference, MR Stewart, editor. Ithaca NY: The CBORD Group, 1990.