DESIGNING A COMPUTERIZED NUTRIENT DATABASE FOR ACADEMIC/EDUCATIONAL APPLICATIONS

Suzanne P. Murphy, Ph.D., R.D.

I. General objectives/applications of the database.

At the University of California, Berkeley, we respond to a variety of users of computerized nutrient data:

- *Students* in our Nutritional Sciences and Public Health Nutrition classes analyze diets as part of various assignments. All nutrition majors are required to analyze three days of their own dietary data, collected according to various methodologies. In addition, there are occasionally specific needs, such as a recent exercise in a laboratory class to compare vitamin C intakes with urinary levels of ascorbic acid.
- *Dietetics* trainees also require access to nutrient data. Their curriculum emphasizes design as well as assessment, so they require a system that allows rapid feedback on changes to proposed diets.
- *Researchers* require accurate data for a variety of nutrients, primarily to assess intakes of survey participants. Occasionally nutrient data are used to calculate the content of research diets, although investigators are more likely to have these chemically analyzed. Traditionally, our researchers need nutrient information that is not commonly available, particularly for trace elements such as zinc and copper, or the various forms of iron.

After several years of working with researchers who use nutrient data (including my own research) and of teaching both nutrition and dietetics students, I have come to the conclusion that no nutrient data base that we currently have on our campus is ideal for everyone. Researchers want a data base with a large number of foods and nutrients, and high specificity of food items. They are less concerned with ease of coding or with the costs to run analyses. The dietetics students have almost opposite requirements - they want a fairly small nutrient data base so sample diets can easily be changed and costs of analysis are low. The nutrition students fall somewhere in between - perhaps one of the most important considerations for them is that the coding and data entry scheme not be too complex, as they are usually novice computer users (1).

II. What nutrient data are collected and how.

We now maintain three nutrient data bases, one for each of the three types of users. All are derived primarily from the various versions of Agriculture Handbook No. 8 (2) or Home and Garden Bulletin No. 72 (3).

1. For classroom dietary assessment assignments we use the diskette version of Home and Garden Bulletin No. 72 (1985 version) with 908 foods and 19 nutrients. This academic year we switched all our assessment laboratories to microcomputers, and have been pleased with the results. Many students have their own microcomputers now, and the vast majority of scientists will be using them in the future, so it is logical to use microcomputers in our classrooms. We have written a Fortran program for the IBM PC that interactively accepts the Home and Garden codes as well as serving sizes in multiples of the standard serving in the bulletin. Although this data base has fewer foods than the Agriculture Handbook No. 8 data base used previously, the foods are more representative of those in students' diets, and there is considerably less frustration about missing food items. Another advantage is the ability to purchase the "coding manual" for $2.75, and have a useful food composition table as well.

2. For dietary design applications we wanted a nutrient data base that would fit on a microcomputer spread sheet, so students could see immediately the effect of entering and deleting food items, and of changing serving sizes. After searching unsuccessfully for an existing database, we chose to develop our own with help from a campus group called the Instructional Technology Program. We entered 85 foods and 11 nutrients from the Home and Garden Bulletin No. 72 into Microsoft Multiplan for the Apple Macintosh. Students used this program last semester for the first time, with considerable success. Spread sheets can become unwieldy if they greatly exceed the size of the screen, so it is best to keep the number of foods small, and add items as needed for a specific diet assignment. If assignments could be solved using only foods already on the data base, the system worked very well. However, learning how to add food items for specific assignments was difficult for some students. I'd be very interested in suggestions any of you in the audience might have on how to simplify the use of spread sheets for this purpose.

3. Our most demanding needs come from our researchers. They want high specificity *and* a large number of nutrients. We have addressed the latter problem for many years with the UCB Minilist (4). This nutrient data base has relatively few food items (235) but a fairly extensive set of nutrients (49). The data come from many sources, including the USDA Standard Reference tapes and journal articles. There are no missing values, which means a relatively high maintenance cost, especially when new foods or nutrients are added. In order to achieve the high specificity required, we have developed an automated system of recipes and multipliers to allow substitutions of Minilist foods, or combinations of foods, for items reported in diets. Some work I did a few years ago showed that nutrients present on the Minilist and also on the NHANES II nutrient data base correlated well for individual diets when this system was used (5).

III. What data are available to the public.

All three of our data bases are available to the public. The Home and Garden Bulletin No. 72 data base can be ordered from the National Technical Information Service (order forms are usually available at this conference). Anyone who would like to have a copy of the spread sheet data base for classroom use may contact me. It is copyrighted by the U.C. Regents, which means it can't be sold by others, or further distributed without permission. The UCB Minilist is also copyrighted, and is available on a diskette for the IBM PC for a fee which helps cover the cost of maintaining this data base. Contact me for more information.

IV. Nutrition labeling and appropriate uses of label values.

We use very few data directly from nutrition labels. The major exception is for ready-to-eat breakfast cereals--since we have been particularly interested in trace mineral levels, it is important to distinguish levels of fortification in these cereals. The labels are used as a guide to the amount of fortification that is currently being used by a manufacturer. Some of the data from USDA's revised Handbook 8 (6) are already out-of-date (and were also based on label values in a number of cases), due to the rapid reformulations and introduction of new products. In this situation, I believe a trip to the supermarket to read the labels is the fastest way to be sure a nutrient data base reflects current fortification levels in a specific area. Although label values are likely to be lower than actual values, the level of accuracy is adequate for most purposes.

We also use label information to determine proportions of ingredients for mixed dishes. For example, if the nutrient content of macaroni and cheese is given (for several nutrients), the proportion of the ingredients can be estimated. We would probably carry this item as a recipe on our data base, rather than add the nutrient values, even if they were all known. The usual case is that only a few nutrients are given on the label, and the rest must be estimated from the ingredient proportions. Thus, we find it easier to just carry the recipe.

V. Impact of missing nutrient values.

We do not have any missing values on any of our data bases. The ones used for nutrition and dietetics students are taken directly from USDA data bases, which come without missing values (the staff at USDA calculate and impute values if necessary). For the UCB Minilist, a nutritionist fills in the missing values for the 49 nutrients that we carry. This is done via literature searches, calculations, and substitutions, as appropriate. Our policy has been that a reasonable estimate is better than always assuming (by default) that a missing nutrient value is zero.

VI. Summary.

The task of developing nutrient databases for academic and educational applications is a constantly challenging one. I appreciate the opportunity offered by this conference to share with you my experiences, and to learn from yours.

REFERENCES

1. Murphy, SP, King, JC, Calloway, DH. Choosing a diet analysis system for classroom use. J Nutr Educ 1984;16:73-5.
2. United States Department of Agriculture. Composition of Foods. Agriculture Handbook No. 8 and Revisions. Washington, DC: U.S. Government Printing Office. 1963-1986.
3. United States Department of Agriculture, Gebhardt SE, Matthews RH. Nutritive Value of Foods. Home and Garden Bulletin No. 72. Washington, DC: U.S. Government Printing Office. 1985.
4. Pennington, JA. Dietary Nutrient Guide. Westport, CT: Avi Publishing Co. 1976.
5. Murphy, SP, Calloway, DH. Nutrient intakes of women in NHANES II, emphasizing trace minerals, fiber, and phytate. J Am Diet Assoc 1986;86:1366-72.
6. United States Department of Agriculture. Composition of Foods. Breakfast cereals. Agriculture Handbook No. 8-8. Washington, DC: U.S. Government Printing Office. 1982.

Partial funding for development of the IBM PC-XT diet analysis systems was obtained through an IBM DACE grant.