Getting Started: An Overview of Nutrient Databases

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

The purpose of this overview is to acquaint new users with the basic concepts and terminology necessary to become knowledgeable about nutrient databases. I will cover some information about the definition, format, terminology, and features of nutrient databases. I realize this will be at least a partial review for many of you, but I hope it will be useful to establish a baseline of knowledge for everyone.

There is increasing awareness of the link between diet and disease, and therefore increasing interest in dietary nutrient data by consumers and nutrition professionals alike. Many consumers are reading the nutrition information on labels, and the upcoming revisions to the labeling standards have certainly increased interest even more. As a result, nutritionists are being called upon as a resource to guide colleagues in the health professions, as well as the public, in determining the nutrient content of the food we eat. We all know this is a complex task. Thus, this conference is held annually and provides a forum for the exchange of ideas and methods for better performing tasks concerned with compiling and using nutrient data.

To begin, I'd like to comment on the distinction between a food composition table and a nutrient database. Generally, we reserve the former term for printed tables showing the nutrient composition of foods (e.g., books such as USDA Handbook No. 8 (1)), and the latter for the same type of data, but in electronic form (e.g., disks and tapes). The amount of data available currently make it impractical for most people to keep paper copies of all the information. Furthermore, it is a tedious and time-consuming task to try to determine the nutrient content of multiple foods (such as in a diet) by looking up all the foods in a printed table. As a result, this conference focuses primarily on electronic nutrient databases although in many cases the data also exist in printed form as well.

Why are nutrient databases so crucial for dietary assessment and dietary design? Primarily because, given the complexity of our food supply, and the expense of chemical analyses, few nutritionists can afford to actually analyze food samples. As a result, we must rely on published and electronic food composition data.

If one is to keep data in electronic form, one must have both computer hardware and software to access it. As for the hardware—the computer itself—you all are very aware that the choice of computers currently available is almost overwhelming, and that it changes daily. Ten years ago, it was necessary to have access to a large (mainframe) computer system to do much with nutrient databases; and since it was very expensive to keep data on the mainframe tapes or disks, you saw people walking around with decks (or boxes) of computer cards in their hands. Today, it is much more common to work with personal computers. Not only has our dependence on cards disappeared, but the use of display screens saves reams of paper. The switch to personal computers (PC's) has had an enormous impact on dietary assessment technology. Most of us have PC's on our desks, so, with the proper programs, we can analyze a subject's diet in a few minutes. Furthermore, many of our subjects can analyze their own diets in an equally short time. Through a modem, we can be in touch with other systems and users, using phone lines. There are many other hardware features we could discuss, but time doesn't permit me give a hardware overview. As you will see from the exhibits at this conference, almost all of the nutrient databases available today use PC's, but there is a wide variety of specific types in use.

What does a typical nutrient database contain? Generally the same information as is in a typical printed food composition table: usually there is a record (line of data) for each food item, and a field (data value) for each nutrient. In addition, there is some kind of identifier for the food item: usually an identifying code, and often a full alphabetic name as well. Thus, at a minimum, the record for each food item will contain the food identifier and the amount of each nutrient in a specified amount of the food. Often the default amount is 100 grams (i.e., the nutrient values are the amounts contained in 100 grams of the food item). This is useful for people compiling nutrient data, but not very useful for people using it; we want to be able to determine how much is in a cup or a tablespoon of the food. Thus, virtually all databases have some information of the weight of a usual "portion". Sometimes, the nutrient data will be for one of these usual portions, rather than for 100 grams (i.e., the vitamin C in a cup of orange juice), sometimes the data will still be for 100 grams, but the weight of a cup will be contained in a separate field on the database; the nutrient calculation program can then multiply all the nutrient values by the appropriate weight. As you can imagine, there are many ways of organizing portion size data on the database. For some databases, the same food is carried multiple times, once for each portion size; there might be an entry for one medium peach, one cup of peach slices, and 3 ounces of peaches. This is somewhat a holdover from printed tables, since computer programs can readily recalculate the nutrient content of differing portions. Duplicate entries for the same food item will inflate the apparent size of a nutrient database; there may be 1000 food records, but only 500 different food items.

Databases often contain other information about the foods and nutrients. Some examples would be: food group codes; the number of dietetic exchanges; flags for allergens like lactose and gluten; codes for the sources of the nutrient data; refuse factors for converting "as purchased" weights to "edible portion" weights.

This leads us to consider the size of a nutrient database. Generally the size is a direct function of the number of food items (rows) and the number of nutrients (columns). With the price of computers, and particularly of disk storage, falling, it is often the case that bigger is better. It's very frustrating not to have the exact food that was in the diet, and have to figure out something close to it as a substitute; furthermore, it can be quite inaccurate. Precision (how closely you can match the food consumed with a food on the database) can be increased by having larger databases. Even more important is having all the nutrients that interest you and your subjects. The key here is to decide if you need to pay for the additional accuracy of a large database, <u>for your application</u>. For some types of research, accuracy is very important; for classroom use, it may be less critical. If your hardware can support a large nutrient database (i.e., it has enough storage space and doesn't take a long time to find each food item), then you may want to invest in a large database, particularly for research projects. However, large databases do have some disadvantages: they usually cost more to purchase and keep current, and the potential for error is much greater. Thus, it is particularly important to select one that is carefully compiled and maintained.

When selecting nutrient databases, it's obviously crucial to find out the sources of the nutrient data. The "core" of most U.S. databases is from data compiled by the U.S. Department of Agriculture. The next speaker will tell us more about these data. Many database compilers also use other sources; commonly, information from the food industry is also incorporated (often called "brand-name" data). Many food manufacturers supply these data directly to database compilers; in other cases, the data are taken from the nutrition labels on the product. Be aware that these may not be the same numbers! The data on the labels are conservative; that is, they somewhat underestimate the "good" nutrients and overestimate the "bad" ones for a variety of regulatory reasons. Thus, a compiler who is obtaining the actual data from the manufacturer will have a more accurate database. There are other sources of nutrient data as well: sometimes data are gathered from published literature such as journal articles; sometimes data will come from other nutrient databases (such as international databases); and sometimes a compiler will have access to unpublished data from analytic laboratories. Whatever the sources of the data, developing and maintaining a database is not a trivial task. Our food supply is constantly changing, so information that is accurate today may be completely incorrect tomorrow (as the manufacturer decides to change the fortification profile of a breakfast cereal, for example). If accuracy is important for your application, then the method of keeping the database current should be carefully examined.

The handling of missing nutrient data is another issue of importance. As the size grows, it is often the case that the number of data points without nutrient values grows; this is especially a problem with brand-name data when the manufacturer only gives information for a few nutrients. If the missing nutrient values are left blank for a food item, then programs that add up the nutrients in a diet will assume there is none of that nutrient in the food (even though the food may actually be a good source of that nutrient). Thus it is key to know what a compiler is doing with missing data; any advantage of having a large database with many food items may be offset by the disadvantage of incorrectly assuming missing nutrient values are zero. It is my opinion that it is always better for compilers to impute (estimate) a value from a similar food, if it is done by an expert and well documented. However, correctly imputing values is a time-consuming task, especially for large databases.

It's important to remember that just because numbers come from a computer, they aren't always right! In addition to errors due to missing values, nutrient values could be incorrect for a variety of reasons, including out-of-date values and data-entry errors.

Most people who work with nutrient databases make a distinction between the database itself and the programs that access it; obviously, it's not very useful to have a nutrient database but no convenient way to summarize dietary totals and compare the results to a standard (often the Recommended Dietary Allowances). The combinations of programs and databases that allow us to do all these things are referred to as nutrient calculation systems (or dietary assessment systems). The range of features available for these systems is astounding, and the time and expertise it takes to develop a flexible system is huge. The method of displaying the results often takes advantage of the excellent graphic capabilities of computers and printers. Evaluating these systems is by no means a trivial task, and guidelines will be offered by several of the other speakers today.

One important feature of a nutrient calculation system is the method by which food items are accessed: how the identifying information entered by the user from the keyboard of the computer is matched to the proper item. The access (or coding) scheme may be dictated by the organization of the nutrient database. Often each food item has a numeric code. In the simplest data entry scheme, you enter this numeric code for each food item. Thus, if you want to specify "white bread", you will have to look up the code for white bread (in some kind of coding manual) and enter the its number (e.g., "346"). This isn't very time-consuming for a small database, but the coding manuals are extensive for a large one. Therefore, many systems now take other approaches. Some access programs will search on the name of the food ("white bread"); but problems can sometimes arise with the format and spelling (e.g., if you are supposed to enter "bread, white"). More elaborate schemes let you search for the first few characters, or for combinations of characters that are anywhere in the food name. The other, more common, approach is to use hierarchical menus. In this case, the access program shows you several options on the screen (typically some broad categories of foods) and you select the proper one (e.g., grain products); the next menu might let you select breads vs. cereals, pasta, rice, etc.; the final menu lets you select the exact type of bread. Sometimes a combination of methods is available—you can enter the code (or part of the code) if you know it, and then use the menus to get to the specific item.

Cost of nutrient databases and nutrient calculation systems vary widely. A few systems are public domain—you may have to pay an initial amount to cover the cost of copying and mailing, but you can then make as many additional copies as you wish; or if you know someone who already has the system, you can make yourself a copy without any payment. For example, there is a very nice system provided by USDA (Diet Analysis Program) that is public domain (2). You can order it for about $60, or you can download it from USDA's bulletin board at no cost. One can expect to pay more for systems with extensive manufacturer data, uncommon nutrients, imputed data, frequent maintenance, a wide variety of nutrient calculation options, and sophisticated data entry and display. Also, when you're thinking about costs, remember to find out the fee for updates. In some cases, your institution may have a site license for a system, so you only pay a small fee (or no additional fee). On our campus, we have developed our own system for classroom use, which has the great advantage that we can supply it to our colleagues and our students without any fee.

I'll conclude with some comments on documentation for nutrient databases. I feel this is a topic that often is ignored when evaluating and selecting systems. Of course, you want a system with a good users manual that explains exactly how to use the system. Sometimes these kinds of instructions are available on the computer by selecting a "help" screen. However, the documentation shouldn't stop with instructions on how to use the system. Be certain you know how to cite the nutrient database: at a minimum you will need to give the database name, version number, date, and the name and address of the vendor. There also should be documentation (either electronic or printed) that gives you details about the database itself: how many foods it contains, the exact definitions of all the nutrients, what were the sources of the data, what was done about missing values, etc. This information isn't always readily available, but I think users should insist that it's necessary.

Attached is an annotated bibliography which I use in the classes that I teach at the University of California. I've listed several comprehensive references that I find most useful, and also a number of journal publications that address methods for evaluating nutrient databases and dietary assessment systems.

I hope these comments have given you some background information to use throughout the rest of this session, as well as throughout the upcoming conference.

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. Human Nutrition Information Service, USDA. Dietary Analysis Program. Springfield VA: National Technical Information Service. 1989.

Nutrient Data Bases and Computerized Diet Analysis Systems: An Annotated Bibliography

Suzanne Murphy

January, 1993

Comprehensive references:

Nutrient Data Bank Directory, 8th Edition, 1992. Smith, J.L. (Ed.). Order from: Department of Nutrition and Dietetics, Alison Hall, University of Delaware, Newark, DE 19715-3360. (Lists in detail the features of 51 nutrient data bases.)

International Directory of Food Composition Tables. Rand, W.M. (Ed.) 2nd Edition, June, 1988. Order from: INFOODS Secretariat, UNU, P.O. Box 500, Charles Street Station, Boston, MA 02114-0500. (List of 200 food composition tables from Africa, Asia, Europe, Middle East, Oceania, North and South America.)

The 1988 Annual Journal of Dietetic Software. P.O. Box 2565, Norman, OK. (Reviews programs of interest to dietitians; descriptions of software from approximately 75 organizations.)

Computer Programs and Databases in the Field of Nutrition. 1989. 6th Edition. Seattle, WA: Computing Information Center, University of Washington. (Abstracts of over 200 diet analysis systems.)

Food Composition Data: A User's Perspective. Rand, W.M., Windham, C.T., Wyse, B.W. & Young, V.R. (Eds.). 1987. The United Nations University. Order from: UNIPUB, 4611-F Assembly Drive, Lanham, MD 20706-4391. (21 papers on experiences, uses, and management of food composition data.)

Proceedings of the National Nutrient Data Bank Conference. This annual conference brings together users (both current and potential) and developers of nutrient data banks. The proceedings from many of the 18 conferences have been published.

Compiling Data for Food Composition Data Bases. Rand WM, Pennington JAT, Murphy SP, Klensin JC. 1991. Tokyo: United Nations University Press. A comprehensive discussion of the process of compiling nutrient composition data.

Microcomputer Software Collection. Food and Nutrition Information Center, National Agricultural Library, USDA. 1991. Lists cost and brief description of 58 diet analysis/diet planning programs, plus many other programs for food service management, nutrition education, and recipe analysis. All are available for demonstration at the NAL in Beltsville MD. To order, call 301-344-3719.

Other helpful references:

Buzzard IM, Price KS, Warren RA. 1991. Considerations for selecting nutrient-calculation software: evaluation of the nutrient database. American Journal of Clinical Nutrition 54:7-9.

Dare D, Al-Bander SY. 1987. A computerized diet analysis system for the research nutritionist. Journal of the American Dietetic Association 87:629-632.

Eck LH, Klesges RC, Hanson CL, Baranowski T, Henske J. 1988. A comparison of commonly used nutrient database programs. Journal of the American Dietetic Association 88:602-4. (Compares DINE, U. Texas, NCC, Short Report.)

Frank GC, Pelican S. 1986. Guidelines for selecting a dietary analysis system. Journal of the American Dietetic Association 86:72-75.

LaComb RP, Taylor ML, Noble JM. 1992. Comparative evaluation of four microcomputer nutrient analysis software packages using 24-hour dietary recalls of homeless children. Journal of the American Dietetic Association 92:1391-92. (Compares USDA's Diet Analysis Program, Food Processor II, Nutritionist III, and Univ. of Texas Food Intake Analysis System with USDA's Nutrient Database for Standard Reference).

Mattes RD, Gabriel SJ. 1988. A comparison of results from two microcomputer nutrient analysis software packages and a mainframe system. Journal of Nutrition Education 20:70-75. (Compares U. Mass., Nutritionist III and Nutriquest.)

Nieman DC, Nieman CM. 1987. A comparative study of two microcomputer nutrient data bases with the USDA Nutrient Data Base for Standard Reference. Journal of the American Dietetic Association 87:930-32. (Compares USDA, Nutritionist III, and The Food Processor.)

Nieman DC, Butterworth DE, Nieman CM, Lee KE, Lee RD. Comparison of six microcomputer dietary analysis systems with the USDA Nutrient Data Base for Standard Reference. Journal of the

American Dietetic Association 1992;92:48-56. (Compares DINE, Food Processor II, Minnesota NDS, Nutri-Calc HD, Nutritionist III and Professional Dietitian with USDA.)

Penfield MP, Costello CA. 1988. Microcomputer programs for diet analysis: A comparative evaluation. Journal of the American Dietetic Association 88:209-11. (Compares U. Tenn., Apple Diet, Eat Smart, Nutrichec, and Idaho.)

Ralston CE, Matthews ME. 1988. Software selection: Can a demonstration computer package help? Journal of the American Dietetic Association 88:1087-89.

Schakel SF, Sievert YA, Buzzard IM. 1988. Sources of data for developing and maintaining a nutrient database. Journal of the American Dietetic Association 88:1268-71. An excellent list of references for obtaining nutrient data.

Taylor ML, Kozlowski BW, Baer MT. 1985. Energy and nutrient values from different computerized data bases. Journal of the American Dietetic Association 85:1136-1138.

Windham CT, Helm AA, Wyse BW. 1990. Integrity of small data bases in computer analysis of dietary data. Critical Reviews in Food Science and Nutrition 29:149-66.