Food Composition and Nutrient Data Sources

Grace J. Petot

Virginia McMasters, in her comprehensive review of the history of food composition tables of the world, published in 1963, stated that "human knowledge is often accepted and used with little thought of its origins". I would like to paraphrase that statement to say that computerized food and dietary nutrient analyses are often accepted with little thought of the origins of the information provided. This discussion will first review and describe the process of accumulation and identification of the sources of food composition data in the HVH-CWRU Nutrient Data Base and secondly, identify some problems and issues which must be considered by those who are maintaining nutrient data bases, and by the various users of nutrient data bases.

All nutrient data bases in the United States are based on one or more versions of food composition tables compiled by USDA. The procedures, status and food composition data sets available from USDA were reviewed at last year's conference. In the meantime, Handbook 8-3, Composition of Baby Foods, has been published. Dr. Rizek brought us up to date in this morning's discussion. The USDA Food Composition Tables continue to be the best sources of information for conventional, relatively unprocessed foods, especially for meats, most dairy products, fruits, vegetables and many cereal grain products. Much information is still needed for these groups of foods. The situation has not changed appreciably since these needs were reviewed by Watt and Murphy in their article published in Food Technology in 1970. Therefore, in order to meet the ever increasing demands for food composition and dietary data by researchers, clinical practitioners, food service management and government rulemaking agencies, other sources of data are used. These sources may be one or more of the following:

1. USDA data reported in widely read journals

2. Composition data for foods or groups of foods for one or more nutrients published in a wide variety of journals

3. Directly from the food manufacturers and/or industry associations

4. Tables published in a number of books, for example, Bowes and Church, and the Geigy Tables

5. Other independently compiled nutrient data bases

6. From nutrition lables on food products

7. One's own laboratory analyses of one or more samples for one or more nutrients

8. Calculated values, especially for recipes, based on one or more sources

9. Imputed values based on varying criteria

With only a few minutes of thought devoted to this list, it does not take long for one to realize that there are numerous questions to be asked about the sources and that there are many problems associated with evaluating, interpreting and comparing output from the various nutrient data bases. These questions and problems are not new. There is just a more widespread awareness of them because the computer has made so much more information available, so much quicker and at much less cost to more diversified groups of people. Therefore, it is axiomatic that the user of any data have access to knowledge of the data sources so that he or she may be able to evaluate that data relative to a specific use.

The original use of the HVH-CWRU Nutrient Data Base was in conjunction with a diet diary method of data collection from free-living subjects as part of several research studies in the Cleveland area. Continuous development has resulted in the inclusion of more than 2300 food, recipe and therapeutic product items as a result of processing several thousand 24 hour intake records of free-living and hospitalized persons from many areas of the United States. Efforts have been maintained to accumulate as much reliable information as is possible about frequently consumed foods.

The major sources of nutrient data are USDA Handbooks 8, 456, 8-1, and 8-2. Handbook 8-3 data will be incorporated as soon as the tape becomes available, as will subsequent sections as they are published. Another major source, but one which is becoming outdated, is Bowes and Church, 12th Edition, which in the past served as one of the few tables which included food manufacturers' data. However, with better availability of this information directly from food manufacturers, it is becoming a less significant source. These sources, USDA, Bowes and Church and food manufacturers, because they provide in all cases, values for the major nutrients, are identified in the data base with a 7 character alpha-numeric major source code. For example, USDA item number 01-077 in Handbook 8-1, Milk, whole, 3.3% fat, fluid, has a source identification code in our food table of UD01077. The name of a food manufacturer is also indexed with an alpha-

numeric code and the source identification for a single food item includes the code number for the manufacturer. Original source materials are then filed for easy reference.

Space is also allowed within the data base for recording a source identification for each single nutrient. An alpha-numeric code is assigned to the source of data which may be one of the major sources, a journal article, government regulation, etc. These indexed materials are then filed for easy reference. A bibliography of references is included in the code manual. Insofar as is possible, primary sources of information are used. A few frequently used recipes, precalculated, are included. However, it has been our experience that when one ingredient in a recipe may be substituted for another by the cook, such as butter for margarine for example, it is preferable to obtain recipes or ingredient information and use these individual recipe ingredient items within a dietary or recipe analysis as needed. It is also possible to create an associated recipe file which accesses the ingredients in the Data Base, thereby always using the most current nutrient data.

Needless to day, with a rapidly changing food supply and with the many variations in life-styles and dietary patterns among individuals and groups, it is necessary to continually revise the list of food items included in the Data Base. Indeed, it is possible to respond almost immediately to the frantic dietitian who must have nutrient values included for a new product or an unusual food, provided we can obtain the composition data. Values for individual nutrients are added as they become available and as they are judged to be reliable. Thus, the nutrient data base is dynamic in nature.

Food items for which USDA values are used are described in the code manual as they are listed by USDA. Brand names are used for commercial products. All nutrient data are stored in the Data Base for 100 grams of product. Volumetric household and metric measure equivalent weight factors are included for conversion in the computer.

There are several problems associated with using food compostion data from many different sources. I would like to elaborate on just a few.

1. The reporting of data by food manufacturers.

Most of us who are maintaining data bases use values per 100 grams of food. We are finding that many manufacturers, especially those who have had analyses made for nutrition labeling, are cooperative in providing data in this manner. However, in many cases, we do not receive volumetric equivalent weights or portion size with weight which are more frequently used in recording intakes and for recipes. Analyses for many fluid products are customarily reported for a volume measure such as milligrams of nutrient/liter or quart. Specific gravity or another

volume/weight equivalent is necessary to convert to gram weights. U.S. RDA values are not satisfactory and are not consistent with the accuracy of other values being used. The percentages of U.S. RDA's are rounded for use on nutrition labels.

Careful evaluations must be made for calculated food composition values which are distributed by some manufacturers. Depending on the food product and its subsequent preparation for serving, these values may or may not be appropriately used in a data base.

It has been estimated that there are some 10,000 food products in the markets. Keeping up to date with just those which have been analyzed for nutrition labeling is a formidable task. For many products, it may be far into the future before analyses for all nutrients and other components are made. If interest and need is demonstrated, this task may need to be undertaken by a government agency.

We must admit that, at times, we are pesty and nagging in our requests for data, although many manufacturers are cooperative. Some have even been known to supply it quickly when a large institutional order is being considered on the basis of availability of nutrient information. We here today might make a plea for a centralized collection system and brand name identification in USDA tables.

2. Composition values from different sources for a single food item.

USDA handbooks do not provide lactose data. We know that lactose is present in milk and many milk products. In our search for lactose content information, we discovered that for many milk products, the lactose values reported in the literature did not report total carbohydrate for the same samples. These lactose values in many cases were greater than or much less than total carbohydrate values for the same products in USDA Handbook 8-1. Perhaps access to USDA Data Base II could help resolve this problem or we could suggest to USDA that this issue should be placed high on their priority list for publication.

A similar situation exists with fatty acids in the tables published in series in the Journal of the American Dietetic Association. For many food items, total lipid values do not match those in Handbooks 8 and 456. Therefore, when we include values for individual fatty acids, their sum may be greater or less than Handbook 8 total lipid values. Individual fatty acid values can be scaled to total lipid, but is this an acceptable procedure?

The HVH-CWRU Nutrient Data Base does include several table checking programs which allow us to sum components so that they may be compared with a reported total. For example, total water, carbohydrate, protein, fat and ash may be added. If they do not add up to 100 grams, a message is received and errors may be

traced. The same check may be performed for other breakdowns of nutrients.

3. Values for nutrients which have low priority for analysis in some foods and groups of foods.

As dietitians we make generalized assumptions that some nutrients and food components are either not present or are present in only trace quantities in some foods and groups of foods. For example, we generally assume that cholesterol is present in animal foods only. The identification of sterols in plant materials is currently being reported. Cholesterol may be present in only trace amounts or not at all. Instead of designating cholesterol content as being unknown because an analysis has not been or may not be made, can we agree that a zero or trace value may be assigned to these foods. Similar cases may be illustrated for several other nutrients such as sucrose in meats, lactose and B12 in plant foods. Chemical analyses are expensive and may indeed be unnecessary for some groups of foods.

4. Separate identification of nutrients naturally present in a food and the same nutrients added to the food.

In some studies it is desirable to separately identify the quantity of a substance which has been added. For example, with increasing use of fructose and lactose as food ingredients in foods which already contain the same substance. The first occurance of the substance should be identified as one which occurs naturally; the second could be identified as a refined sugar. At the present time, we do not have adequate quantitative information to identify these or enrichment and fortification nutrients separately.

5. Calculations of nutrients in recipes with consideration of nutrient losses in preparation.

Increasing use is being made of nutrient data bases for calculating nutrients in recipes for home or institutional use. Allowances for nutrient loss in preparation may or may not be considered. Resolution of this issue depends on the state of our knowledge of nutrient losses and agreement on the use of common factors for specific nutrients and preparation methods.

This discussion has briefly reviewed the origins of food composition and nutrient information in nutrient data bases and has discussed only a few problems associated with using data which is acquired from a variety of sources. Evaluation, interpretation and acceptance of the data and information processed by nutrient data bases requires close, ongoing communication among all of us here today and among all who contribute to, maintain, use and evaluate nutrient data bases.

Questions and Answers:

Q: Do you make an attempt to fill in blank spaces with nutrient values? As you said, we know, for instance, that "trace" is used to indicate a detectable quantity, but do you use zeros if you can assume that none of the nutrient is present?

A: If we can determine, and sometimes we can from an ingredient list, that the value is zero, then, yes, we do put a zero in. However, if there is any doubt we leave it as "unknown. Who is interested in chemically analyzing meats for sucrose, for instance? It is very low on a priority list. Yet we retain sucrose as "unknown" on our list of nutrients in the calculations for meats.

Q: How do you handle manufacturer's data which are based on nutrient label claims instead of analysis?

A: We do not use nutrient label information which is expressed as per cent of U.S.RDA. If the manufacturer will supply us with an analysis per 100 grams of product, we do use it.

Q: You mentioned a number that you have which corresponds to the item number in Handbook 8. Does this number link together the information about certain parts of the identification, the process?

A: Yes. This number is used as a source code and consists of alpha-numerics. For example, UDA0258 refers to a Handbook 8 number for a food item.