JOURNAL OF FOOD COMPOSITION AND ANALYSIS **10,** 102–114 (1997)
ARTICLE NO. FC970527

Procedures for Estimating Nutrient Values for Food Composition Databases

Sally F. Schakel,*[,1] I. Marilyn Buzzard,† and Susan E. Gebhardt‡

**Nutrition Coordinating Center, Division of Epidemiology, University of Minnesota, Minneapolis, Minnesota; †Department of Preventive Medicine and Community Health, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia; and ‡Nutrient Data Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Riverdale, Maryland*

Received July 10, 1996

When used to assess the nutrient content of diets, recipes, or commercial food products, a nutrient database should provide a complete nutrient profile for each food in the database. Chemical analyses for a wide range of nutrients in the many foods included in a database are not always practical. Therefore, some nutrient values must be estimated. Common methods for estimating nutrient values include (1) using values from a different but similar food, (2) calculating values from different forms of the same food, (3) calculating values from other components in the same food, (4) calculating values from household recipes or commercial product formulations for multicomponent foods, (5) converting values from information on the nutrient label of a commercial food product, (6) calculating values from a product standard, and (7) assuming a zero value. Quality control procedures and nutrient validation programs should be implemented to verify that appropriate data selection, calculation methods, and data entry were used. Estimated nutrient values should be identified in the database. Some referencing systems also indicate the method used to derive each estimated value.

INTRODUCTION

Nutrient databases provide food composition data that are used in a variety of ways (Rand *et al.*, 1991). Health researchers and epidemiologists use nutrient intake studies of individuals or groups to correlate food components with causes or prevention of disease. Dietitians counsel patients in dietary changes based on an analysis of their usual dietary habits. Food manufacturers determine the nutrient content of their products for food labels. Cookbook authors calculate nutrients in a serving of each recipe. Food service managers plan menus for schools, hospitals, and other institutions based on their nutrient content. All of these uses of nutrient databases require that the nutrient profile for each food in the database be complete (i.e., have no missing values) so that the nutrient content in a diet, a recipe, or on a food label is not underestimated.

PROCEDURES FOR ESTIMATING NUTRIENT VALUES

A nutrient database may contain some values obtained from chemical analysis of the food as well as other nonanalytic values which are calculated from conversion factors or estimated from other knowledge about the food. Chemical analyses of nutrients in foods are costly and often unavailable for foods not commonly consumed

[1] To whom correspondence should be addressed at 1300 S. Second St., Suite 300, Minneapolis, MN 55454. Fax: 612-626-9444.

0889-1575/97 $25.00

or for nutrients of lesser interest. To provide a complete nutrient database, it is often necessary to calculate nutrient values from available analytic data using defined algorithms or estimate values using other sources of data. Nonanalytic values may be described as imputed, calculated, or estimated (Rand *et al.*, 1991). Estimated values frequently are calculated values, but they may also require additional assumptions, such as the proportion of ingredients in a frozen entree and the nutrient contribution of each ingredient (Gebhardt, 1992; Schakel *et al.*, 1989; Westrich *et al.*, 1994). For the purposes of this paper, we will use the term ''estimated'' to represent all nonanalytic values.

Common methods for estimating nutrient values include (1) using values from a different but similar food, (2) calculating values from different forms of the same food, (3) calculating values from other components in the same food, (4) calculating values from household recipes or commercial product formulations for multicomponent foods, (5) converting values from information on the nutrient label of a commercial food product, (6) calculating values from a product standard, and (7) assuming a zero value (Sievert *et al.*, 1989; Gebhardt, 1995).

General guidelines and examples for each of the common methods for estimating nutrient values are described below.

1. Using Nutrient Values from a Different, but Similar, Food

Nutrient values from another food within the same genus or within the same family of foods may be used. Gebhardt (1992) provides the following examples of nutrient values imputed in this manner in the USDA Nutrient Database for Standard Reference:

> Vitamin B-6 and folate in black walnuts (*Juglans nigra*) are calculated from English walnuts (*Juglans regia*) on a solids basis. Magnesium, potassium, zinc, copper, vitamin B-6, and folate in Asian pears (*Pyrus pyrifolia*) were taken directly from these nutrients for pears (*Pyrus communis*). Folate and vitamin B-6 values for beef heart were used for veal heart based on protein content since these are the same species but of different maturity.

It is important to consider the part of the plant (e.g., leaf, root, stem) when selecting a similar food for estimating nutrient values (Gebhardt, 1992). For example in the *Brassica* genus, it is appropriate to use nutrients from turnips for rutabagas since both are root vegetables, whereas it is not appropriate to use nutrient values from cabbage, a leaf vegetable. Color of the vegetable is also an important consideration when imputing the carotenoid or vitamin A content of a vegetable since the level of carotenoid is often related to the green or orange color. The vitamin A content of a dark green vegetable, such as broccoli, would not be similar to the amount of vitamin A in a white vegetable, such as cauliflower, even though both are in the same genus. Other factors that may contribute to nutrient variability among foods of the same family or genus are growing conditions, geographic location, plant maturity, processing or preparation, fortification or other additives, or cut of meat (Rand *et al.*, 1991).

Compilers of the USDA nutrient databases have developed generic nutrient profiles for some groups of food. These groupings facilitate the estimation of nutrient values for foods within the group that have not yet been chemically analyzed for some nutrients that may be of interest. For example, selected vitamins and minerals for the tropical fruits acerola, carambola, passionfruit, and sapodilla were derived from the available values for other tropical and subtropical fruits (Gebhardt, 1992).

2. Calculating Nutrient Values from a Different Form of the Same Food

(a) Nutrient values calculated for a cooked or processed food from its raw values. Frequently, a nutrient value is known for a raw food, but the nutrient has not been chemically analyzed for the same food after cooking. To estimate the nutrient content of the cooked food, consideration must be given to both the cooked yield and the nutrient retention (McCarthy, 1992). Nutrient retention factors for vitamins and minerals have been determined for groups of foods and various cooking methods by the USDA, using data from paired samples of raw and cooked foods and the following formula (Murphy *et al.*, 1975):

$$\%\text{ true nutrient retention} = \frac{\text{nutrient content per g cooked food} \times \text{g cooked food}}{\text{nutrient content per g raw food} \times \text{g raw food}} \times 100.$$

A table of nutrient retention factors for nine vitamins and nine minerals in 260 food groups has been published by the USDA for use with their survey database (USDA, 1993). Additional nutrient retention tables are included in editions of the USDA Agriculture Handbook 8 Series, *Composition of Foods* (USDA, 1976–1993), dealing with beef, pork, lamb, veal, game, fish, poultry, vegetables, and legumes. The retention factors include nutrient losses due to heating and losses due to draining. If there is an evaporative loss or moisture gain, a yield factor must also be applied (Rand *et al.*, 1991). Sources of yield data include the USDA publications, Agriculture Handbook 102: *Food Yields Summarized by Different Stages of Preparation* (USDA, 1975) and the Agriculture Handbook 8 Series, *Composition of Foods* (USDA, 1976–1993), as well as international databases, such as McCance and Widdowson's *The Composition of Foods* (Holland *et al.*, 1991). The following example illustrates use of the nutrient retention factors and cooked yields in estimating nutrient values: raw halibut contains 0.344 mg vit B6 per 100 g; nutrient retention of vit B6 in broiled halibut is 90%; yield of broiled halibut is 73%; 0.344 mg vit B-6 × 90% retention = 0.310 mg vit B6; 0.310 mg vit B-6 per 73 g cooked yield = 0.424 mg vit B6 per 100 g broiled halibut.

For some nutrients, the retention is assumed to be 100% during cooking. In these cases, only the yield factor needs to be applied to determine the nutrient content per 100 g of cooked food: assume no protein is lost during cooking of halibut; raw halibut contains 20.81 g protein per 100 g; yield of broiled halibut is 73%; 20.81 g protein remains in 73 g broiled halibut = 28.51 g protein per 100 g broiled halibut.

Percentage solids of both the raw and cooked food can also be used to estimate a nutrient value of a cooked food from that of a raw food: raw kidney beans contain 11.75% water or 88.25% solids; cooked kidney beans contain 66.94% water or 33.06% solids; raw kidney beans contain 394.1 mcg folacin per 100 g; nutrient retention of folacin (boiled 45–75 min, drained) = 45%; 394.1 mcg folacin per 88.25 g solids in raw beans × 33.06 g solids in ckd beans per 100 g cooked beans = 147.6 mcg folacin per 100 g cooked beans; 147.6 mcg folacin × 45% retention = 66.4 mcg folacin per 100 g cooked beans.

(b) Nutrient values calculated from data on separable parts. Nutrient values for

cuts of meat or poultry may be calculated from proportions of the separable flesh, fat, and/or skin based on physical composition data. The USDA Agriculture Handbook 8 Series (USDA, 1976–1993) provides yield data for percentage of flesh, fat, and/or skin for poultry, pork, beef, lamb, and veal in various meat cuts or poultry parts. Posati (1985) describes a typical nutrient calculation based on the physical composition data in raw chicken thigh with skin:

Tissue	Percentage of cut (edible portion) (g/100 g)	Cholesterol (mg/100 g)
Thigh meat	73.0	83
Skin	16.1	109
Separable fat	10.9	58

mg cholesterol per 100 g raw chicken thigh with skin

$$= (83 \times 0.730) + (109 \times 0.161) + (58 \times 0.109) = 84 \text{ mg}$$

3. Calculating Nutrient Values from Other Components in the Same Food

Specific calculations may be used to derive a nutrient value from one or more nutrients with a known value. Examples of calculations for selected nutrients are listed below.

(a) Energy. The Atwater method of energy calculation uses factors to calculate energy from protein, fat, carbohydrate, and alcohol (Merrill *et al.*, 1973).

(1) The energy calculation using general Atwater factors is the following:

$$\begin{aligned} \text{energy (kcal)} &= (4 \text{ kcal/g protein} \times \text{g protein}) \\ &+ (9 \text{ kcal/g fat} \times \text{g fat}) \\ &+ (4 \text{ kcal/g carbohydrate} \times \text{g carbohydrate}) \\ &+ (7 \text{ kcal/g alcohol} \times \text{g alcohol}) \end{aligned}$$

(2) The British food tables use 3.75 kcal per g carbohydrate instead of the 4 kcal per g (Holland *et al.*, 1991) in the above calculation.

(3) More specific energy factors based on energy availability for different types of foods are used by the USDA to calculate energy values in their nutrient tables (USDA, 1976–1993). Following are a few examples of specific Atwater factors:

	Protein (kcal/g)	Fat (kcal/g)	Carbohydrate (kcal/g)
oats	3.62	8.37	4.12
oat bran	1.82	8.37	2.35
milk	4.27	8.79	3.87
carrots	2.78	8.37	3.84
beef	4.27	9.02	3.68

(b) Protein. Protein values are commonly calculated from the analyzed nitrogen content of a food. The general conversion factor is 6.25 g protein per gram of nitrogen. Protein values in the USDA nutrient tables are calculated from nitrogen using the factors of Jones which are more specific for different food categories (Jones, 1941).

(c) Carbohydrate. Nutrient databases may provide values for ''total carbohydrate'' or for ''available carbohydrate.'' Total carbohydrate values in the USDA tables are calculated by difference using the following formula for 100 g of food (Merrill *et al.*, 1973):

$$\text{carbohydrate} = 100\text{ g} - (\text{g protein} + \text{g fat} + \text{g alcohol} + \text{g ash} + \text{g water}).$$

Carbohydrate calculated in this manner includes dietary fiber, as well as other components of a food that are not protein, fat, alcohol, ash, or water.

Some nutrient database compilers calculate total carbohydrate from the sum of sugars, starch, oligosaccharides, and dietary fiber (Klensin *et al.*, 1989). Others calculate available carbohydrate from the sum of sugars, dextrins, starch, and glycogen (Holland *et al.*, 1991).

(d) Dietary fiber. Dietary fiber may be calculated from the sum of soluble and insoluble fiber. Likewise, if values for dietary fiber and one fiber fraction are known, the other fraction can be calculated by subtraction. There is no accepted method for calculating dietary fiber from crude fiber (Rand *et al.*, 1991).

(e) Vitamin A. Vitamin A can be calculated as the sum of vitamin A activity of retinol and the provitamin A carotenoids which are expressed below as beta-carotene equivalents (FAO/WHO, 1967; National Research Council, 1989).

$$\text{vitamin A (RE)} = \text{mcg retinol} + 1/6\ (\text{mcg beta-carotene equivalents})\ \textit{or}$$

$$\text{vitamin A (IU)} = (\text{mcg retinol} \times 3.33) + (\text{mcg beta-carotene equivalents} \times 1.67).$$

(f) Beta-carotene equivalents. Beta-carotene equivalents describe the vitamin A activity of the provitamin A carotenoids in a food in terms of beta-carotene. When the amount of each carotenoid is known, beta-carotene equivlents may be calculated as follows (FAO/WHO, 1967):

$$\text{beta-carotene equivalents} =$$

$$\text{mcg beta-carotene} + 1/2(\text{mcg alpha-carotene} + \text{mcg beta-cryptoxanthin}).$$

Beta-carotene equivalents may also be calculated from total vitamin A when the amount of retinol is known or when it is assumed to be zero (see vitamin A calculations above). The proportion of retinol and beta-carotene equivalents in a food can be estimated using Table 1, ''Estimated Distribution of Sources of Vitamin A Activity in Various Foods'' developed by INCAP/ICNND (Leung *et al.*, 1961).

(g) Vitamin E. Vitamin E, expressed as alpha-tocopherol equivalents, can be calculated from the activity of tocopherols and tocotrienols using various conversion factors derived from different assay methods. Two common formulas used to calculate vitamin E in terms of alpha-tocopherol equivalents are the following:

For individual foods (McLaughlin and Weihrauch, 1979):

$$\text{mg alpha-tocopherol equivalents} = \text{mg alpha-tocopherol} + 0.4\ (\text{mg beta-tocopherol})$$

TABLE 1

ESTIMATED DISTRIBUTION OF SOURCES OF VITAMIN A ACTIVITY IN VARIOUS FOODS

	From Vitamin A	From Vitamin A Precursors:	
		Beta-carotene	Carotenoids other than Beta
Animal Origin:			
Meat and meat organs	90	10	
Poultry	70	30	
Fish and Shellfish	90	10	
Eggs	70	30	
Milk and milk products	70	30	
Animal or fish oil	90	10	
Plant origin:			
Cereals:			
Maize, yellow		40	60
Others		50	50
Legumes and seeds		50	50
Vegetables:			
Green vegetables		75	25
Deep yellow/orange		85	15
Pale yellow/orange		50	50
Other vegetables		50	50
Fruits:			
Deep yellow		85	15
Other fruits		75	25
Vegetable oils:			
Red palm oil		65	35
Other vegetable or seed oils		50	50

Note. Adapted from Leung, W. W., and Flores, M. (1961) Food composition table for use in Latin America. The Institute of Nutrition of Central America and Panama and the Interdepartmental Committee on Nutrition for National Defense, National Institutes of Health.

+ 0.1 (mg gamma-tocopherol) + 0.01 (mg delta-tocopherol)
+ 0.3 (mg alpha-tocotrienol) + 0.05 (mg beta-tocotrienol)
+ 0.01 (mg gamma-tocotrienol).

For mixed diets (National Research Council, 1989):
mg alpha-tocopherol equivalents =
mg alpha-tocopherol + 0.5 (mg beta-tocopherol)
+ 0.1 (mg gamma-tocopherol) + 0.3 (mg alpha-tocotrienol).

The following formula can be used to convert IU of vitamin E to mg of alpha-tocopherol equivalents (American Pharmaceutical Association, 1975):

1.49 IU vitamin E = 1 mg alpha-tocopherol equivalents *(h) Fatty acids.* For a food which contains only one ingredient fat, individual fatty acid values can be calculated from the proportion of the ingredient fat in the food and its fatty acid profile.

Example: salad dressing made with soybean oil.
100 g salad dressing = 9.8 g total fat
100 g soybean oil = 100 g total fat
100 g soybean oil = 10.3 g palmitic acid, 3.8 g stearic acid, 22.8 g oleic acid, etc.

palmitic acid/100 g salad dressing = 10.3 g palmitic acid per 100 g fat in soybean oil × 9.8 g fat/100 g dressing
= 1.0 g palmitic acid per 100 g salad dressing.

(i) Amino acids. Similar to the fatty acid calculation shown above, the amino acid profile for a food with a single protein source can be calculated from the protein-containing ingredient. In this case, the amount of the protein contributor is calculated from the total protein of the food and that of the ingredient (Tietz, 1990).

Example: flour tortilla contains flour as the only protein-containing ingredient.
100 g flour tortilla = 7.32 g protein
100 g flour = 10.50 g protein
100 g flour = 1.34 g proline, 0.46 g serine, etc.
proline/100 g tortilla
= 1.34 g proline per 10.5 g protein in flour × 7.32 g protein in tortilla
= 0.93 mg proline per 100 g tortilla

4. Calculating Nutrient Values from Household Recipes or Commercial Product Formulations for Multicomponent Foods

Nutrient values for multicomponent foods can be calculated by summing nutrients of all ingredients contained in the food. For some multicomponent foods, recipes which provide a list of ingredients and their amounts are available. For other foods, such as commercial products, ingredient lists, but not ingredient amounts, are available from the food manufacturer. The following procedures are used to calculate nutrients from recipes and commercial product formulations:

(a) Recipes. The most frequently used methods for calculation of nutrients from recipes are the Yield Factor Method and the Retention Factor Method (Powers *et al.*, 1989). USDA uses the retention factor method for recipes in both Agriculture Handbook No. 8 and the Survey Nutrient Database (Perloff, 1985). Rand *et al.* (1991) provide the following detailed procedure which combines the steps used in both methods and give specific examples of their use:

(1) Select or develop an appropriate recipe.
(2) Collect weight and nutrient content data for each ingredient.
(3) Correct ingredient nutrient levels for weight of edible portions where appropriate.
(4) Correct ingredients for effects of cooking:
if data for cooked ingredients are available, use yield factors to adjust from raw to cooked weights; if data for cooked ingredients are not available, use

TABLE 2

DETERMINATION OF A FORMULATION FOR WHOLE WHEAT CRACKERS FROM INGREDIENTS AND KNOWN NUTRIENT VALUES

ingredient	amount	kcal	protein (g)	fat (g)	carbohydrate (g)	sodium (mg)
whole wheat	88 g	301	9.9	1.8	66.3	2
soybean oil	10.2 g	90	0	10.2	0	0
salt	0.4g	0	0	0	0	155
total based on formulation	**98.6g**	**391**	**9.9**	**12.0**	**66.3**	**157**
total based on product label	**100g**	**400**	**9.7**	**12.0**	**67.7**	**161**

data for uncooked ingredients, applying yield factors to adjust for weight changes, and retention factors for nutrient losses or gains during cooking.

(5) Sum weights of ingredients to get weight of recipe.
(6) Sum nutrient values of ingredients to obtain nutrient value of recipe.
(7) Adjust recipe weight and nutrient levels to reflect changes in fat/water when whole mixture is cooked; make any additional refuse adjustments; apply retention factors if available for whole recipe.
(8) Determine the quantity of prepared food produced by the recipe.
(9) Determine the final values per weight (e.g., per 100 g), volume (e.g., per cup), or serving portions as desired.

Variations of these two basic methods of recipe calculation have been described by several authors (Petot, 1993; Powers *et al.*, 1989). Hoover (1994) has proposed an Integrated Recipe Model which combines the attributes of both the Yield Factor Method and the Retention Factor Method into a single database structure to provide the flexibility of using either procedure for calculation of recipe nutrients from one software program.

(b) Commercial product formulations. The first step in calculating missing nutrient values for multicomponent commercial food products is to obtain a list of ingredients in the food product and any known nutrient values for the food. This information may be provided by the food manufacturer or obtained from the product label. Using this information, the second step is to estimate the amount of each ingredient in the product so that the ingredient nutrients can be totaled.

An example is a commercial cracker made from whole wheat, partially hydrogenated soybean oil, and salt. Table 2 illustrates determination of the cracker formulation. Ingredients are listed in descending order by weight as found on the product label. The nutrient information provided by the manufacturer is entered in the last row titled ‘‘total based on product label.’’ The goal is to determine the proportion of each ingredient in the cracker so that nutrient totals for the combined ingredients approximate the known values for the cracker. Since whole wheat provides all of the protein and carbohydrate, we can estimate that the cracker contains between 86.6 and 90.0% whole wheat based on the following considerations:

Nutrient content per 100 g of crackers is 400 kcal, 9.7 g protein, 12.0 g fat, 67.7 g carbohydrate, and 161 mg sodium. Whole wheat, which has 11.2 g protein per 100 g and 75.2 g carbohydrate per 100 g, provides all of the protein and carbohydrate in the cracker. Therefore, the whole wheat is between 86.6 and 90.0% of the cracker (9.7/11.2 = 86.6% whole wheat based on protein content or 67.7/75.2 = 90.0% whole wheat based on carbohydrate content). Thus, an estimated 88 g of whole wheat was used in the formulation.

The soybean oil provides the remainder of the fat to equal the 12.0 g total fat in the cracker, and the salt provides most of the necessary sodium. Once the formulation of ingredients and ingredient amounts has been determined, it can be used to calculate the other nutrient values for the product.

Commercial food products often contain many more ingredients than in the example shown above. Computer programs have been developed to estimate product formulations using mathematical optimization techniques of linear programming (Marcoe *et al.*, 1992; Westrich, 1993; Westrich *et al.*, 1994). Without such a program, estimates of ingredient proportions may be done by trial-and-error until the calculated nutrients are sufficiently close to the known nutrient values of the food (Westrich *et al.*, 1994). To make the trial-and-error approach feasible to use, it is often necessary to limit the number of ingredients to those expected to be the major contributors of nutrients in the product.

5. *Converting Nutrient Values from Nutrient Label Information of a Commercial Food Product*

In countries where there are voluntary or mandatory nutrition labeling regulations, nutrient information may be available from the package label of a commercial food product. If the original analytic data cannot be obtained from the manufacturer, the information provided by the package label can be used to convert nutrient values to a 100-g basis. Nutrition labeling regulations can be quite specific and vary by country. Labeling values may have been manipulated to be in compliance with labeling regulations for rounding and therefore may not be in precise agreement with the original analytic data (Gebhardt, 1993; Rand *et al.*, 1991).

Example: iron in XYZ brand of peanut butter cookies.

From the package label, iron content is 4% of the Daily Value (DV) or Reference Daily Intake (RDI) for one serving of cookies. The RDI for iron in the U.S. is 18 mg. Serving size is two cookies weighing 26 g.

$$\text{mg iron per serving of cookies} = 4\% \times 18 \text{ mg iron} = 0.72 \text{ mg}$$
$$\text{mg iron per 100 g of cookies} = 0.72 \text{ mg iron per 26 g cookies} \times 100 \text{ g} = 2.77 \text{ mg}$$

6. *Calculating Nutrient Values from a Product Standard*

When a food item has nutrients added for enrichment or fortification purposes, the value for the nutrient can be estimated from a product standard if the analytic data are not available. Product standards will vary by country. In the United States, the standards of identity are published yearly by the Food and Drug Administration in the *Code of*

Federal Regulations (Office of the Federal Register, 1995). For added nutrients in certain foods, a minimum and maximum level is given. For example, macaroni which is labeled as ''enriched'' must contain not less than 4 mg or more than 5 mg of thiamin per pound. If it is unknown whether the nutrient is usually added near the minimum or maximum level, the midpoint of the range can be used. For added nutrients in other foods, a single figure may be the minimum level, with averages left to good manufacturing practices. Vitamin A in milk must be added at a level not less than 2000 International Units per quart. If a nutrient is listed as ''optional,'' the usual practice of the industry in the particular country must be known to estimate the nutrient in a typical food sample. This may change over time. For example, in the U.S. enriched flour must contain thiamin, riboflavin, niacin, and iron (as well as folic acid after Jan. 1, 1998) at specified levels. Calcium, however, is an optional added nutrient. During a period of the 1980s, most of the flour produced for the retail market contained added calcium; however, currently calcium is not usually added to enriched flour.

7. *Using an Assumed Zero*

Assumed zero values that are based on knowledge that a food would not contain the nutrient are not considered to be ''estimated'' values. Assumed zeros used for components such as cholesterol and vitamin B12 in plant products or dietary fiber in animal products fit into this category. However, in some cases, assumed zeros are used as estimates of nutrients that are not expected to be present in significant amounts, but by sensitive analytical techniques are not absolute zero. For example, vitamin C in cheese and vitamin A in white cornmeal may be estimated as zeroes although very small amounts may be detected through chemical assay.

TECHNIQUES FOR VALIDATION OF ESTIMATED VALUES

The update of nutrient values in a food composition database requires implementation of quality control procedures to verify that appropriate data selection, calculation methods, and data entry were used. In addition, procedures are needed to ensure the consistency of the values within the entire database. It is especially important to validate estimated values, since their derivation often depends on assumptions made by the database compiler, as well as a series of calculation procedures. Buzzard *et al.* (1995) describe quality control procedures applicable to validation of estimated values. These include manual review of the rationale and calculations used to derive each value; edit checks at the point of data entry to identify keying errors; consistency among nutrient values of similar foods; and consistency of related nutrient values within each food.

Many of these quality control procedures can be automated. For example, edit limits, which define the usual minimum and maximum nutrient values per 100 g of food within a food group, can be used during data entry to flag values that fall outside the limits (Buzzard *et al.*, 1986). Flagged values are investigated for possible errors. Murphy (1989) describes the development of edit limits for 29 nutrients in the University of California-Berkeley's nutrient database (UCB Minilist), using data from the USDA Survey Database. She selected the 1st and 99th percentile nutrient values of foods in each of 12 food groups as the minimum and maximum edit limits for the UCB Minilist.

Before release of a new version of the database, validation programs should be run to check for consistency within the database. Reports that compare nutrient values of similar foods (e.g., calcium in all types of yogurt or iron in various beef cuts) can be reviewed for outliers. Other quality control programs verify database consistency by comparing calculated algorithms with expected values for each food in the database (Sievert *et al.*, 1989). For example, the sum of protein, carbohydrate, fat, alcohol, water, and ash is compared with 100 g. Calculated energy values based on 4 kcal per gram of protein and carbohydrate, 9 kcal per gram of fat, and 7 kcal per gram of alcohol are compared with energy values in the database. Other comparisons include the sum of fatty acids with total fat; the sum of sugars, starch, and dietary fiber with total carbohydrate; and the sum of amino acids with total protein. Discrepancies between the calculated algorithms and the database values that fall outside acceptable ranges are investigated. Examples of acceptable differences established for the nutrient database at the University of Minnesota are given by Buzzard *et al.* (1995).

The most desirable validation of nutrient estimation methods is direct comparison of the estimated values with values obtained by chemical analysis for the same food. The Human Nutrition Information Service (HNIS) of USDA checked the reliability of their recipe calculation procedures by comparing calculated and analyzed values for proximate nutrients, 8 minerals, 6 vitamins, and cholesterol in six mixed dishes (Matthews, 1988). Differences between calculated and analyzed values for proximate nutrients and minerals were generally less than 10%, except for copper which showed greater differences and was most often higher when chemically analyzed. Vitamins showed differences of less than 20% in most cases with the exception of vitamin B12. Analyzed values for cholesterol were always lower than calculated values. However, in this case, Matthews suggests that the problem is not with the calculation method, but rather with the difficulty in extraction of cholesterol from food mixtures during chemical assay.

Other studies have compared estimated and analyzed values for dietary fiber (Thorsen, 1991; Westrich *et al.*, 1994), linoleic acid (Westrich *et al.*, 1994), and sucrose (Thorsen, 1991). Differences between the estimated and analyzed values for these food components were attributed mostly to lack of appropriate commercial ingredients in the database used to estimate nutrient values. Thorsen and Westrich showed their calculation methods underestimated dietary fiber by about 10%. Some high fiber commercial ingredients (e.g., corn bran) were not available in the database and were substituted in the estimation procedures with similar ingredients, but containing less fiber (e.g., cornmeal). Linoleic acid was overestimated by approximately 10% using the linear programming techniques of Westrich. This overestimation was most likely due to a lack of appropriate commercial oils in the ingredient database used to calculate the fatty acid content of commercial food products. Commercial oils are often more hydrogenated than the retail oils included in the database. A similar problem was encountered by Thorsen in calculating sucrose values for commercial food products made with commercial corn syrups, which were not available in the ingredient database used to estimate nutrient values. Commercial corn syrups often differ from retail varieties with respect to proportions of mono- and disaccharide content.

DOCUMENTING AND REFERENCING ESTIMATED VALUES

All estimated nutrient values should be referenced as such in the database. This enables the database compiler to identify foods and nutrients that might require chemical assay at some future time. It also provides database users with an indication of the types of data in the database and whether the data are appropriate for their needs.

A single reference code may be used to identify all estimated values or individual reference codes for each nutrient estimation method may be assigned. An example of a database reference system has been described by Schakel *et al.* (1988).

A four position letter code is being developed at USDA to give more specific information about procedures used to estimate values when analytic data are not available (Gebhardt, 1995). The first position identifies the major method of estimation used, such as based on a different but similar food or estimated from a product formulation. The second, third, and fourth positions, if needed, identify the procedures used in progressively more specific detail, such as, ''the use of retention factors'' or ''calculations based on total solids content.''

For some nutrients there may be more than one accepted method of calculation; therefore, it is important for the database compiler to specify the method used when reporting these data. A system for standardized identification of food components and methods used in their derivation has been established by INFOODS of the United Nations University (Klensin *et al.*, 1989). These ''tagnames'' define the food component and, if relevant, reference the calculation procedure used to determine the value. Detailed documentation of all assumptions and calculations used in the derivation of the values should be maintained by the database compiler and included either within the structure of the database or made available to users of the database upon request.

REFERENCES

American Pharmaceutical Association (1975). *National Formulary XIV.* Washington, DC.

Buzzard, I. M., Sievert, Y. A., and Schakel, S. F. (1986). Database validation procedures. *11th National Nutrient Databank Conference Proceedings,* June 29–July 2, 1986. Athens, GA.

Buzzard, I. M., Schakel, S. F., and Ditter-Johnson, J. (1995). Quality control in the use of food and nutrient databases for epidemiologic studies. In *Quality and Accessibility of Food-Related Data* (H. Greenfield, Ed.), pp. 241–252. AOAC International, Arlington, VA.

FAO/WHO (Food and Agriculture Organization/World Health Organization) (1967). *Requirements of Vitamin A, Thiamine, Riboflavin, and Niacin.* Report of a Joint Food and Agriculture Organization/World Health Organization Expert Committee. FAO Nutrition Meetings Report Series No. 41. WHO Technical Report Series No. 362. World Health Organization, Geneva.

Gebhardt, S. E. (1995). Derivation codes to identify sources of nonanalytical nutrient data. Abstract. *Second International Conference on Dietary Assessment Methods,* January 22–24, 1995. Boston, MA.

Gebhardt, S. E. (1992). Imputing values for nutrient data bases. *17th National Nutrient Databank Conference Proceedings,* June 7–10, 1992. Baltimore, MD.

Gebhardt, S. E. (1993). Issues related to increasing brand names in the survey nutrient data base. *18th National Nutrient Databank Conference Proceedings,* May 23–26, 1993. Baton Rouge, LA.

Holland, B., Welch, A. A., Unwin, I. D., Buss, D. H., Paul, A. A., and Southgate, D. A. T. (1991). *McCance and Widdowson's The Composition of Foods,* 5th ed. Royal Society of Chemistry, Cambridge, UK.

Hoover, L. W. (1994). A new recipe calculation model. *19th National Nutrient Databank Conference Proceedings,* May 22–24, 1994. St. Louis, MO.

Jones, D. B. (1941). Factors for converting percentages of nitrogen in foods and feeds into percentages of protein. USDA Cir. 183.

Klensin, J. C., Feskanich, D., Lin, V., Truswell, A. S., and Southgate, D. A. T. (1989). *Identification of Food Components for INFOODS Data Interchange.* United Nations University, Tokyo.

Leung, W. W., and Flores, M. (1961). *Food Composition Table for Use in Latin America.* The Institute of Nutrition of Central America and Panama, and the Interdepartmental Committee on Nutrition for National Defense. National Institutes of Health, Bethesda, MD.

Marcoe, K. K., and Haytowitz, D. B. (1992). A method for estimating nutrient values of mixed dishes based on ingredient and nutrition label information. *17th National Nutrient Databank Conference Proceedings,* June 7–10, 1992. Baltimore, MD.

Matthews, R. H. (1988). Recipe calculations vs analytical values. *Proceedings of the Thirteenth National Nutrient Databank Conference,* June 6–8, 1988. Framingham, MA.

Merrill, A. L., and Watt, B. K. (1973). Energy value of foods—Basis and derivation. U.S. Department of Agriculture, Agriculture Handbook No. 74.

McCarthy, M. A. (1992). Retention and yield factors. *17th National Nutrient Databank Conference Proceedings,* June 7–10, 1992. Baltimore, MD.

McLaughlin, P. J., and Weihrauch, J. L. (1979). Vitamin E content of foods. *J. Am. Diet. Assoc.* **75,** 647–665.

Murphy, E. W., Criner, P. E., and Gray, B. C. (1975). Comparisons of methods for calculating retentions of nutrients in cooked foods. *J. Agric. Food Chem.* **23,** 1153–1157.

Murphy, S. P. (1989). Integrity checks for nutrient data. *Proceedings of the Fourteenth National Nutrient Databank Conference,* June 19–21, 1989. Iowa city, IA.

National Research Council (U.S.) (1989). *Recommended Dietary Allowances,* 10th ed. National Academy Press, Washington, DC.

Office of the Federal Register (1995) *Code of Federal Regulations,* 21CFR. U.S. Government Printing Office, Washington, DC.

Perloff, B. P. (1985). Recipe calculations for the NFCS data base. *Proceedings of the Tenth National Nutrient Data Bank Conference Proceedings,* July 22–24, 1985. San Francisco, CA.

Petot, G. J. (1993). Yield factor and summation methods. *18th National Nutrient Databank Conference Proceedings,* May 23–26, 1993. Baton Rouge, LA.

Posati, L. P. (1985). Procedures for imputing values. *Proceedings of Tenth National Nutrient Data Bank Conference,* July 22–24, 1985. San Francisco, CA.

Powers, P. M., and Hoover, L. W. (1989). Calculating the nutrient composition of recipes with computers. *J. Am. Diet. Assoc.* **89,** 224–232.

Rand, W. M., Pennington, J. A. T., Murphy, S. P., and Klensin, J. C. (1991). *Compiling Data for Food Composition Data Bases.* United Nations University Press, Tokyo.

Schakel, S. F., Sievert, Y. A., and Buzzard, I. M. (1988). Sources of data for developing and maintaining a nutrient database. *J. Am. Diet. Assoc.* **88,** 1268–1271.

Schakel, S. F., Warren, R. A., and Buzzard, I. M. (1989). Imputing nutrient values from manufacturers' data. *Proceedings of the Fourteenth National Nutrient Databank Conference,* June 19–21, 1989. Iowa City, IA.

Sievert, Y. A., Schakel, S. F., and Buzzard, I. M. (1989). Maintenance of a nutrient database for clinical trials. *Contr. Clin. Trials* **10,** 416–425.

Tietz, S. P. (1990). Incorporation of Amino Acid Values and Animal and Vegetable Protein Source Values into the Nutrient Database at the Nutrition Coordinating Center. Thesis for the Degree of Master of Science, University of Minnesota.

Thorsen, J. A. (1991). Validity of Imputed Nutrient Values in a Nutrient Database. Master's Research Project, Nutrition Coordinating Center, University of Minnesota.

USDA (1976–1993). *Composition of Foods—Raw, Processed, Prepared.* Agriculture Handbook No. 8, revisions.

USDA (1993). *Computerized Data Sets Used to Create Release 6 of the USDA Nutrient Data Base for Individual Food Intake Surveys.* USDA Table of Nutrient Retention Factors.

USDA (1975). *Food Yields Summarized by Different Stages of Preparation.* Agriculture Handbook No. 102.

Westrich, B. J. (1993). Development and Evaluation of Mathematical Optimization Software for Estimation of Nutrient Values in Food Products. Thesis for the Degree of Doctor of Philosophy, University of Minnesota.

Westrich, B. J., Buzzard, I. M., Gatewood, L. C., and McGovern, P. G. (1994). Accuracy and efficiency of estimating nutrient values in commercial food products using mathematical optimization. *J. Food Comp. Anal.* **7,** 223–239.