Imputing Values for Nutrient Data Bases

Susan E. Gebhardt
Human Nutrition Information Service

The Nutrient Data Research Branch supplies nutrient values for the various nutrient data bases that are available from the Human Nutrition Information Service. The number of nutrients and food items for which data must be supplied is ever increasing. There are currently about 3,300 items in the Primary Nutrient Data Set (PDS) that is used to generate the Survey Nutrient Data Base. Twenty-eight nutrients for each of these 3,300 foods means approximately 99,000 separate values. Analytical data are not available for every nutrient in every food. Some unusual foods have never been analyzed, and even some common foods have not been analyzed for all of these nutrients. However, to analyze food intake data for nutrient content requires a complete data base with nutrient values for all foods consumed to ensure that nutrient intake will not be underestimated. Therefore, to provide a complete data base we must calculate or impute nutrient values for many foods.

Data bases may contain three types of nutrient values--analytical values, calculated values, and imputed values. The distinction between analytical and calculated values and between calculated and imputed values is not always clear cut. Analytical values are arithmetic or weighted means of values obtained by chemical or biological assays.

Calculated values are not themselves averages of analytical data but are derived from analytical data using specific procedures. Calculated values are no less valid than the analytical data on which they are based and the soundness of the calculation procedures. A paper delivered at the 10th Nutrient Data Bank Conference discussed categories for calculating nutrient values from analytical data based on:

Physical composition data
Weighting factors
Content of a nutrient in a nutrient fraction of a similar food
Regression analysis
Retention and yield data
Recipes

Imputed values are quite often calculated values, but they also require additional assumptions. Nutrients for a multi-ingredient food, such as a frozen entree, would be calculated by estimating the proportion and nutrient contribution of each ingredient. This paper will concentrate on imputed values for single- ingredient foods. For foods that are major contributors of a nutrient to the diet--those foods in the group that contribute 80 percent or more of a nutrient to the American diet--our goal is to have analytical data. These foods have been identified, and where we used imputed values, these items have been or will be included for analysis on key food contracts. However, that leaves hundreds of nutrients in other foods that we cannot afford to have analyzed but for which a nutrient value must be supplied. Each food specialist is responsible for filling in missing nutrient values for items in their food group.

Several general procedures are used to fill in missing values:

1) Search for nutrient values in foreign food tables.

If a food is more commonly consumed in another country, nutrient values may have been published in a table of food composition developed for that country or area of the world. This approach has been used to find the thiamin content of crayfish (taken from the German table of food composition), nutrients in Asian pears (Japanese table of food composition), and nutrients in hearts of palm (East Asia table of food composition). Often, however, some of the nutrients that are needed are not reported in the foreign table. Vitamin B-6, folate, zinc, and copper are nutrients that are frequently missing.

2) Impute values from another form of the food.

Many times nutrient values are available for the raw form of a food but not for the cooked or processed forms. When values for a cooked form of a food are calculated from the raw form, differences in yield and nutrient retention must be taken into account. In this example (Table 1), the amounts of vitamins C and B-6 in cooked edible-podded peas are calculated from these nutrients in raw edible-podded peas by using a factor based on the change in total solids and applying retention factors of 80 percent for vitamin C and 90 percent for vitamin B-6.

Retention factors are not available for all of the many different types of processing that are used for different types of foods. In the following examples, more assumptions have to be made than in the previous example. Values for a similarly processed food are used without taking additional losses for retention. Values for vitamins C and B-6 in dry-roasted pecans were taken from dried pecans (Table 2). The nutrients are calculated on the total solids basis. Grams of solids in the dry-roasted pecans are divided by the grams of solids in the dried pecans to give the factor for calculating nutrients from dried pecans to dry-roasted pecans.

Folate in smoked haddock is calculated from folate in raw haddock (Table 3). A factor based on the change in protein content is used.

3) Impute values, based on taxonomy, from another species within a genus.

Vitamin B-6 and folate in black walnuts (*Juglans nigra*) are calculated from English walnuts (*Juglans regia*) on the solids basis (Table 4).

Magnesium, potassium, zinc, copper, vitamin B-6, and folate in Asian pears (*Pyrus pyrifolia*) were taken directly from these nutrients for pears (*Pyrus communis*) (Table 5). These were the nutrients that were not available from the Japanese table.

The iron value for cisco (*Coregonus artedii*) was taken directly from whitefish, mixed species (*Coregonus spp.*) (Table 6).

Folate and vitamin B-6 values for beef heart were used for veal heart based on protein content (Table 7). These are the same species but the maturity of the animal is different.

When using this procedure for imputing nutrients for vegetables, the part of the plant--leaf, root, or inflorescence--also should be considered. Even though rutabaga and cabbage are both in the *Brassica* genus, different parts of the plant are consumed, so it would be more appropriate to impute values for rutabaga from turnips, another root vegetable in the *Brassica* genus. For some nutrients such as vitamin A, which is based on the provitamin A carotenoid content of a plant and is known to be related to color, the color of the food must be considered. It would be inappropriate to use a vitamin A value for broccoli (dark green) for cauliflower (white), even though they are both *Brassica oleracea.*

4) Impute values, based on taxonomy, from another genus within the same family.

The cholesterol value for Atlantic herring was used for European anchovy (both *Clupeidae*) (Table 8). Another consideration to take into account when choosing a similar food is to be conservative in the estimate. A food that is a particularly high source of a nutrient is probably not the best choice to use for a similar food if there are other options available. For example, the zinc content of Eastern oysters, which is extremely high, would not be the best choice for zinc content in another mollusk.

5) Impute values from a generic nutrient profile developed for a food group.

Many nutrient values were missing for tropical fruits, and there were no other species within the same genus. Rather than try to decide which other tropical fruit might be similar, a generic profile for tropical fruits was developed by taking the mean of the available values for tropical and subtropical fruits for those nutrients that were frequently missing (Table 9). This same procedure was also used for fish.

Separate generic profiles were developed for finfish, mollusks, and crustaceans (Table 10).

Asian pears are one food for which values were imputed but analytical values are now available. Table 11 shows how the imputed values--some from the Japanese table and some taken from pears--compare with the mean of two analyzed samples. For most nutrients there is fairly good agreement. If looked at on a percentage difference basis, some nutrients--such as fat, thiamin, zinc, and copper--would have large differences. However, when the differences in nutrient contribution compared to the U.S. RDA are considered, the differences are inconsequential.

It is important to recognize that imputed values developed by these procedures are estimates. If it is necessary to have a more precise nutrient value and estimate of variability, the food should be analyzed.

	Peas, edible-podded, raw		Peas, edible-podded, cooked
	per 100 g		
Moisture:	88.885 g		88.910 g
Solids:	11.115 g		11.090 g
Vitamin C:	(60 mg x .998 x .80)	→	47.9 mg
Vitamin B6:	(.160 mg x .998 x .90)	→	.144 mg

$$\frac{\text{g solids } 11.090}{\text{g solids } 11.115} = .998$$

Retention

Table 1

	Pecans, dried		Pecans, dry-roasted
	per 100 g		
Moisture:	4.817 g		1.100 g
Solids:	95.813 g		98.900 g
Vitamin C:	(1.971 mg x 1.039)	→	2.048 mg
Vitamin B6:	(.188 mg x 1.039)	→	.195 mg

$$\frac{\text{g solids } 98.900}{\text{g solids } 95.183} = 1.039$$

Table 2

	Haddock, raw		Haddock, smoked
	per 100 g		
Protein:	18.91 g		25.23 g
Folate:	(11.5 mcg x 1.33)	$\longrightarrow$	15.3 mcg

$$\frac{\text{g protein } 25.23}{\text{g protein } 18.91} = 1.33$$

Table 3

	Juglans regia English walnuts		*Juglans nigra* Black walnuts
	per 100 g		
Moisture:	3.645 g		4.360 g
Solids:	96.355 g		95.640 g
Vitamin B6:	(.558 mg x .993)	$\longrightarrow$	.554 mg
Folate:	(66.0 mcg x .993)	$\longrightarrow$	65.5 mcg

$$\frac{\text{g solids } 95.640}{\text{g solids } 96.355} = .993$$

Table 4

	Pyrus communis Pear		*Pyrus pyrifolia* Asian pear
		per 100 g	
Moisture:	83.81 g		88.6 g
Magnesium:	6 mg	→	6 mg
Potassium:	125 mg	→	125 mg
Zinc:	.12 mg	→	.12 mg
Copper:	.113 mg	→	.113 mg
Vitamin B6:	.018 mg	→	.018 mg
Folate:	7.3 mcg	→	7.3 mcg

Table 5

	Coregonus spp. Whitefish, Mixed species		*Coregonus artedii* Cisco
Moisture:	72.77 g/100 g		78.93 g/100 g
Iron:	.37 mg/100 g	→	.37 mg/100 g

Table 6

	Beef heart		Veal heart
		per 100 g	
Protein:	17.05 g		17.18 g
Vitamin B6:	(.43 mg x 1.01)	→	.43 mg
Folate:	(2 mcg x 1.01)	→	2 mcg

$$\frac{\text{g protein } 17.18}{\text{g protein } 17.05} = 1.01$$

Table 7

Family: Clupeidae

	Clupea harengus harengus Herring, Atlantic, raw:		*Engraulis encrasicholus* Anchovy, European, raw:
Moisture:	72.05 g/100 g		73.37 g/100 g
Cholesterol:	60 mg/100 g	→	60 mg/100 g

Table 8

Generic Nutrient Profile for Tropical and Subtropical Fruits

	MEAN PER 100 g	NO. OF SAMPLES
Vitamin B6	0.97 mg	20
Folate	14.0 mcg	7
Copper	.086 mg	24
Zinc	.10 mg	21

Used for acerola, carambola, passionfruit, and sapodilla.

Table 9

Generic Nutrient Profile for Finfish

	MEAN PER 100 g	RANGE
Vitamin B6	0.4 mg	0.1 - 1 mg
Folate	5 mcg	1 - 10 mcg
Magnesium	30 mg	20 - 60 mg

Used for Dolphinfish, Milkfish, Sablefish, and Sea Bass.

Table 10

		Asian Pear (imputed) *per 100 g*	Asian Pear (analyzed) *per 100 g*	USRDA
from Japanese Table	Moisture:	88.6 g	88.2 g	
	Protein:	.3 g	.5 g	
	Fat:	.1 g	.2 g	
	Carbohydrate:	10.1 g	10.6 g	
	Calcium:	3 mg	4 mg	1,000 mg
	Thiamin:	.03 mg	.01 mg	1.5 mg
	Niacin:	.20 mg	.22 mg	20 mg
from Pears	Magnesium:	6 mg	8 mg	400 mg
	Potassium:	125 mg	121 mg	--
	Zinc:	.12 mg	.02 mg	15 mg
	Copper:	.113 mg	.05 mg	2 mg
	Vitamin B6	.018 mg	.022 mg	2 mg
	Folate:	7.3 mcg	8.0 mcg	400 mcg

Table 11