

## **Documentation for the USDA-NCC Carotenoid Database for U.S. Foods**

The USDA-NCC Carotenoid Database was created through a collaborative effort between the USDA and the Nutrition Coordinating Center (NCC) at the University of Minnesota and was partially funded with grants from the National Cancer Institute (NCI). This is an update of the 1993 USDA-NCI Carotenoid Database. New literature values from 1992 - 1996 including international sources, as well as analytical data generated from foods analyzed at the Food Composition Laboratory, ARS, USDA (unpublished data) were used for the update. Analytical values are reported for alpha-carotene, beta-carotene, beta-cryptoxanthin, lycopene, and lutein+zeaxanthin. The values represent the total isomer content of each carotenoid; isomers were either not separated or, when reported separately, the level of each isomer was combined to give a total value. Since most analytical systems measure lutein and zeaxanthin together, selected foods were analyzed for zeaxanthin separately. The resulting data were combined with literature values and reported in a separate table.

Values for each food from each reference were evaluated using the expert system described by Mangels, et al. ( J. Am. Diet. Assoc. 93:284-296, 1993). The evaluation criteria included five categories: analytic method, analytic quality control, number of samples, sample handling and sampling plan. Each mean is weighted by the sampling plan rating and the aggregated mean is assigned a Confidence Code (CC) of a, b or c. The Confidence Code is an indicator of relative quality of the data and the reliability of a given mean value. A Confidence Code of "a" indicates considerable reliability, due either to a few exemplary studies or to a large number of studies of varying quality.

In the 1993 release, different forms of a single food (e.g., raw, cooked, canned, etc.) with similar values were aggregated, and median values were presented. In the new version, food forms are disaggregated to be consistent with the food description in the USDA Nutrient Database for Standard Reference. Data for U.S. and non-U.S. foods are separated; data for non-U.S. foods will be finalized and released at a later date. The U.S. foods are defined as those foods consumed in the United States, including imported foods. Acceptable data for U.S. foods have been combined to generate mean values. Although the 1993 USDA-NCI Carotenoid Database and Mangels, et al. reported medians as the estimate of central tendency for each food-carotenoid combination, a workshop of statisticians convened by the USDA in 1996 concluded that arithmetic means, due to their additive property, were more appropriate and hence are reported here.

The data for U.S. foods were obtained from many sources and may represent different growing years, growing areas, cultivars, processing techniques, lengths and conditions of storage, and possibly different methods of analysis. All of these factors influence the carotenoid content of foods. It should be noted that the values for different forms of a food were not necessarily derived from the same sample. That is, a single sample of carrots was not analyzed in each of the forms presented, i.e., raw, cooked, canned, and frozen. Therefore, in a comparison of different forms of a food, differences in carotenoid contents should not be ascribed solely to the effect of processing or preparation methods.

The completed database contains three files:

1. Car\_tble is the table of analytical carotenoid values for U.S. foods.
2. Zea\_tble is the table of analytical zeaxanthin values for 22 selected U.S. foods.
3. Car\_ref is the list of references/studies from which carotenoid values were obtained.

These files are available for viewing and/or printing from this web site. A compressed file, containing the complete database and documentation, also is available for downloading from this web site (car98.zip). This will allow you to use the database on your computer with other programs. A detailed description of each of the files follows:

### Car\_tble - Analytical Carotenoid Values for U.S. Foods

Car\_tble contains carotenoid values for 215 U.S. foods.

The fields in the car\_tble are as follows:

NDB	USDA Nutrient Data Bank number <sup>1</sup>
Desc	Food description
Carot	Name of the carotenoid
a_car	alpha-Carotene
b_car	beta-Carotene
b_cryp	beta-Cryptoxanthin
lut+zea	Lutein+Zeaxanthin
lyc	Lycopene
Mean	Weighted mean value (mcg/100g edible portion) <sup>2</sup>
SEM	Standard error of the mean <sup>3</sup>
#S	Number of means/individual values <sup>4</sup>
Min	Minimum value (mcg/100g edible portion)
Max	Maximum value (mcg/100g edible portion)
CC	Confidence Code <sup>5</sup>
Ref. No.	Reference(s) from which carotenoid values were obtained <sup>6</sup>

#### Footnotes:

<sup>1</sup>The NDB number is a five-digit numerical code used in the USDA Nutrient Database for Standard Reference, the electronic version of Agriculture Handbook No. 8, which can be downloaded from this site. Foods in the USDA-NCC Carotenoid Database which do not have corresponding entries in the USDA Nutrient Database for Standard Reference, are indicated by 'xxxxx' in the NDB column. For more information on these files contact the Nutrient Data Laboratory, 4700 River Road, Unit 89, Riverdale, MD 20737. Tel. 301-734-8491.

<sup>2</sup>To calculate the means, carotenoid values from each reference/study were weighted by their respective sampling plan ratings during the data evaluation process. Sampling plan reflects the representativeness of the sample regarding the brand or cultivar, method of preparation and the geographic origin. A higher rating for a sampling plan indicates the national representativeness

for that particular food. Lower ratings were assigned to data for foods grown or prepared under experimental conditions and which are not consumed by the general population. Therefore mean values derive more weight from those values which were more representative of the food supply. A zero value is a true analytical zero value (<detection limit), while an unreported carotenoid value does not necessarily mean a zero value.

<sup>3</sup>Standard Error of the Means (SEMs) are reported when values for #S are  $\geq 3$ . Where only one weight class (sampling plan rating) exists, the SEM was calculated using the standard equation (SEM = standard deviation divided by the square root of the number of observations in the data set). In data sets with more than one weight class and with two or more observations in each weight class, the SEM of the weighted mean was calculated as the square root of the sum of the square of the weight times the variance of the mean for each weight class over all weight classes (W. G. Cochran, 1977, Sampling Techniques, John Wiley & Sons, Chapter 5). Several data sets had only a single observation in one or more weight classes. In this case the weight class means were replaced with the weighted mean for the data set and the 1 in the n-1 of the denominator was replaced with the sum of weights in that weight class to compute the variance of the mean for the weight class, where the sum of weights across the data set equals 1. With these variances of the means for each weight class, the calculation of the standard error of the weighted mean proceeded as above. Dr. Larry Douglass, Biometrics Program, Animal and Avian Sciences, University of Maryland, conducted simulation studies to evaluate the characteristics of these and a number of other alternative equations for estimating the SE of the weighted mean.

<sup>4</sup>#S is the total number of means/individual values used to calculate the mean and SEM for each carotenoid in the database. In the scientific literature each value can be a mean of many values (depending on the number of samples used in the study) or an individual value. Furthermore there may be more than one value for a single food in one reference. As a result, the total number of references may not equal #S. Since the data have been compiled from various sources, #S does not necessarily equal “n” in statistical terms.

<sup>5</sup>The Confidence Code designated as a, b, or c is a general indicator of the quality of the data (a=best). The procedure for determining Confidence Codes is described in Mangels, et al. (J. Am. Diet. Assoc. 93:284-296, 1993).

<sup>6</sup>Documentation for each reference can be found in the Car\_ref file.

### Zea\_tble - Analytical Zeaxanthin Values for 22 U.S. Foods

Zea\_tble contains zeaxanthin values for 22 U.S. foods. The fields in the Zea\_tble are as follows:

NDB	USDA Nutrient Data Bank number
Desc	Food description
Mean	Mean value for zeaxanthin (mcg/100g edible portion)
#S	Number of means/individual values
Ref. No.	Reference from which zeaxanthin value was obtained (Refer to Car_ref file).

There were very limited data for the zeaxanthin content of foods. For each food in this file, a single value from a reference constituted the 'mean'. As a result, SEM (standard error of the mean) could not be calculated and a Confidence Code was not assigned.

### Car\_ref - Carotenoid References

Car\_ref provides a list of 49 references from which values for U.S. foods were obtained. The reference numbers from the reference file correspond with the Ref. No. column. Published references list authors, title, and journal citation, as well as a brief description of the analytical method used, and the foods and carotenoids analyzed. The sources of unpublished data are also provided.