Programmer's Perspective - Use Of Formats In Other Systems

Lois Steinfeldt, University of Texas, Health Science Center at Houston, Houston, TX

The two major changes in the Survey Nutrient Database System described here today present both an opportunity and a challenge. An opportunity to use the food and nutrient data which will now be available in an easily accessible fashion to improve nutrition related applications and a challenge to use this data accurately and appropriately.

The conversion of the Survey Codebook from a word processing file to a set of normalized data base files; the creation of the measure description file to standardize measures for the gram weights of foods; and the creation of codebooks for the PDS, retention and moisture fat change data sets represents a significant amount of work which greatly enhances the usefulness of the Survey Nutrient Database System. These files can now be easily imported into a variety of software packages providing much greater flexibility in the use of the data. In data base management, statistical or spread sheet software these files can easily be queried and analyzed.

These codebooks are now a resource which can be used not only by programmers, but by anyone with some knowledge and experience in software packages. They should be considered when evaluating and selecting sources of food and nutrient data for specific projects. Listed here are a few examples of what types of information can be obtained from the codebook files.

Table 1
Potential Uses For Survey And Pds Codebooks

- Create subset databases using food codes and/or descriptions
  - Foods whose descriptions contain the word chicken
  - Foods whose codes begin with 57 (cereals)
- Select and calculate gram weights for selected foods
  - Average grams per surface inch of pizza
- Calculate nutrient values for standard measures for selected foods
  - Sodium content for 1 cup of chicken soups
  - Iron content of 3 oz raw weight of meats
  - Vitamin A content for 1 cup of cold cereals

The Nutrient Database System for Trend Analysis presents both a conceptual and a technical challenge. While it solves problems which are inherent in sequential versions of nutrient data bases, the size and complexity of the system are increased. It will require an additional step in order to retrieve data accurately. Dates must be taken into account for each use of the data files. When data is requested for a specific date a single record is returned. If no date or a range of dates is requested, decisions must be made about the meaning and processing of multiple records. For example, the links shown earlier between the files which comprise the survey codebook must be made not only on the data items which link the files such as food code and subcode, but also with reference to a specific date. The date may be the date of the food intake, the current date, or any other date of interest.

Table 2 shows the data for a sample food from the codebook description and gram weight files. The food description has an effective date range of 4/1/1985 to 12/31/2010. However there are two gram weights for measure number 61528 with different effective dates. The first gram weight of 46.8 is effective from 04/01/1985 to 12/31/1992. The second gram weight of 35.0 is effective from 01/01/1993 to 12/31/2010. This change in the gram weight of 1 package is a food change. It could result from changes in packaging or from changes in the food itself.

Table 2

SURVEY CODEBOOK

FOOD CHANGES IN GRAM WEIGHTS

CODEBOOK DESCRIPTION FILE

Survey Code	Description	Starting Date	Ending Date
1010101	Sample food description	04/01/1985	12/31/2010

CODEBOOK GRAM WEIGHT FILE

Survey Code	Sequence Number	Measure Number	Gram Weight	Starting Date	Ending Date
1010101	1	61528	46.8	04/01/1985	12/31/1992
1010101	1	61528	35.0	01/01/1993	12/31/2010

In order to retrieve the correct gram weight, the date for which the data is applicable must be part of the selection of the data records. For example, in Table 3 a simple selection command is shown to retrieve the codebook description and gram weight data for the sample food. First using the codebook description file, select the record with food code = 1010101 and starting date less than or equal to the current date and an ending date greater than or equal to the current date. Second using the gram weight file select the record using the same criteria.

Table 3

SURVEY CODEBOOK
DATA SELECTION PROCESS

Use codebook description file

Select the record with food code = 1010101 and
starting date <= current date and
ending date >= current date

Use codebook gram weight file

Select the record with food code = 1010101 and
starting date <= current date and
ending date >= current date

If the current date were taken from the computer, different records would be retrieved on different days. If this retrieval were done today, the gram weight would be 35.0. However, if the retrieval were done 1 year ago, the gram weight would have been 46.8. In each of these cases only 1 gram weight record would meet the selection criteria and that record would contain the correct gram weight for the current date. This same concept and the same general method of retrieval applies to all the codebook files. Each codebook file, including the description file, may have multiple records with different effective dates.

Table 4

SURVEY CODEBOOK
DATA SELECTION RESULTS

For current date = 05/24/1993

1010101 Sample Food

61528 1 package 35.0

For current date = 05/24/1992

1010101 Sample Food

61528 1 package 46.8

The file which is most likely to have the most food changes is the Survey Nutrient File. Since the nutrient values for a survey food are represented by a recipe, the nutrient values will include the combination of food changes to each component of the recipe.

As shown in Table 5 this includes food changes to ingredients, the gram weight and nutrient values for PDS and Survey foods which are used as ingredients, retention factors, moisture and fat changes and type of fat.

Table 5

POTENTIAL SOURCES OF FOOD CHANGES TO SURVEY NUTRIENT VALUES

Recipe components:

- Ingredients
  - Adding and deleting ingredients
  - Changing ingredients and ingredient amounts
  - PDS foods - gram weight and nutrient values
  - Survey foods - gram weight and nutrient values
- Retention Factors
- Moisture and fat changes
- Type of fat

Food changes to different recipe components could occur at any time and in many different combinations. Some food changes, such as a vitamin fortification in a cereal, may change only the value for a single nutrient, while others, such as a change in an ingredient in a recipe, may change all the nutrient values.

Table 6 shows an example of a breakfast cereal which has food changes occurring on three different dates affecting five nutrients. Food changes result from real differences in the food while data changes result from improvements in food composition data. In this food, the changes are due to fortification and reformulation. The Vitamin A value, nutrient code 392, decreases on 04/01/1989 from 1324 to 794. The Vitamin C value, nutrient code 401, increases from 53.0 to 211.6 on 04/01/1987. The values for saturated fat, monounsaturated fat and polyunsaturated fat, nutrient codes 606, 645 and 646, change on 10/01/1989. The rest of the nutrient values remain the same.

Table 6
SURVEY NUTRIENT FILE
MULTIPLE FOOD CHANGES TO NUTRIENTS

Survey Code	Nutrient Code	Nutrient Amount	Starting Date	Ending Date
5721300	392	1324.0	04/01/1985	03/31/1989
5721300	392	794.0	04/01/1989	12/31/2010
5721300	401	53.0	04/01/1985	03/31/1987
5721300	401	211.6	04/01/1987	12/31/2010
5721300	606	0.81	04/01/1985	09/30/1989
5721300	606	0.82	10/01/1989	12/31/2010
5721300	645	0.37	04/01/1985	09/30/1989
5721300	645	0.42	10/01/1989	12/31/2010
5721300	646	0.52	04/01/1985	09/30/1989
5721300	646	0.58	10/01/1989	12/31/2010

Table 7 shows a simple selection command using the date of intake as the criteria used to select records. Using the Survey Nutrient File, records are selected for food code equal to 5721300; nutrient codes equal to 392 (Vitamin A), 401 (Vitamin C), 606 (Saturated Fat), 645 (Monounsaturated Fat) and 646 (Polyunsaturated Fat); and starting date less than or equal to the date of intake and ending date greater than or equal to the date of intake.

Table 7

SURVEY NUTRIENT FILE
DATA SELECTION PROCESS

Use survey nutrient file

Select the records with food code = 5721300 and

nutrient code = 392 or 401 or 606 or 645 or 646 and

starting date <= intake date and ending date >= intake date

Substituting in different dates for the intake date will retrieve different nutrient values as shown in Table 8. In this example, none of the nutrients changed value more than once. However, that can and certainly will happen.

Table 8

SURVEY NUTRIENT FILE
DATA SELECTION RESULTS

Date of Intake	VIT A 392	VIT C 401	SAFA 606	MUFA 645	PUFA 646
03/15/1987	1324.0	53.0	0.81	0.37	0.52
03/15/1989	1324.0	211.6	0.81	0.37	0.52
06/15/1989	794.0	211.6	0.81	0.37	0.52
10/15/1989	794.0	211.6	0.82	0.42	0.58

These examples from the survey codebook and nutrient files clearly demonstrate the critical role that dates now play in retrieving data from the Survey Nutrient Database System for Trend Analysis. Dates must now be used to select both food and nutrient data. When there are multiple records for a data item and no selection based on dates is done, which record is selected or how many records are selected may depend on the software used, how the retrieval is programmed, the indexing of the data file, and the physical order in which the records are stored. If dates are not taken into account, an incorrect record may be retrieved and the wrong value used.

However, there may be times when all the records are required. When this is done, provision must be made for the retrieval of multiple values and decisions made as to how the multiple values are processed.

Version data bases can be extracted for any date. However the date used must always be a single date to insure that multiple values are not retrieved. For example, there are many ways a version data base could be set up for the calendar year 1993. The data could be selected based on the beginning, ending or middle of the year. Data could be selected based on the values in effect for the longest time period during the year or other more complex algorithms using weighted times, etc.

These are just a few examples of how the improvements in the Survey Nutrient Database can be used in other systems. The conversion of the codebooks from word processing to database files is a major step which now provides more data in a format which is much easier to use. Adding the fields and data to monitor food and data changes over time represents a substantial amount of work which will produce a more comprehensive database which can be used in either a continuous or version format. These new formats also provide the flexibility needed to adapt to changing needs for food and nutrient data in the future.