Using Blaise in a Nationwide Food Consumption Survey

Lois Steinfeldt, Ellen Anderson, Jaswinder Anand, Nancy Raper

Introduction

The Food Surveys Research Group (FSRG) at the U.S. Department of Agriculture, working with Westat, Inc., is in the process of converting the nationwide food consumption survey from a paper and pencil questionnaire to a computer-assisted interview programmed in Blaise 4. Collecting and coding information about the foods consumed by individuals in large-scale food consumption surveys requires asking specific questions for each food. There are numerous factors affecting the nutrient content of foods that must be captured. The questions required across all foods can easily reach into the thousands and the number of responses to an individual question can reach into the hundreds. The large number of questions and responses has presented many challenges in developing and managing questionnaire specifications and testing the instruments.

Background

The U.S. Department of Agriculture (USDA) has conducted surveys to collect national information on food consumption since the 1930's. These surveys monitor food use and food consumption patterns in the U.S. and provide data used to address economic, nutrition and food safety issues. For example, the data are used to evaluate the nutritional adequacy of the American diet and the impact of food assistance programs. The data are also used to estimate exposure to pesticide residues and to study the impact of food fortification, enrichment, and food labeling policies.

The Continuing Survey of Food Intakes by Individuals (CSFII) is a nationally representative sample of individuals of all ages. The screener questionnaire is used to identify eligible households. A household questionnaire is administered to the household member most knowledgeable about household characteristics such as income, education and food shopping practices. Selected individuals are asked to provide food intakes for 2 days, spaced 3-10 days apart. One respondent over age 19 within a household, who provided at least 1 day of food intake, is selected to complete the Diet and Health Knowledge Survey (DHKS).

During the intake interview, individuals recall the foods and beverages that were consumed the day before the interview. Details about each food and beverage are collected as well as an estimate of the amount consumed. Information is also collected on the time of day the food was eaten, the name of the eating occasion, whether the food was eaten at home or away from home, and where the food was obtained. Each food reported is then linked to nutritive values in order to calculate how much of each nutrient the individual consumed. Fifty-two different nutrients are calculated including calories, fat, protein, carbohydrate, vitamins, and minerals. Plans are underway to integrate the CSFII with the National Health and Nutrition Examination Survey (NHANES) conducted by the National Center for Health Statistics (NCHS), U.S. Department of Health and Human Services (DHHS). Both surveys will use the USDA dietary collection method and nutrient database. Each of the surveys will produce a core set of variables for estimating and interpreting dietary intakes in a combined yearly sample.

Description of the Instruments

Table 1 shows the number of questions, enumerated responses, lookup files and responses in lookup files for each of the instruments after the conversion from paper and pencil to the computer-assisted interview version. The table clearly shows that the food intake instrument is the largest of the instruments in terms of the numbers of questions and responses. Most of the food intake instrument consists of questions about specific food details, including the amount of food eaten. The large number of questions and responses produce an even larger number of skip patterns because the questions asked about a food depends on the responses to the previous questions. The number of possible paths through the food detail questions is roughly estimated to be over four hundred thousand. The size and complexity of the food intake instrument is what makes this application unusual and challenging.

Table 1
Pilot Study I CATI

Instrument	Questions	Enumerated Responses	Lookup Files	Responses in Lookup Files
Screener	52	74		
Household	112	248		
Food Detail Section	2389	11932	93	9352
Total Food Intake	2515	12414	96	9959
Diet and Health Knowledge Survey	137	429		

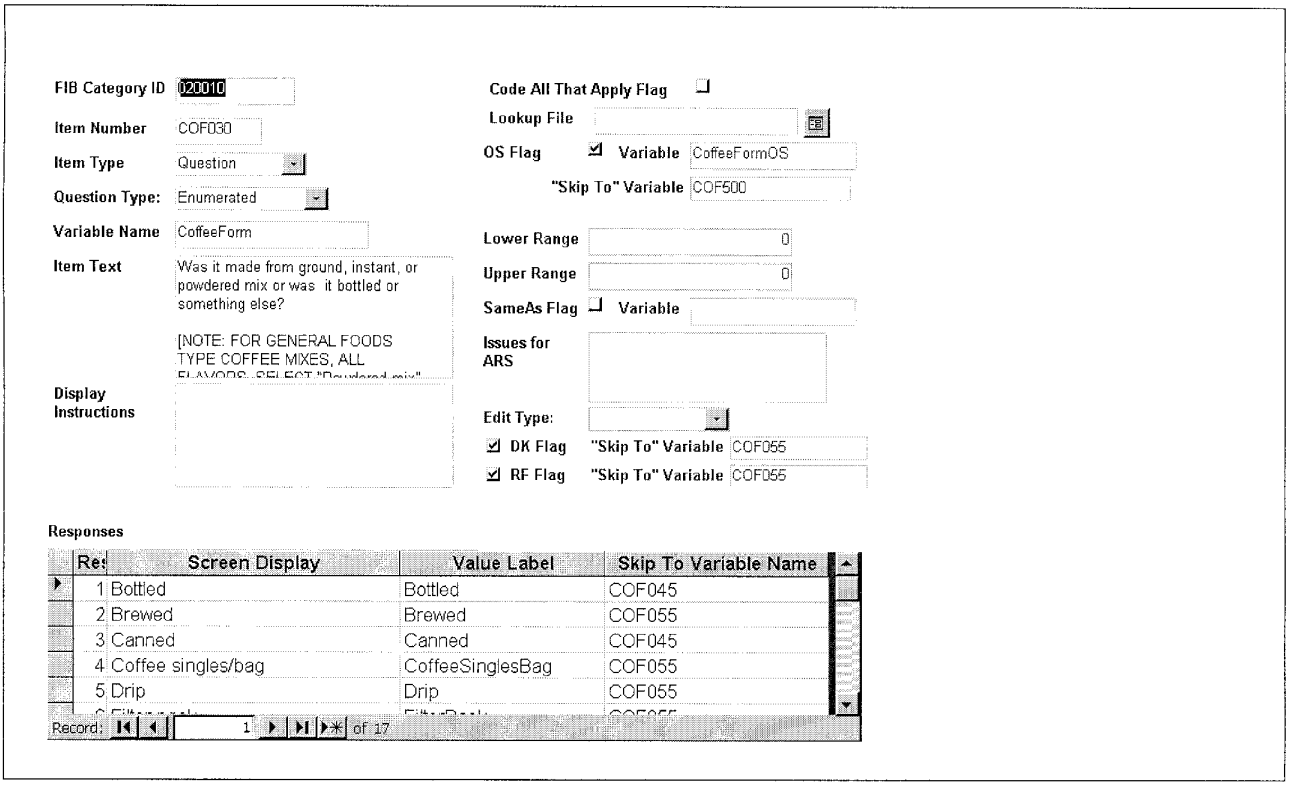
Food Detail Specifications

Creating the food detail specifications began with the Food Instruction Booklet (FIB). The FIB was an 80-page booklet used as an interviewer aid in earlier food intake surveys. It was designed to assist interviewers in collecting detailed descriptions of foods and amounts for the paper and pencil version of the food intake questionnaire. Foods had been grouped into 16 broad categories. Since the questions and responses within these categories were very different, a decision was made to further divide them into 132 categories for the computerized version. Each of the food categories was assigned a unique code. The categories made it possible to ask more specific questions and made writing, programming, and testing the specifications a more manageable task.

Early in the planning stages, the decision was made to use a database approach to writing the specifications. From years of experience maintaining a large and diverse food coding database in a constantly changing food market, it was clear that the food intake specifications would be large and complex and would need to be updated periodically. This approach produced many advantages in the writing, reviewing, and editing of the specifications. When changes were made that affected more than one category, it was easy to review and apply the changes uniformly.

Figure 1 shows the form used to enter and edit the specifications. The specification database has two main tables: Items and Responses. The Item Number uniquely identifies each item and links the Items and Responses tables. Included in the Items table are questions, boxes, and edits. Question types include enumerated, open ended, lookup file, and continuous. The responses for enumerated questions are stored in the Responses table. Enumerated questions are used when the responses fit on one screen. Lookup files are used when the number of responses would exceed one screen (approximately 40). Some of the largest lookup files include candy, cake, cereals, and frozen meals, all of which have more than 300 responses. Questions with continuous responses are used for the amounts of food eaten. Open-ended questions are used for “Other, specify” responses.

Figure 1
Specifications database data entry and edit form



When there is a lookup file for a question, the name of the file is stored in the Items table. Lookup files allow the interviewer to search for a response using the trigram search rather than having to page through multiple screens. Lookup files can be used with more than one question. Table 2 shows examples from the "CheeseKind" lookup file. This lookup file provides response options for the "What kind of cheese was it?" question for the Cheese and Grilled cheese sandwich categories. Each category that uses the question has a column in the lookup file for the Skip to Variable Name field. In this example, when cheddar cheese is chosen in the cheese category, the instrument skips to item CHE010. When cheddar cheese is chosen in the Grilled cheese sandwich category, the instrument skips to item GCS440.

Table 2
CheeseKind Lookup File

Food Name	Cheese Category Skip to Variable Name	Grilled Cheese Sandwich Category Skip to Variable Name
Cheddar cheese	CHE010	GCS440
Cheese spread	CHE017	GCS335
Provolone cheese	CHE020	GCS345

Most questions allow only one response. Questions that allow more than one response are flagged as Code All That Apply. There are 188 Code All That Apply questions in the food detail instrument. DK flag, RF flag, and OS flag fields are checked when "Don't know", "Refused to answer" and "Other, specify" are allowed responses. DK Skip to Variable, RF Skip to Variable, and OS Skip to Variable fields contain the Item Number that identifies the question to skip to for those responses. "Don't know" and "Refused to answer" responses are allowed for every food detail question. "Other, specify" is also allowed for every question with a few exceptions such as questions with only "yes" and "no" as response options. Edit ranges are stored in the Lower Range and Upper Range fields, and Edit Type specifies whether the edit is hard or soft.

The Responses table contains the Response Number, Screen Display, Value Label, and Skip to Variable Name for each of the responses for the enumerated questions. The Skip to Variable Name indicates the question to skip to for that response. Skip patterns were provided for every one of the approximately 21,000 responses in the Responses table and lookup files.

Edit items use the Item Text box to explain complex edits. An example of a complex edit is checking for impossible combinations of answers between two or more questions. Edits are also used for impossible combinations of responses within a Code All That Apply question.

Boxes are used to explain complex skip patterns that can not be defined using the Skip to Variable Name fields in the Responses table and lookup files. The skip instructions are written in the Item Text box. Boxes are also used for instructions on lookup files used by more than one question. As shown in Table 2, these lookup files have a Skip to Variable Name column for each question that uses the file. The box is used to specify which column in the lookup file applies to which question.

The database approach saved time in writing and reviewing specifications. This method made it possible to easily copy questions, edits, boxes, and responses within a category as well as across different categories. Templates were developed for questions that were asked across multiple categories. Text areas that display instructions for the programmer and issues to resolve provided flexibility to specify unusual conditions. A variety of reports were created and used during the review process.

Main Food List

The first step in the food intake interview is to obtain a list of foods consumed. In order to produce the correct set of questions for each food, there must be a link between the reported food and the food category. This link is provided by the Main Food List (MFL), which is a Blaise lookup file containing approximately 2500 food names. Each food name is linked to a food category number that determines which questions are asked for that food. The food names on the MFL must reflect current food supply and food consumption patterns. General descriptions such as lunch, buffet, or unknown food are also included and linked to a special unknown category to collect further detail later in the interview. The trigram search mechanism is used to locate foods on the MFL as the respondent reports them. A complete, easy-to-search MFL that can be updated in a timely manner is extremely important to the success of the survey.

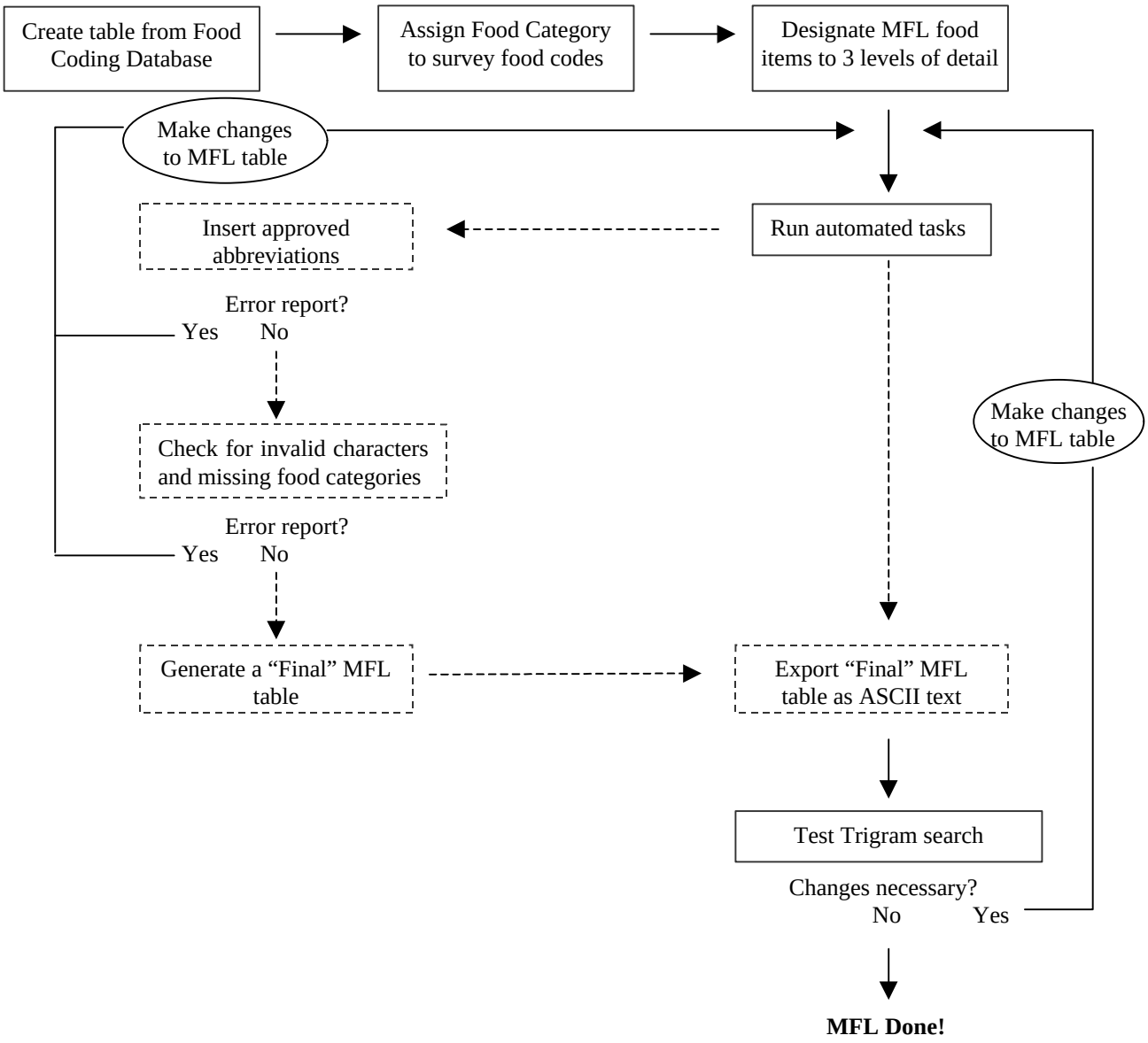
The process of creating the MFL is shown in Figure 2. It was constructed primarily from the CSFII Food Coding Database. The Food Coding Database was a logical starting point because:

1) it is a large and diverse collection of foods and brand name products;
2) frequency data from previous surveys identified the most commonly consumed foods and ensured their inclusion on the MFL;
3) the relational database format of the Food Coding Database files made extracting foods for the MFL very easy and enhanced our ability to manage the progress of a constantly-evolving MFL.

To begin building the MFL, a database table was created that contained all 7300+ survey food codes, their generic food descriptions, and brand name information extracted from the Food Coding Database. Additional fields for Food category, MFL name1, MFL name2, and MFL name3 were added to the table. Each survey food code was assigned one of the 132 food categories by a food specialist. By linking the survey food codes to food categories, it was anticipated that:

1) this information could be utilized for automated coding purposes;
2) maintenance of the MFL would be enhanced by tracking changes and additions to the Food Coding database, which in turn, may affect the MFL and the food category questions.

Figure 2
Creating the Main Food List from the Food Coding Database



As shown in Table 3, three separate fields for MFL name were used to record different levels of detail for MFL food items. For example, "oatmeal" was assigned to MFL name1 representing the least amount of detail, whereas the more specific "instant oatmeal" and "Quaker instant oatmeal" would be found under MFL name2 and MFL name3, respectively. Having these three fields made it possible to generate three different MFLs based on the level of detail desired. A team of food specialists reviewed the survey food codes category by category, assigned food items to a level of detail, and decided what food items would be included on the MFL. In some cases, MFL food items were added to the table but were too general to link to a specific survey food.

Table 3
Food Coding Database table used for constructing the Main Food List

Food Description	Survey Code#	Food Category	MFL name1	MFL name2	MFL name3
Oatmeal, cooked, NS as to regular, quick or instant	56203000	50020	oatmeal		
Oatmeal, cooked, instant	56203080	50020		instant oatmeal	
Oatmeal, cooked, quick	56202970	50020		quick oatmeal	
Oatmeal, cooked, regular	56202980	50020		regular oatmeal	
Quaker Fruit and Cream, all flavor varieties	56203080	50020			Quaker instant oatmeal

Automated tasks included:

1) A list of approved abbreviations (e.g., “bf” for baby food; “lf” for low fat) replaced longer words in some MFL food items;
2) MFL food items were checked for the presence of invalid characters which could hamper the trigram search (e.g., extra blank spaces between words; dashes “-“ and slashes “/”);
3) Every MFL food item was verified to have a food category assigned to it;
4) MFL food items and their corresponding food category were extracted from the original table and copied into a final MFL table;
5) The final MFL table, shown in part in Table 4, was exported as an ASCII text file to be prepared for use in Blaise.

Table 4
Final Main Food List format

MFL Name	Food Category
American cheese	60010
Blue cheese	60010
Brie cheese	60010
Cheddar cheese	60010
Cheese	60010
Cheese ball	60010
Cheese spread	60010
Cheese sticks	60010
Cheez Whiz	60010
Feta cheese	60010

During the development of the MFL, a Blaise program was used to test how well the MFL would work during the interview. Food specialists entered food names and evaluated the results of the trigram search of the MFL. The goal was to have common food names show up at the top or close to the top of the list returned by the trigram search.

Since the MFL information is in a database table, changes can be made with ease, and the programs can be re-run as many times as necessary to generate the final MFL table. This entire process was executed 9 times over a time period of about 6 months during development.

Limitations of Using CAI in Food Surveys

Although the use of database management software was a big help in the specification process, it could not solve all the difficulties in converting the food intake questionnaire from PAPI to CAI. One of the issues is how to balance the need to collect complete and accurate intakes with the size of the data model. When thousands of people, every year, are asked what they eat, their answers will be very different and some will be quite unusual. This occurs in both the numbers and the types of foods and beverages consumed. Although national averages of nutrient intake are not affected by the data for a few foods, for some uses of the data, it is the outliers that are important. An example of this is estimating pesticide exposure from dietary intakes. For this purpose, the tenth fruit or vegetable eaten in a day may be important. Setting array sizes to accommodate the highest intakes of foods per day and foods per category per day does not seem to be possible without compromising the instrument. The ability to dynamically change array sizes would be very useful in this application.

The fixed array sizes also produce a very large, although very sparsely populated data set when the data are extracted from Blaise. This very large set of files (> 900) requires a complex post-interview processing system to extract and condense the food intake data reported in the interview. We are looking forward to seeing how the implementation of Open Blaise Architecture (OBA) can improve the post-interview processing on the food intake data.

Another major concern in using CAI in food surveys is the need to continually update responses and even questions as the food supply, eating patterns, and dietary concerns change. Since the responses are the basis of skip patterns and edits, it appears that the kind of continual update done on Food Coding databases will be difficult to implement on a CAI instrument. Although some changes can wait until for the end of each survey year, some will need to be made while the survey is in the field. We are interested in finding designs or processes which would be able to produce changes in a cost-effective and timely manner.

Summary

The features of Blaise and the extensive use of database management applications made it possible to create a very detailed food intake questionnaire. The database management system made it easy to search for specific questions across all categories, compare questions and responses between categories, and copy specifications from one category to another. The database format also made it easy to find and make changes to the specifications. Utilization of Blaise's trigram search for the Main Food List linking food names to food categories provided a quick way to select the appropriate set of food-specific questions. Creating the Main Food List from a database table proved to be useful for organizing and maintaining different versions of this lookup file and for automating processing tasks. The extensive use of lookup files for long lists of possible responses made it possible to ask more specific questions and will also be easier to maintain as foods change over time.