

Perspectives on Data Quality - Panel Discussion

Lori Borrud, USDA-ARS-HNIS

I'm here representing the Nutrition Monitoring Division of USDA's Human Nutrition Information Service, now a part of the Agricultural Research Service, to present our perspectives on data quality.

The Nutrition Monitoring Division monitors the food intake of the general U.S. population and special high-risk subgroups by planning, conducting, and monitoring USDA's nationwide food surveys; compiles and maintains data on the composition of foods in order to identify the adequacy and quality of U.S. diets; and disseminates data and information through a variety of media. In each of these steps, we have implemented and maintained quality-control procedures to ensure the goal of releasing quality food and nutrient intake data to the public. We believe that data-quality efforts are continuous activities that must be integrated into all survey and post-survey data-processing operations. While these efforts encompass the use of food composition data, that is only one area of data quality we must address.

I would like to discuss our data quality perspectives in relationship to our Continuing Survey of Food Intakes by Individuals, or CSFII, and Diet and Health Knowledge Survey, or DHKS. USDA initiated the CSFII in 1985 to provide continuous dietary intake information on the U.S. population. The current, 1994-96, CSFII is the third in the series of individual intake surveys. The DHKS was initiated in 1989 as a telephone follow-up to the CSFII and is the first national survey designed so that data on individuals' attitudes and knowledge about nutrition can be linked with their food and nutrient intakes.

HNIS is conducting CSFII and DHKS 1994-96 as "What We Eat in America Survey." The survey contract was awarded in September 1992 to Westat, Inc. of Rockville, Maryland. Data collection for the "What We Eat in America" survey began in January of this year and will continue through 1996.

For CSFII and DHKS 1994-96 HNIS has implemented strong management and quality control procedures, both as part of the contract and in its in-house management of the surveys. A team of food survey specialists with experience in survey design, dietary methodology, food coding, data management, and statistics has been assigned specific responsibility for monitoring every task in the contract. The current CSFII and DHKS also include a number of survey design changes and improvements in survey monitoring and data management to speed the release of quality data.

Acceptable response rates are a first step in providing quality data to users. Procedures to achieve and maintain acceptable response rates have been established as part of the CSFII/DHKS 1994-96 contract. For example, the contract has requirements for specified response rates that must be met by the contractor.

In addition, a number of changes in the sampling and data collection methods have been made to reduce respondent burden, which in turn affects response rates.

To manage nationwide collection of data by close to 100 interviewers, the contractor uses an automated field management system that tracks the status of every questionnaire and the activity of every interviewer in the field. In addition, the system allows the supervisor to provide updated assignments and timely feedback to the interviewer.

The contract also specifies an automated forms tracking system to track the information collected in each interview from receipt of the data by the contractor to delivery of the processed data to HNIS. This tracking system is updated daily, allowing HNIS to track the status of any questionnaire. We believe these kinds of automated systems are needed to ensure that the response rates are being met, the sampling design is being followed, and the data are collected and processed according to the survey schedule.

For CSFII 1994-96, HNIS is using Survey Net, an automated food coding and nutrient analysis system designed specifically for use with the CSFII. We believe Survey Net will improve the efficiency of our technical support systems, leading to improved quality and timeliness of results. Survey Net was planned and developed jointly by HNIS and the University of Texas.

Survey Net was designed as a multilevel system for use by both the contractor and HNIS. Survey Net contains a number of user-friendly features that the contractor can use in coding food intake records received from the field. For example, the Survey Net RECIPE option allows the user to modify existing foods and recipes in the Survey Nutrient Data Base for foods reported in the survey. If modifications to the original foods are made, these can be saved to a recipe file. The on-line editing feature of the program permits the user to add, change, or delete foods. The system also contains built-in edit checks and quality control features such as weight or quantity extremes and range of response checks.

In addition, the coder is able to record unknowns. Or put another way, the coder is able to flag any foods or food quantities that cannot be coded for later attention at HNIS. This information is then transmitted electronically to HNIS on a weekly basis together with the coded food data from Survey Net.

On the HNIS level, Survey Net features allow us to monitor the data received from the contractor in a timely fashion, resolve unknowns, update data base files accordingly when new foods are reported, and approve recipe modifications. We can then send updated data base files electronically back to the contractor. In addition, Survey Net's built-in quality-control features provide a continuous review of recipes and food weights.

As part of our quality-control procedures, we have specified maximum error rates that the contractor must not exceed in coding the food intake data. In addition, all coders must pass certification based on test sets of intakes developed by HNIS before they are permitted to code.

In another change instituted for CSFII 1994-96, HNIS receives all survey data by weekly electronic transmissions from the contractor rather than through quarterly or yearly magnetic data tapes. Along with these weekly data transmissions, we receive reports of coding error rates, the status of interviews in the field and of data in-house at the contractor, and other information. All of this information enables us early on to identify and correct any problems in the field or during data processing and will permit the release of data more quickly than in previous surveys.

Data quality activities do not stop with the transmission of the food consumption data to HNIS. We have instituted an automated system to track survey data in-house, since we want to have a firm handle on how long each data processing step takes in order to identify in-house backlogs early. Once received at HNIS, all data are put through a rigorous automated in-house editing process and then review by the various specialists within the Division.

This comprehensive in-house review includes a review of the food intake data by our Survey Systems Team using Survey Net to ensure that recipe modifications were correctly made, codes and weights were correctly assigned, unknowns were resolved, and so on. Corrections are made and feedback is provided to the contractor weekly. Once the editing and review process is completed, the food consumption data are converted to nutrient quantities through the use of the analysis program contained in the Survey Net. Nutrient values per 100 grams of the foods for conversion of foods to nutrients are provided through the Survey Nutrient Data Base.

HNIS conducted a pilot study of all survey operations in the spring of 1993 and was pleased with the performance of Survey Net.

I would like to mention just some of our research activities that relate to data quality. HNIS has an on-going interagency agreement with the Center for Survey Methods Research, Bureau of the Census, to improve the reporting of foods through cognitive research. For example, CSMR recommended the use of a quick list and multiple passes through the day of intake. Intake questionnaires were subsequently revised based on this research.

An issue raised in recent years is the extent of underreporting that may exist in food consumption surveys. Mertz and colleagues have estimated that caloric intakes may be underreported by 18 percent based on their analysis of food records (Mertz et al. 1991). We recognize that underreporting may exist; however, at this time we do not know what foods are underreported or why underreporting occurs--whether it results from the data collection method, the method used for estimation of portion sizes, sociodemographic characteristics of the respondents, or some other reason. HNIS is addressing this issue through its collaborative work with the Center for Survey Methods Research and other planned research projects.

In an earlier paper, Patricia Guenther described our on-going collaborative work with Iowa State University to develop statistical methods that will use food intake data to estimate usual nutrient intake distributions. Iowa State University has now begun the much more difficult task of developing methods for estimating usual intakes of foods.

The Food Instruction Booklet, or FIB, which is used by the interviewers to probe for needed detail on the foods and amounts reported by respondents, has been revised for 1994-96. HNIS staff worked in-house and with our contractor to develop standardized probes written in the FIB to ensure that all interviewers collect the same level of detail in the same way for consistent coding of foods.

HNIS has worked to improve and strengthen every facet of the continuing survey in-house and contractually. But quality data also depend on the food composition data base used to convert reported foods and quantities into nutrients. Betty Perloff, in her paper, described the problems regarding food composition data in trying to compensate for limited resources. We must make sure that funding to implement the GAO recommendations will not draw funds away from needed research to support USDA's food consumption surveys.

For example, implementing GAO's recommendation to analyze each food at a cost of \$2,000 per sample with six samples per food, or \$12,000, is very much a budget issue. Over 1,000 new food codes were added to the Survey Nutrient Data Base during CSFII 1989-91 and we anticipate that as many or more will be added during 1994-96. Analytical costs for new food codes could be very high.

We all recognize that food composition data will continue to improve over time. More data continually become available, better analytic methods are developed, and so on. When many current analytic methods need improvement, questions may arise such as why scarce resources should be spent at this time to increase sample sizes. An option may be to explore innovative ways to make the best use of available data. As one means of doing this, HNIS is developing the Survey Nutrient Data Base for Trend Analysis.

The trends data base will allow previous intakes to be adjusted to account for known improvements in food composition data and will also allow measure of changes in nutrient intake levels resulting from various types of real changes in foods. This data base should be available in about a year.

I would like to briefly mention the issue of brand names. We believe that this issue requires careful consideration before we move ahead quickly to add brand-specific nutrient data to the data base. For example, how often are respondents able to provide accurate brand name information for food items? When is brand name information needed and for what purpose? Do we know which brand name food items may be similar, if not identical, in nutrient content? The addition of brand name information to the nutrient data base will require considerable effort and expense. Are we sure that the addition of this information will result in improved food composition data or, more precisely, in improved estimates of intakes?

To close, I want to reemphasize that while HNIS is actively addressing GAO's concerns, decisions to implement GAO's recommendations will have budget implications, since other program needs must also be met. However, HNIS believes that moving ahead with programs such as the nutrient data base for trends analysis will improve the use of current data for research purposes. Our pursuit of sophisticated statistical methods for combining data from different sources in order to consider quality factors in the redesigned nutrient data bank may be the direction we have to take, not only statistically, but in consideration of the budget constraints we must live with.

HNIS believes that quality nutrient intake data depend not only on a quality food composition data base but also on a quality survey as well. We are moving ahead, taking into consideration the total picture.

Reference

Metz, W., J.C. Tsui, J.T. Judd, S. Reiser, J. Hallfrisch, E.R. Morris, P.D. Steele, and E. Lashley. 1991. What are people really eating? The relation between energy intake derived from estimated diet records and intake determined to maintain body weight. *Amer. J. Clin. Nutr.* 54:291-5.