JOURNAL OF FOOD COMPOSITION AND ANALYSIS (2000) **13,** 379–389
doi:10.1006/jfca.1999.0867
Available online at http://www.idealibrary.com on IDEAL®

ORIGINAL RESEARCH

USDA's National Food and Nutrient Analysis Program[1]: Food Sampling

P. R. Pehrsson[2], D. B. Haytowitz, J. M. Holden, C. R. Perry, and D. G. Beckler

Nutrient Data Laboratories, BHNRC, ARS, USDA, Bldg. 005, Rm. 309A, BARC West, 10300 Baltimore Blvd., Mail Stop 47, Beltsville, MD 20705-2350, U.S.A.

Received June 25, and in revised form December 22, 1999

The National Food and Nutrient Analysis Program (NFNAP) is designed to develop robust estimates of the mean nutrient content of important foods in the food supply and significantly improve the quality of food composition data in the US Department of Agriculture's (USDA) National Nutrient Databank. The program objectives are: (1) evaluation of existing data; (2) identification of Key Foods and nutrients for analysis; (3) development of nationally based sampling plans; (4) analysis of samples; and (5) compilation and calculation of representative food composition data. This paper describes our efforts in development of the sampling plan (objective 3) and presents limited preliminary results. The sampling plan was based on a self-weighting stratified design. First, the U.S. was divided into four regions, then each region was further divided into three implicit strata from which generalized Consolidated Metropolitan Statistical Areas (gCMSAs) were selected. Rural and urban locations were selected within gCMSAs. Commercial supermarket lists were used to select 24 outlets for food pickups; specific brands were selected based on current market share data (pounds consumed). This population-based approach can be applied in the development of other sampling programs for specific ethnic and regional foods. Sampling plans have been developed for margarine, folate-fortified foods (e.g. flours, bread, and pasta), and a number of highly consumed mixed dishes (e.g. pizza and lasagna).

INTRODUCTION

The National Food and Nutrient Analysis Program (NFNAP), an Interagency Agreement between the National Institutes of Health and the US Department of Agriculture (USDA), was initiated in 1997 and has become the most important means of accomplishing a comprehensive update to the National Nutrient Databank. Through this effort, an integrated system for identifying foods and nutrients, food sampling, food preparation and compositing, sample preparation, and chemical analysis has been put into place.

The goals of NFNAP are to provide representative nutrient estimates for foods and selected ingredients, increase data acquisition for important foods, add data for selected new components to the database, and validate factors and algorithms for compilation of nutrient data. The primary objective of NFNAP is to provide the best

[1] In addition to USDA support of staff scientists, the National Institutes of Health and Indian Health Service have provided financial support for this program.

[2] To whom correspondence and reprint requests should be addressed. Tel.: (1-301)504 0716; Fax: (1-301)504 0713. E-mail: Ppehrsson@rbhnrc.usda.gov

0889–1575/00/040379 + 11 $35.00/0

estimate of the nutrient profile—or nutrient means for the population of each food—of foods identified as important in the food supply. Through this process, we continuously redefine the foods in our database to accurately represent what is currently being consumed by the population. For some important contributors of selected nutrients, preliminary estimates of variability will be generated through NFNAP.

To this end and as part of the five aims of the NFNAP, a probability-based (i.e., probability proportional to size or probability proportional to size sampling (PPS)) national sampling plan (Cochran, 1977) was developed to sample and analyse those Key Foods for which existing data were either absent or of poor quality, or where foods or methods have changed (Perry and Beckler, 1999). Through NFNAP, we have identified approximately 1000 foods (660 Key Foods, ethnic foods, mixed dishes, foods used in metabolic studies, and food ingredients) that are important contributors of critical nutrients in the US food supply or to the diet of specific subpopulations (Pehrsson and Haytowitz, 1995; Haytowitz et al., 1996).

Critical nutrients identified for nutrition monitoring include energy, carbohydrate components, sugars, fiber and starch, fat, specific minerals such as sodium, and specific vitamins such as folate (FASEB, 1995). Most of these nutrients appear on the nutrient label, as required by law, on foods distributed in the U.S. A comprehensive list of nutrients for analysis under NFNAP are presented in Table 1. Selecting one of these nutrients—fat—and using the Key Foods approach, we identified nine foods which are primary contributors to fat intake in the U.S. diet (Table 2). The full scope of

TABLE 1

NFNAP nutrients for analysis as appropriate for the food (critical nutrients in bold)

Nutrient description	**Units**	**INFOODS Tagname**
Protein	**g**	**PROCNT**
Total lipid (fat)	**g**	**FAT**
Carbohydrate, by difference	g	CHOCDF
Ash	g	ASH
Energy	**kcal**	**ENERC_KCAL**
Alcohol	g	ALC
Water	g	WATER
Caffeine	mg	CAFFN
Theobromine	mg	THEBRN
Energy	**kj**	**ENERC_KJ**
Sugars, total	**g**	**SUGAR**
Starch	**g**	
Fiber, total dietary	**g**	**FIBTG**
Calcium, Ca	**mg**	**CA**
Iron, Fe	**mg**	**FE**
Magnesium, Mg	**mg**	**MG**
Phosphorus, P	**mg**	**P**
Potassium, K	**mg**	**K**
Sodium, Na	**mg**	**NA**
Zinc, Zn	**mg**	**ZN**
Copper, Cu	**mg**	**CU**
Manganese, Mn	mg	MN
Selenium, Se	mcg	SE
Vitamin A, IU	**IU**	**VITA_IU**
α-carotene	mcg	CARTA
β-carotene	mcg	CARTB
β-cryptoxanthin	mcg	CRYPX
Lycopene	mcg	LYCPN
Lutein	mcg	LUTN

TABLE 1 (Continued)

Nutrient description	Units	INFOODS Tagname
Vitamin E	**mg_ATE**	**VITE**
Vitamin C, ascorbic acid	**mg**	**VITC**
Thiamin	mg	THIA
Riboflavin	mg	RIFB
Niacin	mg	NIA
Pantothenic acid	mg	PANTAC
Vitamin K		
Vitamin B-6	mg	VITB6A
Folate	**mcg**	**FOL**
Vitamin B-12	mcg	VITB12
Tryptophan	g	TRP_G
Threonine	g	THR_G
Isoleucine	g	ILE_G
Leucine	g	LEU_G
Lysine	g	LYS_G
Methionine	g	MET_G
Cystine	g	CYS_G
Phenylalanine	g	PHE_G
Tyrosine	g	TYR_G
Valine	g	VAL_G
Arginine	g	ARG_G
Histidine	g	HIS_G
Alanine	g	ALA_G
Aspartic acid	g	ASP_G
Glutamic acid	g	GLU_G
Glycine	g	GLY_G
Proline	g	PRO_G
Serine	g	SER_G
Cholesterol	**mg**	**CHOLE**
Fatty acids, saturated	**g**	**FASAT**
4:0	g	F4D0
6:0	g	F6D0
8:0	g	F8D0
10:0	g	F10D0
12:0	g	F12D0
14:0	g	F14D0
16:0	g	F16D0
18:0	g	F18D0
20:0	g	F20D0
18:1	g	F18D1
18:2	g	F18D2
18:3	g	F18D3
20:4	g	F20D4
22:6	g	F22D6
22:0	g	F22D0
14:1	g	F14D1
16:1	g	F16D1
18:4	g	F18D4
20:1	g	F20D1
20:5	g	F20D5
22:1	g	F22D1
22:5	g	F22D5
Phytosterols	mg	PHYSTR
Fatty acids, monounsaturated	**g**	**FAMS**
Fatty acids, polyunsaturated	**g**	**FAPU**
15:0	g	F15D0
17:0	g	F17D0
24:0	g	F24D0

TABLE 2

Top contributors of fat in the US diet[1]

Butter
Cheese/cheese food
French fries
Ground beef
Ice cream
Margarine and spreads
Mayonnaise
Milk
Shortening

[1] Quartile 1, accounts for 25% of total intake of fat.
Not listed in order of amount contributed.

these foods actually spans an extensive list of related foods within a food, and include variations due to adjusted fat levels (e.g. vegetable oil spreads and mayonnaise), cooking methods (e.g. ground beef and chicken), and a variety of oil combinations within a food. Margarine and spreads were identified as high-priority foods because they are consumed widely at the table and are also a component of 1090 of the 7352 recipes (about 15%) in the USDA Survey database (USDA, 1998a, b). Margarine and spreads are top contributors of energy, fat, monounsaturated fatty acids, polyunsaturated fatty acids, *trans*-fatty acids, vitamin E, and carotene to the U.S. diet.

This research focuses on the development of an "all-purpose" nationwide sampling plan, the result of which best represents these foods as consumed by the U.S. population. We are currently collecting retail outlet samples but in the future will expand to restaurant foods, and multiple sampling over time to address seasonal variations in select nutrients and foods.

METHODS AND MATERIALS

Our sampling plan was a three-stage design where counties (Consolidated Metropolitan Statistical Areas (CMSAs) or counties) were selected at the first stage, grocery store outlets within the selected counties were selected at the second stage, and specific food products to be purchased and analysed for nutrient content were selected at the third stage. In effect, this gave us a sample of grocery outlets from selected geographical dispersed areas across the United States.

The plan was based on four regions with roughly equal populations, each between 65 and 68 million people and based on geographical proximity of states (Table 3). The exception was the inclusion of Texas with the Midwestern states (Fig. 1). Alaska and Hawaii were excluded for logistical reasons. This was done to best equalize the distribution of the regional populations and because Texas, an agricultural as well as an industrial state, reasonably resembled the Midwestern states. The next step involved selecting three CMSAs from each region. However, since all counties are not included in a CMSA, Generalized Consolidated Metropolitan Statistical Areas (gCMSAs) were developed. The gCMSAs were defined as the standard CMSAs or individual counties for areas not in a CMSA. Once the gCMSAs within a region were sorted in descending order by the population size, a PPS sample of size three was systematically drawn within each region (Cochran, 1977). Once the gCMSAs were selected, the counties that made up each gCMSA were sorted in descending order by

TABLE 3

First stage sampling regions: National Food and Nutrient Analysis Program

Region no.	Region name	1997 estimated population
1	Northeast	66 492 898
2	South	65 245 265
3	Great Lakes and Texas	68 014 743
4	Plains, Rockies and Pacific	67 883 155

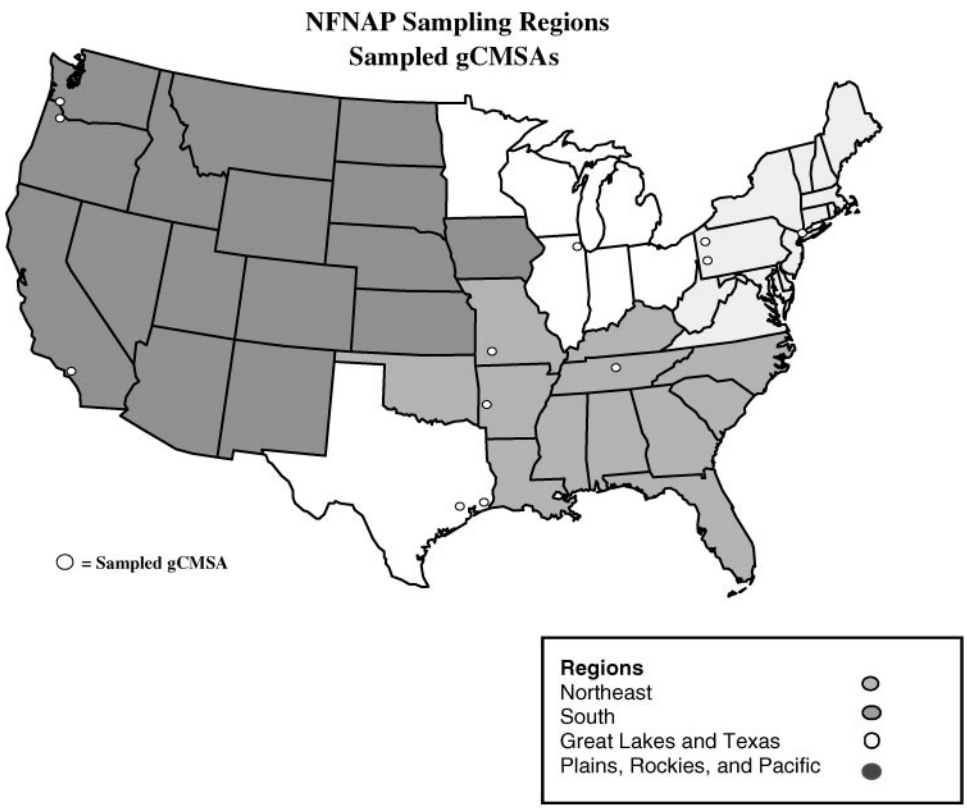

FIGURE 1. NFNAP sampling regions.

their urbanicity (Goodall *et al.*, 1998) and a sample of size two was systematically selected (Cochran, 1977). Sorting counties within gCMSAs by urbanicity ensured that the sample contained both more urban and less urban areas. For those gCMSAs made up of only a single county, the county was selected twice. A summary of the national sampling plan is as follows:

1. US divided into four regions, with roughly equal populations.
2. Each region divided into three strata of high, medium, or low population density.
3. Primary sampling units in each stratum were gCMSAs.

4. Two counties chosen proportional to urbanicity (1 rural and 1 urban) from each selected gCMSA.
5. Selected gCMSAs supplemented with surrounding counties when the gCMSA contained less than 10 grocery stores.
6. Retail stores with sales >$2 million/year selected from grocery store lists for selected counties.
7. Products chosen proportional to package size and market share, using market share information from Nielsen, Inc. or Information Resources, Inc (limiting to manufacturers with 1% or more of market volume); same product sampling applied to all selected outlets.
8. Foods composited by sample number, providing individual product (i.e. brand) data for major brands and overall results for the particular food product. *Results pertain to an average serving from the homogenized food product, not to a typical serving.*

In most cases, 6–12 samples were analysed for each food. Each sample was a composite of packages or units from 24 outlets, yielding in most cases, brand-specific samples of the mean content of selected nutrients. Using this sampling plan, estimates derived from composite samples will be self-weighted. The number of composites are determined by the relative importance of the Key Food. Figure 2 shows how the foods are sampled extensively to achieve a representative picture of what is being consumed nationally; through compositing, we reduce the cost of analysis and still produce a very reliable estimate of the actual nutrient concentrations in the foods. This allows us to sample and analyse more foods. For example, by compositing sample 1 across 24 outlets, sample 2 across 24 outlets, up to sample 12 across 24 outlets, we, in essence, end up with 12 composite samples to be analysed—or an "*n*" of 12 associated with the grand means of the nutrient analysed. However, the mean represents an extensive, representative sampling of the population, a theoretical *n* of 288. The frequency of samples per brand for a given food depends on the extent of the brand name distribution across market shares. The sample may contain many brand names or only a few, depending on their market share and probability to be sampled. For foods eaten cooked or raw and cooked, the Nutrient Data Laboratory (NDL) is analysing a split sample, both forms, to secure yield and retention data for selected nutrients. Some foods, such as ground beef of various fat levels, will be analysed using several cooking methods to represent what is actually consumed in the US diet.

Formulation changes in processed foods, regional differences, shipping and storage conditions, analytical variability, and serving to serving variability may account for considerable differences in certain nutrients in foods; these may introduce health concerns for many individuals. For some foods, where concentrations of specific critical nutrients are very high and where we would expect wide fluctuations of these nutrients, additional, randomly selected samples are analysed individually to determine estimates of variability. Random sampling across gCMSA and independently analysed samples yield information on variability for select nutrients between servings, brands, and geographic areas.

RESULTS

To date, we have sampled and analysed margarine and spreads of all fat levels, a variety of mixed dishes (e.g. pizza, lasagna, spaghetti, pot pies, macaroni and cheese), folate-fortified foods (i.e. retail flour, bread, rolls, rice, pasta, and mixed dishes), American pasteurized processed cheese food, salad dressings, several condensed and

dry soups, cheese, olive and corn oil, and municipal water from 24 locations across the U.S. (Table 4). Individual foods are itemized within food groups and described by a number of analytical samples, either through nationwide or limited sampling, and the relative importance of the food. Each food type includes several variations, so we have actually fully sampled close to 100 foods and 130 foods when including pilot studies. Limited sampling and analysis using reserve samples from other USDA studies has been done on soy-based foods, bottled tea, 2% fat milk, and butter. Nutrient data on these foods are now being reviewed by NDL staff and will be used to update future releases of the USDA nutrient databases.

Through the NFNAP sampling plan, we are able to continuously redefine existing food descriptions and classifications to reflect the market. This ultimately gives us the most representative picture of total nutrient intake contributions by food and overall. For example, based on information available at the time, we had classified margarine by generic fat levels or form of oils contained in the product. The nutrient profiles for these foods were less precise than what has been generated through NFNAP. With more precise descriptions of the food, we can generate more precise, representative nutrient profiles. In future releases of our databases, for example, product descriptions of margarine and spreads will be by the exact fat level, its form, and the oils or oil blends used (Table 5). The NFNAP data for 70% fat spread and 80% fat margarine will replace existing data that no longer reflect products in the market. The data for the 70% were based on limited data for a squeeze bottle product; the new data are comprehensive and representative of one of the most highly consumed spreads across all fat levels. In essence, the nutrient profile for this food did not exist for the popular tub and stick forms prior to NFNAP. The 80% fat proximate data are not considerably different from the older data. The mineral data, however, reflect the current lower sodium (a critical nutrient) content of margarine and spreads.

Some of the foods that will be sampled and analysed in the coming year include commercially prepared fried chicken, ground beef, including many fat levels and cooking procedures, brewed coffee and tea, mayonnaise and oils, additional mixed dishes, and a number of ethnic foods such as American Indian, Hispanic, and African–American dishes. We also plan to sample fresh fruits and vegetables over time as multiple sampling of the same food over time will determine seasonal fluctuations in specific nutrients. We are collecting multiple samples of municipal water (same location and over time) to determine the seasonal variability of trace minerals. The list is by no means exhaustive but represents the breadth of the higher priority foods yet to be analysed.

DISCUSSION

The current sampling plan addresses the 660 foods identified as Key Foods in the food supply. However, other foods or food ingredients are emerging as important contributors of nutrients to the US diet as a whole, or to the diet of specific subpopulations. Therefore, sampling plan options are being developed for foods and food ingredients not consumed on the national level or where variability in the specific nutrients can be assessed better through other methods. For some foods, sampling at the point of production may prove better than sampling nationwide. For example, we may find differences in how grain products such as flour are fortified at the plant level. Foods or food ingredients used only by the food industry and not available at the retail level may also be best sampled at the plant level because of company-specific nutrient specifications. In addition, foods may be consumed by a certain population group or within a specific region such as foods consumed by one of the hundreds of American

TABLE 4

National Food and Nutrient Analysis Program: Sampling completed by March 1999 (SV = additional independently analysed samples for nutrient variability estimation; "Raw/ckd" indicates samples analysed in both raw and cooked forms to determine nutrient retention values)

Foods	All samples (no. foods)[1]	NSP samples (no. foods)[2]	Importance
Margarine/Spreads	43 (30 brands)	42 (30 brands)	Top contributor of fat in U.S. diet. In 15% of survey recipes. Existing data did not reflect proliferation of low- and reduced-fat products. New data will provide values for all levels of fat including the reduced- and low-fat products
80% fat range	8	8	
70% fat range (+ SV)	12	12	
40–60% fat range	12	11	
< 38% fat range	8	8	
fat-free	3	3	
Butter	1 (1)	0	Pilot/monitoring study
Soups, condensed/dry	37 (13)	36 (12)	Will provide new nutrient data for common soups where data were largely unavailable
Mixed dishes	87 (16)	82 (13)	
beef stew, canned	13 (1)	12 (1)	Major contributor for 12 critical nutrients
chili, canned (+ SV)	24 (2)	24 (2)	Top contributor of 11 critical nutrients
pizza, cheese (+ SV), raw/ckd	12 (1)	12 (1)	Top contributor of 11 critical nutrients
lasagna, meat, raw/cooked	13 (1)	12 (1)	Top contributor of 6 critical nutrients
spaghetti, canned	13 (2)	12 (2)	Top contributor of 12 critical nutrients
macaroni and cheese, raw/ckd	12 (4)	12 (4)	Top contributor of 6 critical nutrients
pizza, pepperoni (+ SV)	2 (1)	0	Top contributor of 14 critical nutrients
pot pie, chicken	1 (1)	0	Top contributor of 7 critical nutrients
pizza, vegetable and meat, raw/ckd	1 (1)	0	Top contributor of 12 critical nutrients
ramen noodles	8 (2)	2 (2)	New data where none previously available
Parmesan cheese	6 (1)	6 (1)	Top contributor of 7 critical nutrients

Salad dressings Italian: regular/lite/fat free French: regular/lite/fat free Thousand Island: regular/lite/fat free	36 (9)	36 (9)	Major contributor of fat to U.S. diet recipes. Existing data did not reflect proliferation of low- and reduced-fat products. New data will provide values for these products
Oils olive oil corn oil	8 (2) 4 (1) 4 (1)	8 (2)	Major contributor of fat and fatty acids. Used in many recipes
American cheese	6 (1)	6 (1)	Top contributor of 7 critical nutrients; USDA subsidy programs, fast food/restaurants
Folate-fortified foods[3] flours, wheat spaghetti, raw/ckd rice, regular/instant/parboiled, raw/ckd macaroni, elbow, raw/ckd bread, white (+SV)	54 (16) 18 (5) 8 (2) 12 (6) 8 (2) 8 (1)	54 (16) 18 (5) 8 (2) 12 (6) 8 (2) 8 (1)	Top contributors of folate. Will provide new analytical data to reflect changes in fortification regulations. Used in over 1000 survey recipes USDA subsidy programs
Milk, 2%	1 (1)	1 (1)	Top contributor of 10 critical nutrients; used in numerous survey recipes and USDA subsidy programs. Most prevalent form of milk consumed. (Pilot study)
Tea, bottled	30 (14 brands)	0	New data for high-consumption beverages
Soy products	67 (28 brands)	0	New data for isoflavone-containing foods
Water, municipal (SV only)	24 (1)	24 (1)	Pilot study to explore use of water in preparation of other beverages
TOTAL	399 (133)	294 (86)	

[1] Includes national sampling and limited sampling on reserve samples from other USDA studies.
[2] NSP (National Sampling Plan) includes only nationally sampled foods (24 outlets).
[3] Mixed dishes contain folate fortified ingredients listed in this category and are therefore folate contributors.

TABLE 5

Changes in classifications of margarine and spreads: NFNAP sampling

Previous classification
Types of oils/oil combinations, fat not specified; or
generic fat level (80, 60, 40, 20% fat); or
fat-free; or
form (stick or tup); and/or
soft or hard
NFNAP classification
More precise fat levels (e.g., 80, 70, 60, 53, 48, 37%, fat-free); and
form (stick, tup, squeeze bottle, spray); and
specific oil or oil blend

Indian tribes, or by specific segments of the Hispanic community. Limited analyses of specific nutrients may be conducted on samples reserved from other studies as pilot work.

The new NFNAP sample design provides a geographically dispersed, proportionally representative set of samples for a given food item of the amount of the item consumed by the U.S. population. Hence, data collected under the NFNAP design for a specific food item are geographically and statistically representative of the actual food consumed by the U.S. population. Analysis based on the sample data collected using the new NFNAP sample design provides much more accurate estimates of the nutrient content of a food currently consumed than analysis based on historical nutrient data available in the NDL databases. This sampling plan serves as the template for continuing research at NDL; all stages of the design can be recalculated to reflect the most current population distribution, sales and market share data over time.

In addition to the nationwide sampling of mainstream foods, we will also focus on the development of the database for Native American foods, categorized as mainstream, USDA commodity, traditional, and recipes. Some of the anticipated challenges intrinsic to this type of research are: (1) handling the many variations on ingredients and preparation methods across tribes for a single food (e.g. frybread); and (2) securing reliable population distribution information (given census data are not available). Resources that we plan to use in developing this plan include Indian Health Service (IHS) clinic surveys, surveys in other official institutions, scientific and government publications, the USDA commodity foods distribution program office, personal communications with tribal leaders and IHS officials, and focus groups. Through the informal surveys conducted at the IHS clinics and of the published literature, we are developing a list of the most commonly consumed, more traditional Native American foods (e.g., cornbread, tortillas, frybread, goulashes, wild game stews, soups and goulashes, and region-specific meats). Through NFNAP's umbrella sampling plan, we will be able to analyse the mainstream foods but for the more traditional foods, our sampling plans will need to be less structured and tribe-specific and our compositing of samples even more compressed to accommodate more foods.

CONCLUSION

Through NFNAP and the new database system being developed at NDL, we will continue to update USDA's food composition databases to support nutrition-related research in the scientific community—in analytical methodology, quality control

procedures, statistical sampling, and data quality evaluation with numerous applications in trade, food safety, and research. Compilation and dissemination of self-weighting NFNAP data allows for accurate and representative mean estimates of nutrient profiles in generically described foods as well as brand-specific products. Ultimately, we will link the values for the most important food and nutrients to valid, well-documented analytical values generated in this project. Finally, using state-of-the-art information technology currently being developed in our new database system, we will effectively disseminate these data to the scientific community, the food industry and consumers.

REFERENCES

Cochran, W. G. (1977). *Sampling Techniques*. Wiley, New York, NY.

Federation of American Societies for Experimental Biology (FASEB), Life Sciences Research Office. Prepared for the Interagency Board for Nutrition Monitoring (1995). Third Report on Nutrition Monitoring in the United States. US Gov. Printing Office. Washington, DC.

Goodall, C. R., Kafadar, K., and Tukey, J. W. (1998). Computing and using rural versus urban measures in statistical applications. *The Am. Stat*. **52,** 101–111.

Haytowitz, D. B., Pehrsson, P. S., Smith, J., Gebhardt, S. E., Matthews, R. H., and Anderson, B. A. (1996). Key foods: setting priorities for nutrient analysis. *J. Food Comp. Anal*. **9,** 331–364.

Pehrsson, P. R., and Haytowitz, D. H. (1995). Development of key foods in food composition research. *20th National Nutrient Databank Conference Proceedings,* June 10–13, Buffalo, NY.

Perry, C. R., and Beckler, D. G. (1999). Basic formula for aggregated estimates. Unpublished Working Paper from contract between USDA, Nutrient Data Laboratory and USDA, National Agricultural Statistics Service.

US Department of Agriculture, Agricultural Research Service (1998a). USDA Nutrient Database for Standard Reference, Release 12. Nutrient Data Laboratory Home Page, http://www.nal.usda.gov/fnic/foodcomp.

US Department of Agriculture, Agricultural Research Service (1998b). CSFII/DHKS 1994–96 Data Set, Documentation and Technical Support Files: The 1994–96 Continuing Survey of Food Intakes by Individuals and the 1994–96 Diet and Health Knowledge Survey (CDROM: Accession No. PB98-500457; National Technical Information Service, Springfield, VA).