FROM NUTRIENT DATA TO A DATA BASE FOR A HEALTH AND NUTRITION EXAMINATION SURVEY ORGANIZATION, CODING, AND VALUES--REAL OR IMPUTED

Connie M. Dresser, R.D.
Nutrition Analyst
Division of Health Examination Statistics
National Center for Health Statistics

The National Health and Nutrition Examination Survey (NHANES) and its predecessor, the Health Examination Survey (HBS), collect and utilize data that are obtained by direct physical examination, clinical and laboratory tests, and related measurement procedures. Prevalence data are collected for specifically defined diseases or conditions of ill health; and formative health-related measurement data are collected that show distributions of the total populations with respect to particular parameters such as blood pressure, visual acuity, or serum cholesterol level. Prior to NHANES, the HES programs focused on selected chronic diseases; growth and development and sensory defects of children and youths.

Data on nutrition were first collected in NHANES I and again in NHANES II to monitor changes in nutritional status over time. Data collection methods were selected to record dietary information and a nutrient data base was compiled to process the data. This paper outlines the organization of the original NHANES nutrient data base, its complexities and transitions over the past decade.

ORGANIZATION

The original NHANES dietary component was developed in 1970 by the NHANES nutrition consultant in conjunction with the U.S. Department of Agriculture's Food Composition Research Group. The 24-hour recall method was chosen to record the types and amounts of foods consumed and the vitamin-minerals developments used by the NHANES population on a routine daily basis. The food frequency method was used to depict habitual consumption patterns over the three months prior to their interview. A coding manual and nutrient data base were developed to code and calculate the nutrients contained in the foods reported on the 21-hour recall questionnaires.

After the first year of data collection, the dietary component was evaluated. The format of the nutrient data base was changed to better accommodate the limited headquarters staff and dietary field staff who routinely coded the questionnaires at the collection site. The format was changed to double as a coding manual and the basis for data processing. Numerous problems encountered during the first NHANES were resolved empirically by analysis of the various data collection and processing components.

The nutrient data base consisted of a Model Gram and Nutrient Composition section (Table 1). The model gram section included 1) food group and 2) food code numbers, 3) assigned measure model codes and 4) the food descriptors.

The food groups used in NHANES I were adapted from the Ten-State Survey.

TABLE 1-1 details the groups used to process the 24-hour recall and food frequency forms in both surveys. Foods were assigned to food groups based on descriptive and nutrient content similarities, e.g. milk, cheese, and yogurt are major sources of calcium, phosphorus, and riboflavin. However, in NHANES I, the food items assigned to the 24-hour recall food groups did not necessarily correspond to those foods assigned to the food frequency questionnaire. Discrepancies between the two sets of food groups made it inconvenient to Cross-Check food patterns without first reconstructing food items. Therefore, in NHANES II, the same foods were assigned to the 18 food groups used for both methods.

Food codes available for coding in NHANES I and II, numbered 3630 and 4762 respectively. In NHANES I, the majority of the codes came from USDA food composition handbooks and data tapes, and Tulane University's Dietant Listing, with a minimal number of codes assigned to processed foods from industry or other references (Table 2). Periodically, a recipe was calculated at NHANES headquarters for a mixed dish commonly consumed for which there was no information from USDA. In NHANES II, the number of USDA food codes decreased while the number of processed food codes increased. The decision to add codes was made because food codes provided by USDA, while giving generalized descriptions, did not give sufficient information about packaging, processing, or fortification of brand named foods.

Each food code was assigned a measure model code. This code consisted of 1) an alphabetic symbol used to designate the preferred food model and 2) a gram conversion factor, which is the gram equivalent of one ounce of the corresponding food item. This code was used to convert the ounce equivalents of the food model reported to grams, and then to total nutrients. The alphanumeric description of each food item represented general characteristics, and the method of preparation and packaging, e.g. milk, whole, condensed, canned.

Coding, in NHANES, was conducted at the site of data collection by trained dietary interviewers who held a B.S. in home economics. Their one-week initial training consisted of interviewing techniques, the correct use of questionnaires, food models, the food code manual, and the procedures used to conduct meetings with nutrition professionals in the areas to be surveyed. Followup supervision and interviewer evaluation was conducted by the staff nutrition supervisor.

In both surveys, the Model Gram section of the dietary data base was used as the interviewer coded and verified the majority of food items within 72 hours of collection. Foods left uncoded were identified and coded at headquarters. If the respondent could not recall the exact food consumed or approximate the serving size, no arbitrary substitutes or serving sizes were assigned. Dietary interviewers were responsible for preliminary quality control checks on the collected data. Programmed edits were completed at headquarters to check for missing data or errors in coding, amounts of food or food models reported.

The nutrient composition section of the data base (Table 1) included food group numbers, food code numbers and data on kilocalories, dietary components and nutrients for 100 gram edible portion sizes of respective food codes in the Model Gram section. Data for trace elements and some micronutrients were not added to the data base because the majority of this information was either imputed or available for a limited number of food items.

Updating the NHANES nutrient data base took place twice during both surveys (Table 3). At the beginning of NHANES I (1971), the data sources included:

- food codes from USDA's 1963 data set 8-1. At that time data set 8-1 did not include analyzed nutrient values for total saturated fat, oleic, linoleic acids, and cholesterol. Values for sodium, potassium and phosphorus were limited (code series 0-2700).

- food codes from Tulane University (code series 5000 and 6000).

- unpublished information from USDA which was either imputed data or a recipe calculation (code series 35000).

- nutrient values from USDA's Home and Garden Bulletin #72 and the 1970 edition of Bowes and Church (code series 72000 and 85000).

- a minimal number of nutrient values for foods from IIT Continental Baking Company and other food companies (code series 30000 and 90000).

At the end of NHANES I and before data processing, the nutrient data base was evaluated for 1) programmatic errors, e.g. appropriate decimal placement 2) duplicate codes and nutrient values, 3) missing nutrient values, and 4) appropriate values applicable to the time period. Therefore, in 1975,

- The majority of code series 5000 and 6000 from Tulane University were eliminated because, with few exceptions, the codes were identical to the food codes in the 1963 version of USDA's Handbook #8.

- USDA's 1972 Expansion data tape 8-1-1 of Handbook #8 replaced data set 8-1; however, the adaptation of this tape did not fill in all blanks nor did the literature necessarily provide additional needed information.

- all other remaining code series were updated, where possible, with nutrient values from appropriate sources. On the NHANES I Model Gram/Nutrient Composition data release tape, a symbol of 8888888 was used to indicate that the data were unavailable.

At the beginning of NHANES II (1976), the following changes were made:

- all food codes and nutrient values on the USDA data tape 8-1-1 were replaced with food codes and values from USDA's Handbook #456. The nutrient values were taken from data tapes 456-1 and 456-2 (code series 0-2700).

- remaining Tulane University codes were updated with values from USDA's Home and Garden Bulletin #72, the 12th edition of Bowes and Church, USDA's Handbook #456 and the National Heart Study (code series 5000 and 6000).

- in 1978, NHANES II went to Hawaii and the nutrient values for foods reported were taken from research bulletins and circulars applicable to foods grown on the island (code series 8000).

- NHANES I foods taken from the IIT Continental Baking Company were updated with information from the company or with nutrient values of similar products used in USDA's 1977/78 Individual National Household Food Consumption Survey (code series 30000).

- USDA baby food codes were replaced with nutrient values from Baker/Beechnut and Gerber baby food companies (code series 10000 and 41000).

- USDA's unpublished information was updated with calculated values or recipes (code series 35000).

- foods taken from USDA's Home and Garden Bulletin #72 and Bowes and Church were updated with current editions of both publications (code series 72000 and 85000).

- new commercial foods added to the data base were assigned codes between 90000 and 95000.

At the close of NHANES II, only values for foods reported twenty or more times were updated to reflect knowledge of the nutrient content of foods between 1976 and 1980. The update was limited due to staff and time constraints. If values from industry were unavailable, values were substituted for similar foods in USDA's Handbook #456, and revised editions of Handbook #8, sections, 1-6. If no appropriate values could be found, a symbol of 9999999, not zero, was assigned on the Nutrient Composition data release tape to indicate a thorough search had been made but no reliable information was available.

NUTRIENT VALUES--REAL OR IMPUTED?

Undating the NHANES dietary data base, in both surveys, was hampered by the unavailability of nutrient values applicable to the time periods. This is important to remember when deciding whether the values addressed are "real or imputed". For instance, years ago most of the nutrient data published by the Nutrient Composition Laboratories of the Department of Agriculture were based on direct chemical analyses. The majority of this data has been replaced with composite or composition values from industry, independent laboratories, and scientific or technical literature. Although the nutrient values on USDA data tapes and in handbooks are suitable for determining intakes on kinds and amounts of foods consumed and reported in today's dietary surveys, the composition values represent foods as used throughout the country on a year-round basis. Values appropriate at one time may not apply during a latter period when new or different varieties predominate; when production, manufacturing or preparation processes have changed, or when different ingredients or improved methods of analysis are used. However, while numerous food companies have been most cooperative with the dissemination of their product analyses, users of this data are unaware of methods of analysis or frequency of imputation. It is acknowledged by food companies and most data base developers that the label or fact sheet data provided to the consumer are often based on the percent of the U.S. RDA's or on the NRC RDA standards. When label data are compared to corresponding 100 gram edible portion data, there are discrepancies. Both of these data sources offered limited data on sugars, fiber and nutrients such as zinc, copper, magnesium, folacin, vitamin B6, vitamin B12, and pantothenic acid. Concurrently, the effects of processing and packaging methods and product fortification make it impossible to define the "real values" of foods used in large-scale studies. These unavoidable conditions make it important to know the limitations and caveats of nutrient data bases when analyzing dietary data.

Another important issue to address is the effect of missing nutrient values in the data base. For example, Table 4 suggests that the multitude of missing values accompanying the brand named foods in both the NHANES surveys should have a negative effect on the total nutrient intakes. Before assuming this is true, the data tape user should know how often the foods with missing values were consumed. Table 4 also suggests that the data from sources other than food manufacturers provided more information on nutrient values in foods in NHANES II than in NHANES I. However, without further investigation, the data tape user does not know the quality of this additional data. For example, in NHANES I, 1972 USDA data set 8-1 gave no options for coding foods cooked with salt versus no salt or foods prepared with butter, oil or margarine. This information was available in data set 456-3 during NHANES II. These changes in the data base, in part, contribute to the observed changes in nutrient intakes between the NHANES surveys.

Because data tape users or developers rarely know what percentage of the data are imputed or "real", careful monitoring of the data base is essential. Changes made in the data base will partially explain the shifts in nutrient intakes by population sub-growth in surveys. In NHANES, changes in nutrient intakes between surveys can be attributed to the differences in frequency of food item consumption, total grams of food consumed, changes in food fortification, and to the changes made in the nutrient data base.

It is important to remember that NHANES is not in the "nutrient data base business". Our data base is used primarily by researchers who are addressing questions related to health, nutrition and disease. However, as a data base developer, we must be knowledgeable of the limitations of the nutrient values which will eventually be used by researchers studying the interrelationships of diet and health. Being aware of these caveats is not enough if minimal progress is made to improve the state of the art--a problem which has long faced large-scale surveys. Dietary methodologies and nutrient data bases currently used are still too costly and time consuming. However, work is being done to improve the dietary collection methods and methods of analyzing nutrients. One methodological example is the "core frequency questionnaire" used in the NHANES I Epidemiological Followup Survey at the National Center for Health Statistics. The format includes 100 precoded food items most often consumed in both of the NHANES surveys. The selection of the food items was based on the number of times the food item was reported, total gram consumption and its contribution to the majority of nutrient intake. To calculate median nutrient values for daily and weekly consumption of these foods, a formula using frequency count, reported serving size information and an aggregated nutrient values per food item is necessary. This procedure is relatively simple until nutrient values are addressed. The nutrient data base needed to process this data should be standardized. Standardization of the base is an issue that needs to be evaluated and set into motion before substantial progress can be made in linking dietary intake data to health and disease.

References

1. Baker-Beechnut Baby Food Company, CanoJoharie, New York. 9/1975.

2. Bowes and Church; Food Values of Portions Commonly Used. 1970. 11th edition.

3. Bowes and Church;: Food Values of Portions Commonly Used. 1975. 12th edition.

4. Centers for Disease Control: Ten-State Nutrition Survey in the United States, 1968-1970. Part V. Dietary. DHES Pub. No. (HMS) 72-8133. Atlanta, Georgia.

5. Contemporary Nutrition, Watt, B.K.: Tables of Food Composition: uses and limitations. Vol. 5, No. 2. Feb. 1980.

6. Department of Health, Education and Welfare, Leung, W-T, Butrum R., Roa, M. and Polacchi, W.: Food Composition table for use in East Asia. DHEW Pub. No. (NIH) 79-165. Dec. 1972 - reprint 1978.

7. Gerber Baby Food Company, Fremont, Michigan. 9/1975.

8. ITT Continental Baking Company, P.O. Box 751, Rye, New York. 1970-71.

9. National Center for Health Statistics, Pearce, N.; Data Systems of the National Center for Health Statistics. Vital and Health Statistics. Series 1, No. 16. DHHS Pub. No. (PHS) 82-1318. Public Health Service. Hyattsville, Maryland, December 1981.

10. National Center for Health Statistics: Public users tape documentation: NHANES I - Model Gram/Nutrient Composition Tape Cat. No. 1702, NTIS No. PB-296027. Public Health Service. Hyattsville, Maryland, May 1981.

11. National Center for Health Statistics: Public users tape documentation: NHANES II - Model Gram/Nutrient Composition Tape, Cat. No. 5704, NTIS No. PB-82-142613. Public Health Service. Hyattsville, Maryland, Feb. 1982.

12. National Institutes of Health: The National Heart Study. 1965.

13. Tulane University; Tulane Dietant Listing. 1969.

14. University of Hawaii, Agriculture Experiment Station, Wenkam, N. and Miller, C.; Bulletin 135 - C85 - Composition of Hawaii Fruits. Reprint October 1973.

15. University of Hawaii, Agricultural Experience Station, Standal, R., Bassett, D., Policar, P., and Thom, M.; Bulletin 146 - Fatty Acids, Cholesterol, and proximate composition of certain prepared and unprepared foods in Hawaii. May 1975.

16. University of Hawaii, Agricultural Experiment Station, Miller, C. and Branthoover, B.: Circular 52 - Nutritive values of some Hawaii foods. Reprint Dec. 1963.

17. U.S. Department of Agriculture, Agriculture Research Service, Consumer and Food Economics Research: Composition of foods--raw, processed, prepared. Handbook No. 8 revised. Data set 8-1. Dec. 1963.

18. U.S. Department of Agriculture, Agriculture Research Service, Consumer and Food Economics Research; Composition of foods--raw, processed, prepared. Expansion No. 1 of data published in Agriculture Handbook No. 8. Data sets 8-1-1, 8-2-1. April 1972.

19. U.S. Department of Agriculture, Agriculture Research Service, Consumer and Food Economics Research: Expansion of data published in Nutritive Values of American Foods in Common Units. Handbook No. 456. Data sets 456-1, 456-2. May 1977.

20. U.S. Department of Agriculture, Agriculture Research Service, Consumer and Food Economics Research: Composition of foods. Handbook Nos. 8-1 through 8-6, Nov. 1976-Feb. 1980.

21. U.S. Department of Agriculture, Agriculture Research Service: Nutritive Values of foods. Bulletin 72. 1970.

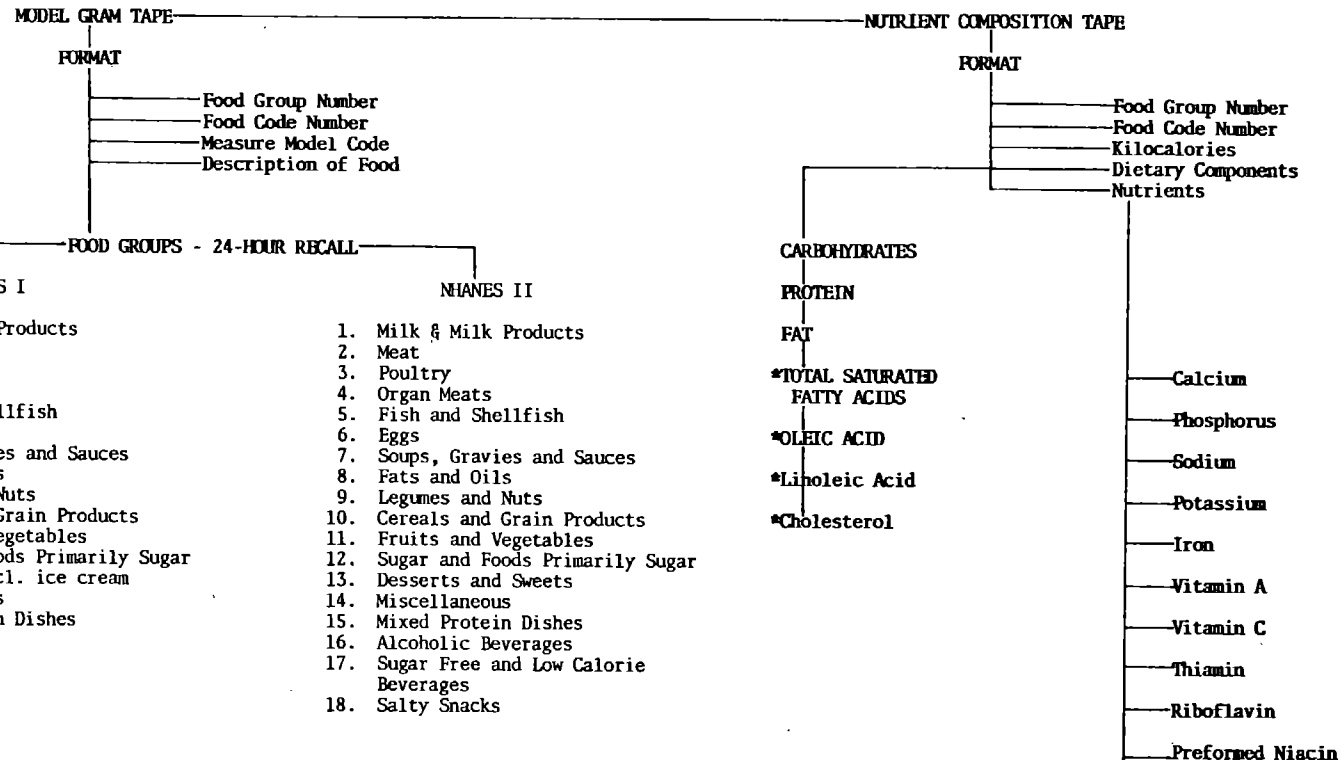
22. U.S. Department of Agriculture, Agriculture Research Service, Science and Education Administration: Nutritive Values of foods. Bulletin 72, 1977.

23. U.S. Department of Agriculture, Agriculture Research Service Consumer and Food Economics Research: 1977/78 Nationwide Food Consumption Survey on Individuals. Tape PB82138504, released 1-1980.

24. U.S. Department of Agriculture, Agriculture Research Service, Consumer and Food Economic Research, Food Composition Research Group: unpublished data. 1971.

TABLE 1

NHANES NUTRIENT DATA BASE



*NHANES II NUTRIENT COMPOSITION TAPE ONLY

TABLE 1-1 NHANES FOOD GROUPS

NHANES I (1971-1974)		NHANES II (1976-1980)
24-Hour Recall Food Groups	**Food Frequency Food Groups**	**24-Hour Recall/Food Frequency Food Groups**
1. Milk and Milk Products	1. Milk Whole Skim	1. Milk and Milk Products Whole Milk Skim Milk Ice Cream Cheese and Cheese Dishes
2. Meats	2. Meat and Poultry	2. Meats
3. Poultry	3. Fish and Shellfish	3. Poultry
4. Organ Meats	4. Eggs	4. Organ Meats
5. Fish and Shellfish	5. Cheese and Cheese Dishes	5. Fish and Shellfish
6. Eggs	6. Dried Beans and Peas	6. Eggs
7. Soups, Gravies and Sauces	7. Fruits and Vegetables those rich in Vitamin A those rich in Vitamin C	7. Soups, Gravies and Sauces
8. Fats and Oils	8. Bread and Cereals	8. Fats and Oils
9. Legumes and Nuts	9. Butter and Margarine	9. Legumes and Nuts
10. Cereal and Grain Products	10. Desserts and Sweets	10. Cereals Grain Products
11. Fruits and Vegetables	11. Candy	11. Fruits and Vegetables those rich in Vitamin A those rich in Vitamin C
12. Sugar and Foods Primarily sugar	12. Beverages	12. Sugar and Primarily Sugar Products
13. Desserts, inc. ice cream	13. Snack Foods	13. Desserts and Sweets
14. Miscellaneous		14. Miscellaneous
15. Mixed Protein Dishes		15. Mixed Protein Dishes
		16. Alcoholic Beverages Beer Wine Distilled liquor
		17. Sugar Free and low Calorie Beverages Artificially sweetened drinks Coffee or Tea
		18. Salty Snacks

TABLE 2

SOURCES OF CODES REPORTED IN NHANES BY CODE SERIES

CODE SERIES		USDA HB.# #8 456 0-2700		Tulane Univ. 5000/6000		Hawaii & East Asian 8000		Continental Baking Co. 30000-31000		USDA Confid. information 35000		Baker/ Beechnut 40000	Gerber's 41000	USDA Bull. #72 72000		Bowes&Church 85000		Commercial 90000-95000	
FOOD GROUPS	NHANES	I	II	I	II	I	II	I	II	I	II	II	II	I	II	I	II	I	II
1. DAIRY		51	50	7	7	-	0	0	0	0	0	2	2	1	1	2	2	5	81
2. MEAT		134	165	22	6	-	0	0	0	1	0	3	9	10	0	0	0	0	31
3. POULTRY		39	43	3	2	-	0	0	0	7	0	3	5	0	0	0	0	0	5
4. ORGAN MEATS		19	23	1	5	-	0	0	0	0	0	0	1	0	0	0	0	0	2
5. FISH & SHELLFISH		81	87	7	5	-	11	0	0	0	0	0	0	0	0	0	0	0	3
6. EGGS		9	11	0	1	-	0	0	0	0	0	0	0	0	0	0	0	0	3
7. SOUPS & SAUCES		66	58	10	9	-	1	0	0	0	0	0	0	0	0	0	0	0	56
8. FATS & OILS		37	34	6	7	-	0	0	0	1	1	0	0	7	7	0	0	0	41
9. LEGUMES/NUTS/SEEDS		53	51	2	4	-	0	0	0	0	0	0	0	0	0	1	1	4	25
10. CEREALS & GRAIN PRODUCTS		191	128	44	23	-	2	1	2	7	6	6	18	2	2	0	0	55	174
11. FRUITS & VEGETABLES		432	447	41	25	-	17	0	0	5	5	43	48	0	0	0	0	8	53
12. SUGAR & SUGAR PRODUCTS		76	77	11	21	-	0	0	0	0	0	0	0	0	0	0	0	2	15
13. DESSERTS		98	84	19	25	-	0	14	0	13	10	6	15	1	1	0	0	37	154
14. MISCELLANEOUS		40	29	7	1	-	0	0	0	2	0	0	0	2	1	2	0	4	26
15. MIXED DISHES		58	37	17	10	-	1	4	0	4	3	14	38	0	0	0	0	6	117
16. ALCOHOLIC BEVERAGES		7	8	1	3	-	0	0	0	0	0	0	0	0	0	0	0	0	0
17. LOW CALORIC BEVERAGES		11	11	1	0	-	0	0	0	2	3	0	0	0	0	0	0	5	21
18. SALTY SNACKS		4	4	3	3	-	0	2	2	0	0	0	1	0	0	0	0	0	0
	TOTALS	1406	1347	202	157	-	32	21	4	42	28	77	137	23	12	5	3	126	807

Survey Totals: NHANES I - Codes available for coding: 3630, reported 1825 = 50%
NHANES II - Codes available for coding: 4762, reported 2604 = 55%

TABLE 3

NHANES NUTRIENT DATA BASE UPDATE

CODE SERIES	NHANES I (1971)	NHANES I(1975)	NHANES II(1976)	NHANES II(1980)
0-02700 USDA's Handbook #8	Data set:8-1, 1963 version	1972 Expansion of Handbook #8, Data set:8-1-1	USDA's Handbook #456-3 Data set: 456-1 and 456-2	
5000/6000 Tulane University's Dietant Listing	Data from 1947-1969 majority of codes were duplicates of USDA's Hb. #8.	Duplicate codes were eliminated.	5000/6000 codes from NHANES I were updated with: with: o USDA's Home & Garden Bulletin #72, 1971 edition o Bowes & Church, 12th ed. o USDA's Handbook #456 o The National Heart Study	Values updated with same sources, where appli-, cable, also the Nationwide Food Consumption Survey data base developed for individual intakes. (HFCS)
08000 Hawaiian Foods and East Asian Foods	**	**	o Research Bull. #135 - Composition of Hawaii Fruit o Research Bull. #146 - Fatty Acids, cholesterol, and proximate composition of certain prepared and unprepared foods in Hawaii o Circular #52 - Nutritive values of some Hawaiian Foods o Food composition table for use in East Asia	
30000 ITT Continental Baking	1971 values from company	update from company	HFCS and commercial data	
35000 USDA Confidential Information	1971 values from USDA's Food Composition Research Group	update from USDA	update from USDA	HFCS data and 456-3

TABLE 3 - continued

NHANES NUTRIENT DATA BASE UPDATE

CODE SERIES	NHANES I(1971)	NHANES I(1974)	NHANES II(1976)	NHANES II(1980)
40000 Baker/Beechnut Baby Food Company	**	**	1975 data from company	update from company
41000 Gerber Baby Food Company	**	**	1975 update from company	update from company
72000 USDA's Home & Garden Bulletin #72	1970 edition	updated with the 1971 slightly revised edition	1971 edition	1977 edition
85000 Bowes & Church	1970 edition(11th)	1975 edition(12th)	1975 edition	
90000 Commercial products	data from companies	update from companies	90000-95000 commercial products	o update from appropriate companies o USDA's Hb. #456-3 o HFCS data base o USDA's Hb. #8, sections 1-6

** Code series not included in survey.

TABLE 4 NUMBER AND PERCENT OF FOODS REPORTED IN NHANES WITH MISSING NUTRIENT VALUES

NUTRIENT	NHANES I				NHANES II			
	BRANDS*(9)		OTHERS		BRANDS*(21)		OTHERS	
	n=147	%	n=1678	%	n=1025	%	n=1579	%
CALORIES	1	0.7	6	0.4	0	0.0	0	0.0
CARBOHYDRATES	3	2.0	19	1.1	6	0.6	1	0.1
PROTEIN	12	8.1	18	1.1	6	0.6	3	0.2
FAT	14	9.5	21	1.3	6	0.6	4	0.2
SATURATED FATTY ACIDS	**	**	**	**	627	61.2	70	4.4
OLEIC	**	**	**	**	643	62.7	70	4.4
LINOLEIC	**	**	**	**	639	62.3	72	4.6
CHOLESTEROL	**	**	**	**	552	53.9	82	5.2
CALCIUM	19	12.9	105	6.3	32	3.1	16	1.0
PHOSPHORUS	20	13.6	145	8.6	267	26.1	31	2.0
SODIUM	18	12.2	269	16.0	184	18.0	67	4.2
POTASSIUM	52	35.4	296	17.6	288	28.1	65	4.1
VITAMIN C	77	52.4	484	28.8	144	14.1	35	2.2
VITAMIN A	74	50.3	285	17.0	117	11.4	32	2.0
THIAMIN	15	10.2	149	8.9	67	6.5	22	1.4
RIBOFLAVIN	13	8.8	154	9.2	57	5.6	19	1.2
PREFORMED NIACIN	19	12.9	164	9.8	74	7.2	24	1.5
IRON	18	12.2	123	7.3	42	4.1	20	1.3
	Total: n = 1825				Total: n = 2604			

* processed foods with known brand names
** nutrients not available for the majority of foods