

7.0 FILE DESCRIPTION FOR SCAN92 DATA FILES

TABLE OF CONTENTS

SECTION 7.0

	<u>Section</u>
SCAN92 Data File Format	7.1,7.2
Coding Manual	7.3

- 7.1** The SCAN92 questionnaire was developed by Drs. Gladys Block and Linda Harlan. The forms were printed by National Computer Systems, Inc. (NCS). The coding and file locations of the variables in this questionnaire were set by the DIETSYS technical support staff in conjunction with NCS personnel. Files in the following format were created when the forms were scanned by NCS. This configuration is stored in SCAN92.CFG. See page iv of the Overview Section for a list of commercial services related to the use of the SCAN92 form (purchasing, scanning, and analyzing).

If you are scanning the questionnaires yourself, it is recommended that you configure your scanner to create files in this format. However, if it is difficult to configure your scanner to create files in this exact format, you may make some format changes. If the Questionnaire Data Files created by your scanner are not in the exact format expressed in Sections 7.2 and 7.3, create a new CFG File which is compatible with the format of your Questionnaire Data Files (Section 11). The location of the Respondent ID, the location of the scanner information (header), and the characters used for Missing and Scanner Error are the specifications which may be changed. Read Sections 7.2 and 7.3 for more information on these scanner-dependent specifications.

- 7.2** The default configuration of the SCAN92 Questionnaire Data Files is described below:

1. Non-Card Format. The data for each SCAN92 questionnaire is stored in 1 line of data. That is, one line of 302 characters is used rather than the 80 character lines used in Card Format questionnaire files.
2. A 10 character Respondent ID must appear in each line. It is recommended that a unique ID be used for each questionnaire.

The location of the Respondent ID field may vary with different scanners. In the SCAN92 default configuration stored in SCAN92.CFG, the Respondent ID is in the first 10 characters of every data line. However, many scanners will automatically store "header information" as the first data segment of each line. If you are scanning the questionnaires yourself, you may either configure your scanner to write the "header information" variables to the locations specified in SCAN92.CFG (at the end of each data line); or create a new CFG File compatible with the Questionnaire Data File created by your scanner. To create a new CFG File, modify SCAN92.CFG by following the instructions in Section 11.4. You may use the **View or Modify Question List** menu item to move the header variables to the beginning of the Question List.

3. The questionnaire must be coded in a Categorical format. All frequencies (for foods, vitamins, etc.) must be coded with a 1 character format. All HHHQ Questions must be coded using their Categorical coding scheme. See Section 10.3 for more information regarding the Categorical format.
4. Missing Code. The CFG File distributed for the SCAN92 questionnaires (SCAN92.CFG) assigns the letter 'M' as the code to be used for missing data. The letter 'E' has been assigned as the scanning error code. 'E' is to be input when 2 or more choices have been shaded for a question expecting only 1 valid response.

**Variable Locations for SCAN92 Questionnaire Data Files
(NCI Version HHHQ,FULL.JAN92)**

If you wish to use characters other than 'M' and 'E' for Missing and Error, you must create a modified version of the SCAN92.CFG file. See Section 11 for details.

5. Each food frequency code must be accompanied by a code for serving size. The 1 digit code for serving size must follow the 1 digit code for food frequency. Small, medium, and large are the only valid serving sizes for the SCAN92 questionnaire.

7.3 Coding Manual. The coding manual in the following pages describes each character in a line of data in SCAN92 Questionnaire Data Files. This configuration is stored in SCAN92.CFG. If the data files you have created do not use these exact locations and coding schemes, create a new CFG File compatible with your files. Section 11 contains instructions for creating a CFG File by modifying an existing file.

Except for Missing and Error codes, you must code the HHHQ Questions on the SCAN92 questionnaire exactly as they are shown here.

**Variable Locations for SCAN92 Questionnaire Data Files
(NCI Version HHHQ,FULL.JAN92)**

COLUMN VARIABLE

1-10 RESPONDENT ID

1-9 Identification Number
10 Blank

The RESPONDENT ID is required by DIETSYS for every questionnaire. DIETSYS uses this 10 character field to identify each questionnaire in the file. Question 1, "Identification Number", is used as the Respondent ID for SCAN92 questionnaires. It is recommended that the nine digit Identification Number be stored in the first nine characters of this field and the tenth position be blank-filled.

11 SEX

1 = Male
2 = Female
M = Missing (unstated or unknown)
E = Error (multiple choices marked)

12 AGE (1 digit format coded as shown below)

<u>Age code</u>	<u>Questionnaire Text</u>	<u>Portion size set</u>
1	Less than 20	Younger
2	20-29	Younger
3	30-39	Middle Aged
4	40-49	Middle Aged
5	50-59	Middle Aged
6	60-69	Older
7	70+	Older
M	Missing (unstated)	Middle Aged
E	Error (multiple choices marked)	Middle Aged

13-18 DATE

13-14 Month (01-12)
15-16 Day (01-31)
17-18 Year (91-00)

Code an 'M' for any column left blank. Code an 'E' for any column in which 2 or more choices were shaded.

**Variable Locations for SCAN92 Questionnaire Data Files
(NCI Version HHHQ,FULL.JAN92)**

COLUMN **VARIABLE**

**Variable Locations for SCAN92 Questionnaire Data Files
(NCI Version HHHQ,FULL.JAN92)**

COLUMN VARIABLE

19-21 WEIGHT (pounds)

Code an 'M' for any column left blank. Code an 'E' for any column in which 2 or more //choices were shaded.

22-24 HEIGHT

22 Feet

0-8 = Feet

M = Missing (unstated or unknown)

E = Error (multiple choices marked)

23-24 Inches

00-11 = Inches

M = Missing (unstated or unknown)

E = Error (multiple choices marked)

Code an 'M' for any column left blank. Code an 'E' for any column in which 2 or more choices were shaded.

25-26 SMOKING

25 Smoke now?

1 = No

2 = Yes

M = Missing (unknown or unstated)

E = Error (multiple choices marked)

26 How many now?

1 = 1-5

2 = 6-14

3 = 15-24

4 = 25-34

5 = 35 or more

M = Missing (unknown or unstated)

E = Error (multiple choices marked)

**Variable Locations for SCAN92 Questionnaire Data Files
(NCI Version HHHQ,FULL.JAN92)**

<u>COLUMN</u>	<u>VARIABLE</u>
27	NUMBER OF TIMES ON WEIGHT LOSS DIET 1 = Never 2 = 1-2 times 3 = 3-5 times 4 = 6-8 times 5 = 9-11 times 6 = 12 or more times M = Missing (unknown or unstated) E = Error (multiple choices marked)
28	DO YOU TAKE VITAMINS/MINERALS REGULARLY? 1 = No 2 = Yes 3 = Yes, but not regularly M = Missing (unstated or unknown) E = Error (multiple choices marked)
29-42	VITAMIN FREQUENCIES

For each supplement coded in columns 29-42, code the two data segments as follows:

Number of Tablets:

- 1 = None
- 2 = 1-3 per week
- 3 = 4-6 per week
- 4 = 1 per day
- 5 = 2 per day
- 6 = 3 per day
- 7 = 4 per day
- 8 = 5+ per day
- M = Missing (unstated or unknown)
- E = Error (multiple choices marked)

For How Many Years?

- 1 = Less than 1 year
- 2 = 1-2 years
- 3 = 3-5 years
- 4 = 6-9 years
- 5 = 10+ years
- M = Missing (unstated or unknown)
- E = Error (multiple choices marked)

**Variable Locations for SCAN92 Questionnaire Data Files
(NCI Version HHHQ,FULL.JAN92)**

<u>COLUMN</u>	<u>VARIABLE</u>
29-42	29 Number of Tablets: Stress-tabs type
	30 How Many Years: Stress-tabs type
	31 Number of Tablets: Therapeutic,Theragran type
	32 How Many Years: Theragran type
	33 Number of Tablets: One-A-Day type
	34 How Many Years: One-A-Day type
	35 Number of Tablets: Vitamin A
	36 How Many Years: Vitamin A
	37 Number of Tablets: Vitamin E
	38 How Many Years: Vitamin E
	39 Number of Tablets: Calcium/Dolomite
	40 How Many Years: Calcium/Dolomite
	41 Number of Tablets: Vitamin C
	42 How Many Years: Vitamin C

43-44 UNITS PER TABLET

- 43 How many units (I.U.) per Vitamin E tablet*?
- 1 = 100 I.U.
 - 2 = 200 I.U.
 - 3 = 400 I.U.
 - 4 = 1000 I.U.
 - 5 = Don't know
 - M = Missing (unstated)
 - E = Error (multiple choices marked)
- 44 How many milligrams per Vitamin C tablet*?
- 1 = 100 mg
 - 2 = 250 mg
 - 3 = 500 mg
 - 4 = 1000 mg
 - 5 = Don't know
 - M = Missing (unstated)
 - E = Error (multiple choices marked)

* The number of Mgs, IUs or a-TEs corresponding to each input code is set in the DIETSYS Foods Database. See Section 12 for information if you wish to change these values.

**Variable Locations for SCAN92 Questionnaire Data Files
(NCI Version HHHQ,FULL.JAN92)**

COLUMN VARIABLE

45-50 OTHER VITAMINS: DO YOU REGULARLY TAKE PILLS CONTAINING ANY OF THESE NUTRIENTS?

Each of the supplements in columns 45-50 must be coded as follows:

- 0 = Not Marked
- 1 = Marked (selected by the respondent)

Each "Other Vitamin" must be coded in the following order:

- 45 No or don't know
- 46 Zinc
- 47 Iron
- 48 Selenium
- 49 Beta-carotene
- 50 Other not mentioned above

51-60 TYPE OF COOKING FAT

Each of the fat types in columns 51-60 must be coded as follows:

- 0 = Not Marked
- 1 = Marked (selected by the respondent)

The fats must be coded in the following order:

- 51 Don't know or don't cook
- 52 Stick margarine
- 53 Half margarine, half butter
- 54 Lard, fatback, baconfat
- 55 Butter
- 56 Low calorie margarine
- 57 Pam or no oil
- 58 Soft tub margarine
- 59 Crisco
- 60 Salad oil, cooking oil

61-69 TYPE OF FAT ADDED AT TABLE

Each of the fat types in columns 61-69 must be coded as follows:

- 0 = Not Marked
- 1 = Marked (selected by the respondent)

**Variable Locations for SCAN92 Questionnaire Data Files
(NCI Version HHHQ,FULL.JAN92)**

COLUMN VARIABLE

61-69 The fats must be coded in the following order:

- 61 Don't add fat
- 62 Stick margarine
- 63 Butter
- 64 Lard, fatback, baconfat
- 65 Soft tub margarine
- 66 Whipped butter
- 67 Low calorie margarine
- 68 Half margarine, Half margarine
- 69 Crisco

70-72 **LOW FAT FOODS**

Each of the foods in columns 70-72 must be coded as follows:

- 1 = Always low-fat
- 2 = Sometimes
- 3 = Rarely low-fat
- M = Missing (unstated or unknown)
- E = Error (multiple choices marked)

The responses must be coded in the following order:

- 70 Cheese
- 71 Ice Cream/Yogurt
- 72 Salad Dressing

73-76 **QUESTION 15**

Question 15 contains four separate HHHQ Questions which use the same coding scheme for frequency. Each must use the following codes:

- 1 = Seldom/Never
- 2 = Sometimes
- 3 = Often/Always
- M = Missing (unstated or unknown)
- E = Error (multiple choices marked)

- 73 How often do you add salt to your food?
- 74 How often do you add pepper to your food?
- 75 How often do you eat the skin on chicken?
- 76 How often do you eat the fat on meat?

**Variable Locations for SCAN92 Questionnaire Data Files
(NCI Version HHHQ,FULL.JAN92)**

<u>COLUMN</u>	<u>VARIABLE</u>
77-82	RESTAURANTS

77	Fried Chicken Restaurants
78	Burger Restaurants
79	Pizza Restaurants
80	Chinese Restaurants
81	Mexican Restaurants
82	Fried Fish Restaurants

The restaurant frequencies are expected by DIETSYS in increasing order. (The frequencies increase across the page.) The following input codes must be used:

- 1 = Never in past year
- 2 = 1-4 times in past year
- 3 = 5-11 times in past year
- 4 = 1-3 times a month
- 5 = Once a week
- 6 = 2-4 times a week
- 7 = Almost every day
- M = Missing (unstated or unknown)
- E = Error (multiple choices marked)

**Variable Locations for SCAN92 Questionnaire Data Files
(NCI Version HHHQ,FULL.JAN92)**

Two characters are used to code the frequency and serving size for each food. The 1st character is the coded frequency, the second is serving size. Extra large is NOT a valid serving size.

Foods on Pages 4-7 of SCAN92 Questionnaire			Foods on Page 8 of SCAN92 Questionnaire		
Code	Questionnaire Text	Value Used	Code	Questionnaire Text	Value Used
1	Never or <1 month	0	1	Never or <1 month	0
2	1 per month	1/month	2	1-3 per month	24/year
3	2-3 per month	28/year	3	1 per week	1/week
4	1 per week	1/week	4	2-4 per week	12/month
5	2 per week	2/week	5	5-6 per week	23/month
6	3-4 per week	13/month	6	1 per day	1/day
7	5-6 per week	22/month	7	2-3 per day	61/month
8	1 per day	1/day	8	4-5 per day	30/week
9	2+ per day	2/day	9	6+ per day	6/day
M	Missing	0	M	Missing	0
E	Error	0	E	Error	0

Coding Scheme for Serving Sizes

- 1 = Small
- 2 = Medium
- 3 = Large
- M = Missing (unstated or unknown)
- E = Error (multiple choices marked)

83-276

FOOD LIST

COLUMN

FOOD ID & NAME

83-84	001 APPLES AND APPLESAUCE, PEARS
85-86	002 BANANAS
87-88	194 RAW PEACHES
89-90	005 CANTALOUPE (IN SEASON)
91-92	006 CANTALOUPE (OUT OF SEASON)
93-94	007 WATERMELON
95-96	009 STRAWBERRIES (FRESH, IN SEAS)
97-98	010 ORANGES, TANGERINES
99-100	012 GRAPEFRUIT
101-102	011 ORANGE JUICE, GRAPEFRUIT JUICE
103-104	163 HI-C
105-106	015 OTHER FRUITS
107-108	066 BRAN AND GRANOLA CEREAL
109-110	067 FORTIFIED CEREALS

FOOD LIST (cont.)

**Variable Locations for SCAN92 Questionnaire Data Files
(NCI Version HHHQ,FULL.JAN92)**

COLUMN	FOOD ID & NAME
111-112	068 DRY CEREAL EXCLUDING FIBER,FORT
113-114	069 COOKED CEREAL
115-116	128 MILK ON CEREAL LINE ITEM
117-118	070 SUGAR
119-120	071 EGGS
121-122	072 BACON
123-124	073 SAUSAGE
125-126	016 GREEN BEANS
127-128	017 PEAS
129-130	102 CHILI WITH BEANS
131-132	018 BEANS (DRIED TYPE)
133-134	019 CORN
135-136	021 WINTER SQUASH
137-138	022 TOMATOES AND TOMATO JUICE
139-140	023 SALSA, RED CHILI SAUCE
141-142	024 BROCCOLI
143-144	025 CAULIFLOWER, BRUSSEL SPROUTS
145-146	026 SPINACH (RAW)
147-148	027 SPINACH (COOKED)
149-150	028 COLLARDS, KALE, GREENS
151-152	029 COLESLAW, CABBAGE
153-154	030 CARROTS, MIXED VEG WITH CARROTS
155-156	031 GREEN SALAD
157-158	064 MAYONNAISE, SALAD DRESSINGS
159-160	032 FRENCH FRIES, FRIED POTATOES
161-162	033 SWEET POTATOES
163-164	034 OTHER POTATOES
165-166	035 RICE
167-168	036 OTHER VEGETABLES
169-170	037 TABLE FAT ADDED --BY PROGRAM
171-172	038 HAMBURGER,BEEF BURRITO,MEATLOAF
173-174	039 BEEF (FAT UNSPECIFIED)
175-176	040 BEEF STEW, POT PIE
177-178	041 LIVER
179-180	042 PORK (FAT UNSPECIFIED)
181-182	043 FRIED CHICKEN (FAT UNSPECIFIED)
183-184	044 OTHER CHICKEN (FAT UNSPECIFIED)
185-186	045 FRIED FISH
187-188	046 TUNA
189-190	114 OYSTERS
191-192	047 SHELLFISH
193-194	048 FISH BROILED OR BAKED

**Variable Locations for SCAN92 Questionnaire Data Files
(NCI Version HHHQ,FULL.JAN92)**

FOOD LIST (cont.)

COLUMN	FOOD ID & NAME
195-196	049 SPAGHETTI
197-198	050 PIZZA
199-200	051 DISHES WITH CHEESE
201-202	052 LIVERWURST
203-204	053 HOT DOGS
205-206	054 HAM, LUNCH MEATS
207-208	055 VEGETABLE SOUP
209-210	056 SOUP EXCLUDING VEGETABLE
211-212	120 BISCUITS, MUFFINS
213-214	057 WHITE BREAD, ROLLS
215-216	058 DARK BREAD
217-218	059 CORNBREAD
219-220	060 SALTY SNACKS (CHIPS, POPCORN)
221-222	061 PEANUTS, PEANUT BUTTER
223-224	063 MARGARINE
225-226	062 BUTTER
227-228	065 GRAVY WITH DRIPPINGS
229-230	080 COTTAGE CHEESE
321-232	081 CHEESE AND CHEESE SPREAD
233-234	082 FLAVORED YOGURT, FROZEN YOGURT
235-236	074 ICE CREAM
237-238	075 DOUGHNUTS, COOKIES, CAKE
239-240	076 PUMPKIN PIE
241-242	077 PIES OTHER THAN PUMPKIN
243-244	078 CHOCOLATE CANDY
245-246	079 CANDY (NON-CHOCOLATE)
247-248	083 WHOLE MILK (BEVERAGE)
249-250	084 2 % MILK
251-252	085 SKIM MILK
253-254	086 SOFT DRINKS
255-256	088 BEER
257-258	089 WINE
259-260	090 LIQUOR
261-262	092 REGULAR COFFEE
263-264	093 TEA
265-266	169 LEMONS, LEMON JUICE
267-268	094 NONDAIRY CREAMER (POWDER)
269-270	096 CREAM
271-272	095 MILK IN COFFEE/TEA (BY PROGRAM)
273-274	097 SUGAR
275-276	099 WATER

**Variable Locations for SCAN92 Questionnaire Data Files
(NCI Version HHHQ,FULL.JAN92)**

COLUMN VARIABLE

277-280 SUMMARY QUESTIONS (QUESTION 18)

Each of the questions in columns 277-280 must be coded as follows:

1 = Less than once per week

2 = 1-2 per week

3 = 3-4 per week

4 = 5-6 per week

5 = 1 per day

6 = 1.5 per day

7 = 2 per day

8 = 3 per day

9 = 4+ per day

M = Missing (unstated or unknown)

E = Error (multiple choices marked)

The responses must be coded in the following order:

277 How often do you use fat or oil in cooking?

278 How many servings of vegetables do you eat?

279 How many servings of fruits do you eat?

280 How many servings of cold cereal do you eat?

281-302 SCANNER INFORMATION

The following fields are included in the Questionnaire Data File if the questionnaires are scanned by NCS. If you are scanning the forms yourself, these variables are optional but recommended for data management. Many scanners will automatically store these variables at the beginning of each data line. If you wish to relocate or delete these variables, you may do so by using a modified version of SCAN92.CFG (Section 11).

281-288 Booklet Number

289-292 Application Number (or "location code")

293-298 Serial Number

299-302 Batch Number

303 DIETSYS STATUS CODE

This code will only be present if the questionnaire file is created using the DIETSYS Data Entry System. It is used by DIETSYS to determine if the questionnaire has been completely entered, is verified, etc. This variable will not be present and is not applicable for data files which have been created by an optical scanner.