

5.0 DATA COLLECTION AND DATA MANAGEMENT

5.1 INTERVIEWER- VS. SELF-ADMINISTRATION

This questionnaire was designed to be capable of being self-administered, and has worked satisfactorily in that format and several validations have used the self-administered format. However, you should be aware that it probably will produce better data if administered by interview, and that the respondent burden is only slightly, not substantially, more by interview.

The improvement probably results not from factors such as probing by a nutritionally sophisticated interviewer, but rather simply because an interviewer is more careful to code the responses correctly. Thus, the interview does not require a nutritionist, but simply someone to read the questions to the respondent and record the answers. The computer-assisted interview in this package provides the further advantage of detecting keying errors and questionable responses on the spot.

The improvement resulting from using an interviewer is fairly substantial, probably at least 10 correlation points (e.g., improvement from $r=0.50$ to $r=0.60$), and perhaps as much as two or three times that in some situations. Studies planning to consider misclassification error in their sample size calculations (see, for example, Walker and Blettner, 1985) will note that an improvement from $r=0.50$ to $r=0.60$ would reduce the required sample size by about 30 percent. That is, getting better data, as through the use of an interviewer, results in a gain in power. Consider that this may offset the additional cost of interviewing.

Another point to consider in weighing cost is the process required with each method. Self-administered versions have to be introduced, checked and possibly queried on return, coded, keyed and edited. For scannable versions, coding and keying is replaced by scanning. An interviewer has to spend approximately 45 minutes on each respondent (for full version, including both non-diet and diet sections). Note that interviewer-administration using the computer-assisted interview program eliminates the need for coding, keying, range-editing and querying, and produces cleaned data on file at the conclusion of the interview. Thus, the cost of interviewer-administration, using the computer-assisted interview, would be limited to the interviewer time.

If you do use the self-administered format, you must instruct the respondent, you must check it for errors and omissions when it is returned and query the respondent about them, you must key-and-verify (double-key) and you must run edit checks on the file after keying, and consider requerying. See Appendix A, Assisting Respondents to Self-Administer the Questionnaire.

5.2 INTERVIEWING AND SELF-ADMINISTRATION SUGGESTIONS

Although the questionnaires may be self-administered, it is clear that it may NOT be very successful to simply mail it to respondents without other contact. There is likely to be a higher proportion of errors and poorer correlations with the truth. This is especially true for the non-categorical formats in FULL87 and BRIEF87. Some personal contact, either in person or by phone, is highly desirable both to make sure that respondents know how to fill it out, and to enroll them into the goals of the study and the need for care and attention. See Appendix A in this section for instructions for introducing the self-administered questionnaire to respondents. In our hands this personal contact need not be very long, literally only 2-5

minutes; but it is probably necessary. We further suggest that this be reinforced by providing written pointers (see Figure 5-1) along with the questionnaire. In addition, we strongly recommend that, if possible, an interviewer go over the questionnaire when the respondent returns it, to clarify omissions, etc. This too need only take 2-5 minutes, except in the infrequent cases in which the whole questionnaire is unsatisfactory. Range and reasonableness edits (Section 13) must ALWAYS be performed.

DO NOT omit the portion size section. Self- or telephone respondents seem to do fine, and accuracy is considerably poorer without it. You need not even read them what a medium portion is in all cases; you should read it for the unitary items, however (2 pork chops, 2 pats butter, etc.). Encourage self-respondents not to simply check all medium. (A high proportion of mediums is okay, but not all mediums.)

Emphasize completeness. They should check "Never", not simply skip foods they rarely or never eat. In the non-categorical FULL87 or BRIEF87 versions, they should not simply place check marks in "day, week", etc; such records are worthless.

Emphasize care and attention. They should be careful which column they're in -- carelessness here can cause serious and often undetectable errors. Discourage them from fixating on a single column -- e.g., translating all responses into "per year". Discourage them from using once per unit time for a lot of the foods (probably substantially more than 70 percent will produce poorer nutrient estimates.)

For self-respondents, show them that there are two pages of foods. If possible, you might even suggest that they take a break after the first page and come back to it later. Some people get careless on the second page, with poor results. This doesn't seem to be a problem with interview respondents.

Specific handouts on self-administration and interviewer administration of the food frequency part of the questionnaire have been developed and are provided in Appendices A and B in this section. These have proved very useful in orienting respondents and training interviewers.

Some investigators have found that the interview goes faster and more smoothly if simple flash-cards are used.

A training video for interviewers is available to borrow, on request. See Section 2.3 for more information.

FIGURE 5-1

SOME POINTS TO WATCH FOR IN FILLING OUT THE HEALTH HABITS AND HISTORY QUESTIONNAIRE*

THANK YOU for filling this out. It provides valuable information about your health habits, and it will also provide a good estimate of your dietary intake. There are instructions about filling out the diet section in the booklet itself. However, here are a few pointers about how to fill it out, or about items which some people have found confusing.

IN THE FOOD SECTION:

1. WRITE NUMBERS in the boxes to indicate how many times per day, week or month you eat a food.
2. DON'T SKIP items. If you rarely or never eat a food, check "Rarely/Never".
3. BE CAREFUL about which column you put your answer in. It will make a big difference in the calculations if you check "Hamburgers once a day" when you mean "Hamburgers once a week".
4. NOTICE that there are three kinds of cereals. Be careful that you don't triple-count here, and wind up with cereal 15 times a week when you really mean cereal 5 times a week.
5. Keep this in mind also for the three kinds of bread, and three kinds of milk.
6. NOTICE that a medium serving of eggs is stated as two eggs. If you normally only have one egg, check "small".

The food list is two pages long. If you get tired at the end of the first page, you might want to put it aside for a while, and come back to it a little later. (But don't forget to come back to it!)

Thanks again.

* May be inserted, loose, in the questionnaire. These instructions apply only to the non-categorical FULL87 or BRIEF87 versions.

5.3 CODING RULES AND SUGGESTIONS

Use the missing code to fill all fields indicated on the questionnaire (except alphabetic fields). That is, do not leave missing fields blank. Blanks may be used in alphabetic fields (Name, Address, etc.).

In the diet section:

- A. If respondent skipped a food altogether code as skipped, not as never. That is, do not impute "never eat" for that food. (A few skips are fairly benign. If many foods are skipped, however, you are left with uncertainty about whether the respondent really never eats those foods, or whether it is a poorly executed questionnaire. Prevent this by careful instruction beforehand that the respondent should check "rarely/never", not simply skip the food. If there are nevertheless a lot of omissions, you should consider querying).
- B. If respondent did not fill in the portion size for a food, do not code as medium, but as skipped. The computer program will impute medium, but will count the number of times this had to be done. (If portion sizes were omitted for only a few foods, this is benign. However, if most or all portion sizes were omitted, considerable precision is lost. Avoid this by careful instruction beforehand, or query afterwards.)

Seasonality Issues:

- C. A few foods in the food list specify "in season". What should be coded is the frequency during the three-month-or-so season. Do not average over the whole year for those foods. The computer program does that by using that food's seasonality factor which is stored in the DIETSYS foods database.
- D. All the rest of the foods are assumed by the computer program to be year-long averages. However, occasionally respondents will write in "in season", e.g., corn "in season". It is appropriate to use that volunteered information in the coding, so DO average those over the whole year. Thus, if the respondent indicates "corn once a week in season", divide the reported frequency by 4 (3 of 12 months). A simple way to do this if s/he responded "per week" is to code "per month".

Fat-type Questions

Two questions ask the type of fat usually used in cooking or on vegetables. Two coding spaces are allowed, so you can code up to two answers. If respondent checked three, our rule of thumb is to code one saturated and one unsaturated fat.

Coding of non-categorical questionnaires:

- A. If respondent placed check-marks in the "how often" columns ("check" times per week, etc.), DO NOT code as "01" times, but as "99". The program will impute an 01, but will count the number of times this had to be done. (Questionnaires in which the respondent simply placed check-marks are virtually worthless. They should be prevented by careful instruction beforehand, and corrected by query or re-interview when they do occur.)

- B. If respondent indicated his/her frequency of consumption as once per week, month, etc. for most or all of the foods, such questionnaires are also virtually worthless. Prevent or query. This does not apply, of course, to records in which the respondent did fill it out properly with 2s and 3s as well as some 1s. Our rule of thumb is that if 70 percent of foods s/he said s/he sometimes eats were checked as 01 per unit time, it is a problem and should be queried.

5.4 DATA MANAGEMENT RECOMMENDATIONS

Coding should be checked routinely by a supervisor. Consider routinely recoding a sample (e.g., 1 in 10) and comparing to detect coding errors. If the data were collected by the DIETSYS Interactive Interview, see suggestions regarding tape recording and other quality control measures, in Section 13.12.

Double-key ("key and verify"), preferably by two different keyers and correct discrepancies. (The same keyer is likely to misread a 7 as a 9 both times, for example.) Neither range edits nor the internal consistency edit described in Section 14 will detect keying errors such as `20' times per month instead of `02' times, or `032', three times per week instead of `023', twice a month. Nor can it detect frameshift errors, such as `223' (22 times per month) instead of `0223' (twice a week, large). Computerized data entry is not an adequate substitute for double-keying; no data entry program can be written which would detect errors such as those mentioned above, since all of them are legitimate and reasonable answers. Only double-keying will detect such errors. This is an essential component of adequate data management. Furthermore, it is not substantially more expensive. Commercial keypunch firms are also available at affordable rates.

Perform DIETSYS Edit Checking BEFORE performing nutrient analysis (see Section 5.5). This is essential to adequate data management. This will help you to identify incorrect coding and values which are inconsistent, or which suggest erroneous responses. Correct, query or disqualify. There is little point in including in diet/disease analyses people who simply did it wrong or were obviously careless. Power is reduced by including them. (The Edit Checking system is not a substitute for double-keying. It can detect "5 times per day" when perhaps "5 times per week" was meant, but cannot detect the reverse. Nor can it detect "5 times per week" when "5 times per month" was meant.)

Coding, keying and analysis without these quality controls constitutes use of this material contrary to good management practices and contrary to the recommendations and wishes of the authors. Unless double-keying, range edits, DIETSYS Edit Checking and correction of errors are performed, users are requested not to reference NCI or Gladys Block in connection with any results.

5.5 HHHQ INTERNAL CONSISTENCY ERROR CHECKS

Section 13 of this document contains the documentation for the DIETSYS Edit Checking feature. This system performs range edits for diet-related questions on the questionnaire, and checks for internal consistency of age and dates and for missing information in a number of questions. Of particular importance, it checks for the reasonableness of the reported frequency of each food, and produces a printout identifying individuals with unusual responses, so that they may be queried or handled in some other way. It will also, optionally, produce an edited output file with obvious errors "fixed". This has been shown to result in improved correlations with reference data. Frequency of various error flags and counters are summarized and frequency distributions should be examined using software such as SAS or SPSS.

5.6 ANALYTIC SUGGESTIONS

The first step after running the DIETSYS Edit Checking system should be to run distributions (Proc Freq) of each of the error variables created by the Edit Check and scrutinize records at the extremes of the distributions. Correct coding/keying errors, and query respondents if possible. Use the "FIX" option Edit checking to correct some obvious errors, and run nutrient analyses on the resulting file. Records identified as "Error" or unreliable should be queried, discarded, or analyzed separately.

It would also be a good idea to run distributions on all of the nutrient estimates and examine the records of extreme outliers. DO NOT automatically disqualify them, as these are often the very ones who do in fact eat a lot or a little; but consider querying peculiar responses.

After identifying the records which will be analyzed, another early step should be to determine, for each variable, what transformation will best normalize the data. Diet data are usually very non-normal, and it is usually inappropriate to use the standard parametric statistics (Pearson correlations, t-tests, analysis of variance, etc.) until the data have been transformed to improve their normality. For certain analyses, regression or analysis of variance, the error (residual) structures should be approximately normal, homogeneous and most importantly independent. Data which are skewed right (tail on the right) may be made more normal by the following: $\sqrt{\text{Variable}}$, $\ln(\text{Variable})$, $-1/\sqrt{\text{Variable}}$, $-1/\text{Variable}$. The latter two are rarely needed unless the data are extremely skewed. Data which are skewed left (rare but not unheard of in diet data) may be made more normal by Variable^{**2} , Variable^{**3} , etc. Thus, you could run PROC UNIVARIATE PLOT NORMAL for, e.g., DIETCAL, $\ln(\text{DIETCAL})$, $\sqrt{\text{DIETCAL}}$, and select the one which is most normally distributed and with the least skewness. For certain analyses this procedure can be applied to the residuals (errors) obtained from the relevant models. This should be done for all the numeric variables which will be included in parametric analyses, and should be done AFTER dropping any outliers or careless responders which you intend to drop.

Diet variables tend to correlate better with serum, physiologic or health factors when expressed per kilogram of body weight; e.g., a small person may eat less in absolute terms than a large person, but his/her physiologic needs are probably less as well. Thus, in addition to examining absolute values, diet variables should always be also expressed per kilogram of body weight. Thus $\text{CALKG} = \text{DIETCAL}/(\text{WEIGHT}*.454)$.

An extensive discussion of analytic and statistical issues relevant to dietary data may be found in Hartman and Block, 1991, and Willett, 1989.

The analysis and use of vitamin supplement data is a potentially serious source of error. See Block et al. 1994 for a discussion. Vitamin supplement use and years of use should always be collected. The important analytic variable is the sum of food and supplement sources of a nutrient. But the most informative subsample for that analysis is those who have been taking supplements for several years. Inclusion of those who started taking them more recently will include persons who had already begun to feel ill or who had comorbid conditions, and will falsely weaken or even obscure any associations. Furthermore, analyses of relationships between health outcomes and "nutrients exclusive of vitamins supplements" should always be done only in the subset of persons who do not take vitamins. For example, if you are studying the role of fruits and vegetables in disease prevention, you must do so in non-users of supplements; if you include supplement users, some people with low fruit/vegetable intake will have high tissue levels of nutrients because of supplement use, and your quantiles will be meaningless. Similarly, if you want to study whether high intake from foods alone is enough to provide reduced risk or whether supplement use is needed, you must look at protection afforded by food sources just in non-users of supplements.

APPENDIX A

SELF-ADMINISTRATION

ASSISTING RESPONDENTS TO SELF-ADMINISTER THE QUESTIONNAIRE

This introduction is designed to assist you in helping respondents to self-administer the food frequency part of the interview. Some specific instructions and guidelines are discussed below. First, however, it may be useful for you to understand a bit about the development and rationale of the food frequency section.

The questionnaire was developed based on information from large national surveys, and has been used with thousands of respondents, and shown to produce very useful results. The food items were chosen so as to give us good estimates of a wide range of nutrients, including fat, vitamin C and others. They don't include all of the possible foods a person could have eaten, of course, but they are the most important foods in most people's diets.

The time frame that it covers is "the past year or so". This is deliberately a little vague, because it is not expected that anyone could remember exactly what they ate during exactly the past year. The idea is just to get a usual pattern -- their current diet at this point in their life. Some people raise the objection, "Oh, I can't even remember what I ate yesterday; how could anyone answer what they ate in the past year?" If respondents have this concern, it's important to make clear to them that the idea is not to remember, but to think about their usual pattern of frequency. For example, they don't have to remember how many times they had eggs in the past year. Instead, what they can tell you with reasonable accuracy is, "Oh, I have eggs about twice a week."

The portion size part of the questionnaire was also worked out based on national diet data, and there too what seems quite imprecise actually does the job quite well. For most of the items, all we really need is whether the person's usual portion is "small, medium or large", because very few people really know what a half cup is, or what six ounces is. But they can tell you, "Oh, I don't really like that very much, so when my wife cooks it I only have a small portion." What we will actually use in the calculations is based on the actual portion sizes of these foods in a 10,000-person sample.

Now some of you may be thinking, "But wait a minute, what one person thinks is medium is different from what another person thinks is medium." That's exactly right, and the calculations take this into account. For example, the medium portion amount that is used for a 70-year-old woman is different from the medium portion amount that is used for a 20-year-old man. So, that "small, medium, large" actually gives us considerably more precision than you might expect.

With this questionnaire, it will be possible to classify people, for example, as high consumers of fat, medium, or low consumers of fat. That kind of information (on other nutrients, as well as fat) will permit very useful studies on the relationship between diet and health.

SPECIFIC INSTRUCTIONS FOR INTRODUCING THE SELF-ADMINISTERED DIET QUESTIONNAIRE TO RESPONDENTS

1. The food frequency questionnaire was designed to be capable of being self-administered, and has performed very satisfactorily in that format, as evidenced by good correlations with both serum values and food record data. However, a brief introduction for the respondent is extremely important to obtaining valid results. It need not be long (2-5 minutes), but it is important to inform respondents about the importance of their contribution and the necessity to do it carefully, and make sure they know how to fill it out. This should also be reinforced by providing written pointers (Figure 1) inserted into the questionnaire.
2. Show respondents how long the questionnaire is, and tell them about how long it will take (25 minutes, for the full-length questionnaire, 17 for the brief version, on the average). You might even suggest that they take a break after the first page and come back to it later. Some people get careless on the second page, with poor results.
3. Tell them it is about the food they usually eat, over the period of the past year. If they object that they cannot "remember", explain that they need not remember, but give a usual pattern ("I eat eggs about twice a week").
4. Point out that they should check off a small, medium or large portion size for each food. Respondents who check all "mediums" probably are not thinking very hard about their answers. On the other hand, if a respondent checks all "smalls", the nutrient estimates will be unreasonably low. They should just check "small" or "large" for those foods that they usually eat substantially smaller or larger portions than other people of their age and sex.

As an approximation, a "small" portion is about 1/2 the stated "medium", or less. A "large" is about 1.5 times as much, or more.

Don't worry if there are a lot of "mediums". In many questionnaires up to 70% of the foods may be "medium". It is all "mediums" (or all "smalls") which is a problem.

5. Emphasize completeness. They should check "Rarely/Never", not simply skip foods they rarely or never eat.
6. Emphasize care and attention, especially in the section on frequency of foods. They should be careful which column they put their answers in. It can make a big difference in the nutrient estimate if they put their answer in the wrong column.
7. Point out eggs, and explain that if the respondent usually eats only one egg, s/he would check "small".
8. Point out the three lines for cold cereal. Respondents should be careful not to double or triple-count here. If they usually mix two kinds of cereal together, they should give each one half the total frequency. They should also be careful not to triple-count the three kinds of breads, and the three kinds of milk.

9. Point out the milk items, and tell them that they should not include milk on cereal for those items. Those items refer to milk consumed as a beverage. (Milk on cereal is added by the program, unless a specific line for it is included by the investigator.)

10. For the type of questionnaire in which the food frequency part has five columns labeled Day, Week, Month, Year, Never: Explain that they are to WRITE NUMBERS in columns representing "per week", "per month", etc., indicating the number of times per day, week, month or year they usually eat that food. They should not simply place checkmarks in the "week" or "month" box. Such records are worthless and must be prevented by careful instruction beforehand. They should also not simply write `1's for each food (1/day, 1/week, etc.). Such records are the same as placing checkmarks in the boxes, and are also worthless. (Obviously, some `1's are acceptable; it is all `1's which is worthless.)

SOME OTHER QUESTIONS THAT MAY COME UP

1. A few items ask how often they eat a food "in season". They should respond with the number of times per day, week or month that they eat it, just during the 3-4 months when it is in season. (The program will then average that frequency over the whole year, but they should not.)

For example, if they eat fresh peaches five times a week when they are in season (regardless of how often they eat them during the rest of the year), they should put a `5' in the `week' column for "Peaches, apricots, fresh, in season".

For those four items for which "in season" is printed on the questionnaire (peaches, cantaloupe, watermelon, strawberries), and for those items only, the program assumes that the response is for the 3-4 month season only. For all other items on the questionnaire, the program assumes that the response is for a year-long average or pattern.

2. Two questions ask the type of fat usually used in cooking or on vegetables. They may check either one or two types of fat. (If they check two, the program will divide up the nutrient values 50-50.) If they check three, we handle that in coding by picking one saturated and one unsaturated fat.
3. Two questions at the end ask broadly, "How many vegetables" (or fruits) "do you usually eat per day or per week"? The purpose here is not to get at variability (different kinds of vegetables), but at how often they serve and eat vegetables. So if they never have a vegetable at lunch but always have, at dinner, a potato and a green vegetable, their answer to "Not counting potatoes and salad, how many vegetables do you eat per day or per week"? would be "1 per day" or "7 per week".

This information can be used by the program to adjust their responses to the vegetable questions in the questionnaire. Many people, when they go through that list of vegetables and fruits, wind up with more vegetables per day than they really eat. This question permits the program to adjust that down to more realistic levels, while preserving the proper distribution of vegetables which they prefer.

4. In the vitamin pill question, some people think that when we ask for quantities of vitamins A, C, E, B vitamins and calcium, we want them to go and look at their multiple vitamins and report those quantities. This is not correct. The program will make certain assumptions about what is in their multiple vitamins, depending on which type they use. The values which are wanted in the "mg/pill" are what is in the pills if they take those vitamins separately, not in multiple vitamins. So if they take vitamin C, how many mg in each of the pills they take? For vitamin C, most users will know. For other types of vitamins, they may not, and I don't think we should worry them too much about it. The program will use reasonable default values, if they don't know.

CHECKING THE QUESTIONNAIRE WHEN IT IS RETURNED

When the questionnaire is returned, spend a few minutes (again, usually only 2-5) checking over the questionnaire while the study participant is still there.

1. Make sure the ID number is correct.
2. Check for omissions -- skipped foods, missing information in the vitamin pill section, etc.
3. Check for unlikely frequencies such as liver twice a day. Confirm that the respondent really eats it that frequently.
4. Check the three dry cereals, to see if the respondents triple-counted. If she eats cereal seven times a week, those three items should not add up to more than seven. Check the three milks and the three breads, for the same possible error.

Don't worry if it adds up to a little more than seven. It's hard for the respondent to be exact, and it doesn't make much difference, anyway. What we're looking for here is actual double-counting -- 21/week because "sometimes I eat one kind, sometimes another kind"; or 14/week because "I mix half All-Bran and half Special K".

However, it is not necessary to try to make the rest of the foods add up to something that seems "right", like seven dinners a week. In many cases the foods will not add up to what we consider a "typical" diet, and you need not spend time trying to make them do so.

5. Occasionally respondents will write in "in season" beside one of the fruits or vegetables. In this case, find out whether their reported frequency for that item refers to just the 3-4 month season. For example, if they wrote "Corn in season 2/week", first confirm that they mean they eat it twice a week when it is in season. If that is the case, you should translate their answer into a year-round frequency.

"Twice a week", if left that way, will be assumed by the program to mean 2/week for 52 weeks a year, and will seriously overcount that item. (Don't worry about asking them how long they think a season is. Just assume three months. That will be precise enough for our purposes.) If you do make such changes and corrections, initial the line you have just changed.

The above procedure should only be done for foods where the respondent has written in "in season"; and only for those items which don't already have "in season" printed on the questionnaire.

6. Check the vitamin pill question for omissions, and to see whether the participant has misunderstood and put information in the "Other vitamins" section that really pertains to multiple vitamins.

There are two clues which help you to spot this. First, the frequency given for the specific vitamins is the same as the frequency given for how often s/he takes multiple vitamins. Second, the mg/IUs per pill is not only filled in, but is filled in with the amounts typically used in multiple vitamins -- usually small amounts such as 60 mg vitamin C per pill.

If you see this, ask the participant if s/he takes, e.g., vitamin C as a separate pill, or is s/he referring to what is in her multiple vitamins. If the latter, cross out the information about the specific vitamin types, leaving only the information about the multiple vitamins. Initial the change.

9. For the type of questionnaire in which the food frequency part has five columns labeled Day, Week, Month, Year, Never:
 - a. Check for checkmarks (instead of numbers) in the "How often" section of the food frequency questionnaire.
 - b. Check for "unreasonable" frequencies which suggest that the respondent may have been off by a column -- for example, orange juice 7/day, cole slaw 3/day.

These errors can happen in any column, of course, but they're hard to spot unless the shift was into the "day" column. Just scan down that column for frequencies which suggest there may have been a column shift.

- c. If the questionnaire has mostly checkmarks instead of numbers, or mostly '1's (1/day, 1/wk, etc.), ask the participant if s/he would go over it and fill in the information a little more carefully.

Don't worry if there are a fair number of '1's, mixed in with '2's and '3's, etc. That may be reasonably accurate. The poor estimates arise when people put almost exclusively '1's.

APPENDIX B

INTERVIEWER ADMINISTRATION

INTRODUCTION TO INTERVIEWER ADMINISTRATION OF THE DIET QUESTIONNAIRE

This introduction is designed to assist you in administering the food frequency part of the interview. Some actual scenarios are discussed below. First, however, it may be useful for you to understand a bit about the development and rationale of the food frequency section.

The questionnaire was developed based on information from large national surveys, and has been used with well over 1,000 respondents, and shown to produce very useful results. The food items were chosen so as to give us good estimates of a wide range of nutrients, including fat, vitamin C and others. They don't include all of the possible foods a person could have eaten, of course, but they are the most important foods in most people's diets.

The time frame that it covers is "the past year or so". This is deliberately a little vague, because it is not expected that anyone could remember exactly what they ate during the past year. The idea is just to get a usual pattern -- their current diet at this point in their life. Some people raise the objection, "Oh, I can't even remember what I ate yesterday; how could anyone answer what they ate in the past year?" If respondents have this concern, it's important to make clear to them that the idea is not to remember, but to think about their usual pattern of frequency. For example, they don't have to remember how many times they had eggs in the past year. Instead, what they can tell you with reasonable accuracy is, "Oh, I have eggs about twice a week".

The portion size part of the questionnaire was also worked out based on national diet data, and there too what seems quite imprecise actually does the job quite well. For most of the items, all we really need is whether the person's usual portion is "small, medium or large". If respondents ask you what you mean by medium, it's fine to tell them. But we don't usually read to them what the "medium" portion is for most foods, because very few people really know what a half cup is, or what six ounces are. But they can tell you, "Oh, I don't really like that very much, so when I have it at all, I only have a small portion". What we will actually use in the calculations is based on what a 10,000-person sample chose as their portion sizes for all these foods. (A few medium portions represent units of items, such as two pork chops; for those foods, you should read the medium portion to the respondent. In some cases, such as number of eggs or number of hot dogs, you could also simply ask "How many eggs?", and then check off small, medium, or large as appropriate.)

Now some of you may be thinking, "But wait a minute; what one person thinks is medium is different from what another person thinks is medium." That's exactly right, and the calculations take this into account. For example, the medium portion amount that is used for a 70-year-old woman is different from the medium portion amount that is used for a 20-year-old man. So, that "small, medium, large" actually gives us considerably more precision than you might expect.

It is important, however, that you ask "small, medium or large?" for each food. Don't simply assume medium unless s/he volunteers something else, as that gives poorer estimates.

With this questionnaire, it will be possible to classify people, for example, as high consumers of fat, medium, or low consumers of fat. That kind of information (on other nutrients, as well as fat) will permit very useful studies on the relationship between diet and health.

Now let's turn to some scenarios, some difficult respondents or situations, and how to handle them.

SOME PROBLEM SCENARIOS

1. Respondent volunteers seasonal information. This will not happen terribly often, and you should NOT probe for it. However, when the respondent volunteers the information it is appropriate to use it. Do not probe for how long the respondent thinks a season is, but simply assume three months, 1/4 of a year.

Q: (How often do you eat) corn?

A: Well, most of the time I have it about once a month. But in the summer I eat fresh corn about once a week.

Q: Once a week in season, once a month the rest of the year?

A: That's right.

(Interviewer should write down the response verbatim. Then, either during the interview if there is time, or during field editing after the interview, calculate how many times per year that averages out to. In this example, it would be (1/week)x(4 weeks in a month)x(3 month season)

$$\begin{aligned} &+ (1/\text{month})\times(9 \text{ remaining months in the year}) \\ &= 1\times 4\times 3 + 1\times 9 \\ &= 21 \text{ times per year.} \end{aligned}$$

2. Respondent gives a frequency which seems unreasonable. Obviously, you should not doubt the respondent repeatedly or probe for this regularly. However, it is appropriate to use some common sense if the respondent appears to have misspoken.

Q: (How often do you eat) liver?

A: Four times a day.

Q: Four times a day?

A: Oh, I mean four times a week.

3. Respondent gives a frequency "per year", and it is a high number per year, such as "150 times/year". Such answers are likely to be inaccurate, so ask the respondent to express it per week or per month. Smaller times "per year", such as "15 times/year", need not be questioned.

Q: (How often do you eat) hamburgers, cheeseburgers or meatloaf?

A: 150 times a year.

Q: Could you tell me in terms of number of times per week or per month?

A: Well, I guess it would be about twice a week. (Here, s/he realizes it isn't really three times a week, which is what 150/year would be.)

or

A: Well, I eat it more than twice a week -- maybe 10 times a month.

4. Erroneous frequency of the three dry cereals (or of the three breads, or the three milks). Do not attempt to make other groups of foods add up to something you think is reasonable; however, for the three dry cereals (or the three breads or the three milks), some respondents may triple-count their frequency. Avoid this first by alerting the respondent that there are several types of this food which will be asked about; and be alert if s/he appears to have nevertheless triple-counted.

Q: (How often do you eat) whole milk or drinks made with whole milk, not including on cereal? (I'm going to ask you about 2%, 1% and skim milk separately).

A: Three times a day.

Q: Small, medium or large?

A: Medium.

Q: 2% milk, not including on cereal?

A: Three times a day.

Q: Skim milk, 1% milk or buttermilk, not including on cereal?

A: Three times a day.

Q: Let's see, you've told me you drink each of those different kinds of milk three times a day. That would be milk nine times a day. Is that right?

A: Oh, no, it's just that sometimes I have one kind and sometimes another kind.

Q: Oh, I see. Well, let's go back and find out how often you drink each of those different kinds. How often do you drink whole milk or beverages made with whole milk, not counting on cereal?

A: Well, whole milk I guess I only have about five times a week.

Q: What about 2% milk, not counting on cereal?

A: That would be most of the time, about twice a day.

Q: And what about 1% or skim, not counting on cereal?

A: That would be pretty rarely, maybe twice a week.

(Milk on cereal, by the way, will be added by the program automatically based on the number of times the respondent reports eating cereal and on the types of milk s/he drinks, unless the investigator has added a specific line for it.)