Combination of probing types. The most effective interviews may consist of a combination of scripted and spontaneous probes described above, rather than either type by itself. By way of analogy, a cognitive interview is similar to a session with a clinical psychologist; the "therapist" has certain guiding principles, and perhaps specific questions or comments, to apply during a session with the patient. However, much of the interchange emerges spontaneously during the course of therapy. The clinical session may be approached in ways similar to other sessions, and be somewhat "scripted", but every interview is different, entails its own developmental sequence, and makes a unique contribution as far as the "diagnosis" of problems.

For the remainder of this manual, probing rather than the strict think-aloud procedure is emphasized. However, practitioners of cognitive interviewing techniques often mix these techniques into the same interview. In fact, procedural flexibility, as opposed to rigid adherence to one dogmatic approach, is often viewed as one of the most attractive features of the cognitive interviewing approach.

4. EXAMPLES FROM COGNITIVE INTERVIEWING

In order to better illustrate the above discussion of cognitive techniques, and the use of verbal probing in particular, a list of examples of survey questions that have been tested using verbal probing techniques is presented below[4]. Each example consists of:

1) The question in its original form.

2) A list of several probes that would be appropriate to use in testing that question.

3) A short description of the problems found, through cognitive testing of these questions, using probes of the types suggested. Each of the examples is classed generally according to whether the problems found are representative of the cognitive categories defined earlier. However, some questions may have more than one type of problem, and in some cases it is arguable what class of problem is really being reflected. This type of classification ambiguity may not be problematic, to the extent that the nature of the specific problem (and perhaps its resolution) is clear.

4) Finally, a suggested resolution to the problem is presented, based on the testing results.

---

[4]These questions were developed during the time the author worked in the Questionnaire Design Research Laboratory at the National Center for Health Statistics, CDC, in Hyattsville, MD (see Willis, 1994). The tested questions were mainly intended for use in the National Health Interview Survey (NHIS), a household-interview-based health survey conducted annually by NCHS.

EXAMPLE 1:

1) Original form of survey question:

Has anyone in the household ever received vocational rehabilitation services from-

... The State Vocational Rehabilitation program?

... another vocational rehabilitation program?

---

2) Probes:

    a) **Can you repeat the question in your own words?**
     (To test how well the subject comprehends the question.)

    b) **What, to you, is a "vocational rehabilitation program"?**
     (To test comprehension of a particular term.)

    c) **How sure are you that (person) got this type of service?**
     (To determine the subject's ability to recall information confidently.)

3) Results:

Comprehension problems: Subjects found it difficult to understand the question, because of its length and technical nature. Further, the meaning of "vocational rehabilitation" was not at all clear; some subjects thought this just meant any type of physical therapy. Because of the comprehension problems in the original form, we suggested the following change:

4) Suggested revision:

---

*Has anyone in the household ever received job
rehabilitation services?*

If YES, ask WHO, and:

*Was (person's) rehabilitation from the state, or from
another job rehabilitation program?*

    Note: The question is "decomposed", or divided up, to make it easier to understand. The term "vocational" is also changed to the more understandable form "job".

EXAMPLE 2:

1) Original form of question:

How long has (name) used the (cane, wheelchair, walker...)?

---

2) Probes:

a) **How did you get the answer of (x) years?**
(To determine the overall cognitive strategy used.)

b) **When did (x) first use the (device)?**
(To test comprehension/interpretation of the question.)

c) **How well do you remember this?**
(To test recall of the relevant information.)

3) Results:

It was found that for target individuals whose use was intermittent over a long period of time, the question was interpreted in two distinctly different ways:

1) "How long has it been since (person) first used the (device)? For example, the subject may say: "since 1960, so about 30 years".

2) "For how long, overall, has (person) actually used the device since first having it? The subject counts up periods of use within a longer time- for example: "For two five-year periods since 1960, so 10 years".

Note that the problem identified can be considered a type of "comprehension" problem, but doesn't involve a failure of comprehension of a key term, as did the last example. Rather, subjects simply have alternate, but reasonable, interpretations of the question intent.

4) Suggested revision:

This required consultation with the client, in order to clarify the objective of the question. It became clear that the desired expression was:

How long ago did (person) first use a (device)?

EXAMPLE 3:

1) Original form:

About how many miles from here is the home (child) lived in before (he/she) moved to this home?

(THE RESPONSE CATEGORIES ARE PRINTED ON THE QUESTIONNAIRE, BUT NOT READ):

[   ] less than 1 mile
[   ] 1-50 miles
[   ] 50+ miles

---

2) Probes:

    a) **How sure are you of your answer?**
       (to determine overall level of confidence)

    b) **How hard was this to answer?**
       (to determine level of difficulty, and likelihood of
        estimation/guessing)

3) Results:

No one had difficulty understanding the question as posed. However, some subjects needed to think for a fairly long time before giving an answer. Further, some subjects struggled needlessly with the level of specificity they thought was required (for example, deciding whether the distance was closer to 20 or to 25 miles, when this information was ultimately irrelevant, as the interviewer would mark "1-50 miles" in either case).

The problem can be described as one involving a difficult recall task, as opposed to comprehension. A rephrasing of the question that incorporated response alternatives was necessary to make clear to subjects the degree of precision that was necessary in their answer.

4) Suggested revision:

About how far from here is the home _____ lived in before (he/she) moved to this home- less than a mile, 1 to 50 miles, or more than 50 miles?

EXAMPLE 4:

1) Original form:

_____

We are interested in your lifetime exercise patterns.
First, when you were 14 to 19 years old:

How many hours a week of brisk walking did you do?

How many hours a week of vigorous exercise such as
running, cycling, swimming, or aerobics did you do?

How many hours a week of activities that required you
to be on your feet (excluding running or walking) such
as dancing, hiking, ... did you do?
_____

2) Probes:

    a) **Was this hard or easy to answer?**
          (to determine comprehension, and overall ability to recall)

    b) **How do you remember this?**
          (to study recall strategy)

    c) **How sure are you of your answer?**
          (confidence probe)

    d) **What, to you, is "vigorous exercise?"**
          (comprehension/interpretation of a specific term)

3) Results:

Subjects found it very difficult to remember back to the time period specified, at the required level of detail. In fact, it seemed that some subjects really could not even answer this with respect to their current behavior, let alone their behavior many years ago. Recall of information (assuming it was ever "learned" in the first place) seemed to be the dominant problem.

As for the previous example, the cognitive interviewing staff needed to confer with the sponsor/client to clarify question objectives. We were able to determine that use of a broad scale of level of activity, comparing past and present behavior, would satisfy the data objectives:

4) Suggested revision:

We are interested in your lifetime exercise patterns. When you were 14 to 19 years old, were you more active than you are now, less active than now, or about as active as now?

EXAMPLE 5:

1) Original version:

During a typical work day at your job as an (occupation) for (employer), how much time do you spend doing strenuous physical activities such as lifting, pushing, or pulling?
[CATEGORIES ARE CONTAINED ON A CARD SHOWN TO RESPONDENT]

___ *None*
___ *Less than 1 hour*
___ *1-4 hours*
___ *4+ hours*

---

2) Probes:

a) **What type of work do you do? Describe a typical workday.**
b) **How did you arrive at the answer of X hours?**

3) Results:

Careful probing revealed that people who gave reports of 1-4 hours often were office workers who did little or no heavy physical work. This appeared to be due to biasing characteristics of the question; saying "none" makes one appear to be entirely "non-physical", and is therefore somewhat socially undesirable. This problem was seen as related to respondent <u>decision</u> processes, rather than to comprehension or recall. A resolution was needed to make it "easier" for someone to report little work-related physical activity:

4) Suggested revision:

The next questions are about your job as a _____ for _____.

Does your job require you to do repeated strenuous physical activities such as lifting, pushing, or pulling heavy objects?

(IF YES:) During a typical work day, how many minutes or hours altogether do you spend doing strenuous physical activities?

Note that the results of a field-based survey experiment by Willis and Schechter (1997) have supported the contention that the revised question form is very likely a better expression than was the initial version.

EXAMPLE 6:

1) Original:

_____

Do you believe that prolonged exposure to high levels of radon gas can cause:

	YES	*NO*	*Don't Know*
Headaches?	__	__	__
Asthma?	__	__	__
Arthritis?	__	__	__
Lung Cancer?	__	__	__
Other cancers?	__	__	__

_____

2) Probes:

    a) **Why do you believe this?**

    b) **How sure are you of this?**

    c) **Is it difficult to answer these?**

3) Results:

Simple observation of subjects made it clear that this question is difficult to answer. Subjects required a long time to respond to each item, and tended to be unsure about several of the items. Further, probing revealed that the format encouraged a "guessing" strategy, rather than actual retrieval of information. Finally, for people who <u>do not</u> believe that exposure to radon is harmful, it became very tedious, and sometimes even offensive, to repeatedly ask about the specific harmful effects of radon.

In this case, it appeared that the subject's <u>decision processes</u> were again excessively burdened by the phrasing of the question.

4) Suggested revision:

Do you believe that prolonged exposure to radon is unhealthy, or do you believe that it has little or no effect on health?

(IF radon believed unhealthy:)

[SHOW CARD TO RESPONDENT] Which, if any, of these conditions do you believe can be caused by radon exposure?

___*Headaches* ___ *Lung cancer*
___*Asthma* ___ *Other cancers*
___*Arthritis* ___ *Don't Know*

The revised phrasing provides the respondent with a way to respond, once, that he or she does not believe that radon is harmful. Then, if he/she does believe it to be harmful, the next question simply allows him/her to "pick and choose" the items that seem appropriate. The burden on decision processes appeared to be reduced, using this alternative.

EXAMPLE 7:

1) Original:

> ***What is the primary reason you have not tested your home for radon?***

---

2) Probes:

   a) Is it hard to think of the main reason?

   b) Can you think of any other reasons?

   c) How much have you thought about having your home tested?

3) Results:

Although the question is easily enough understood, it was very difficult for subjects to produce a reasonable answer, especially if they had never given the issue much thought. Instead of simply saying "I never thought about it", or "I haven't gotten around to it", subjects tried to think of more "appropriate" answers, that appear to be more defensible. Here both recall and decision processes appeared to be operating.

4) Suggested solution: --- **DELETE QUESTION** ---

   The sponsor/client agreed that it was not especially useful to ask the reason that someone <u>had not</u> carried out this activity.

This example demonstrates an important point worth emphasizing; sometimes, there is no obvious "correction" to a survey question. Especially when subjects simply don't have information that we want, it is better to acknowledge that we may not want to ask that question. Thus, one effect of lab testing is to test the boundaries of "what can be asked and what can't."

5. DETECTION OF STRUCTURAL PROBLEMS IN SURVEY QUESTIONS

The discussion above has focused almost completely on cognitive problems in questionnaires; that is, problems involving the comprehension, recall, decision, or response processes necessary to adequately answer the question. However, cognitive interviewing has several overall positive effects, in addition to the understanding of specific cognitive processes:
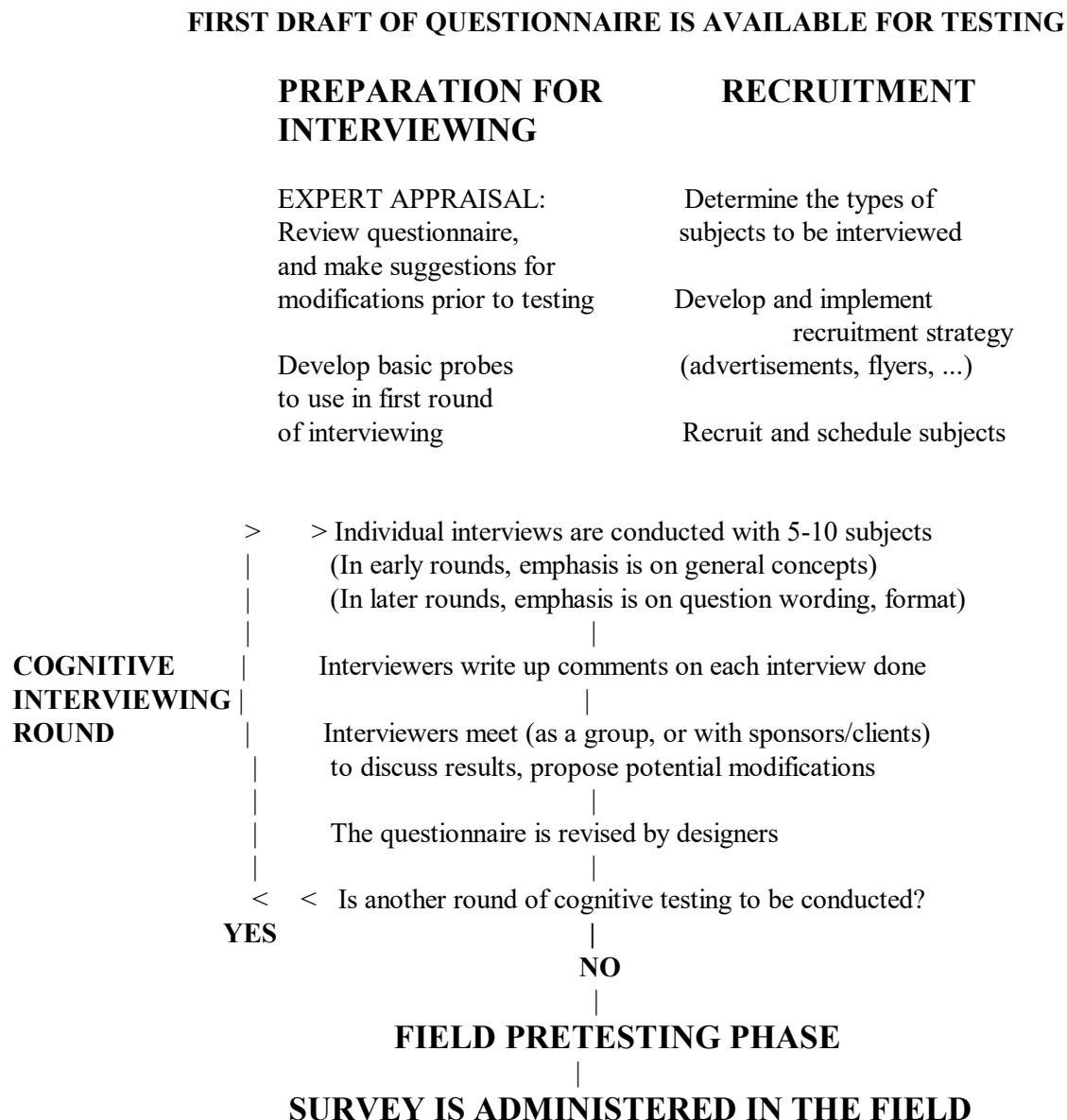
a) *Learning about the topic*: One can explore the nature of the underlying concepts to be measured in the survey, and the specific topical material, by relying on lab subjects as substantive "experts". For example, no one is more knowledgeable about the topic of illicit drug use than those individuals who have used them, and the basic logic of questions on the use of assistive devices can best be assessed through intensive discussions with individuals who use canes, wheelchairs, walkers, and so on.

b) *Learning about <u>non-cognitive</u> defects in the questionnaire*. An important beneficial effect of lab testing is to detect <u>structural</u>, or logical problems, not normally viewed as relevant to the cognitive approach. Structural problems are those features of the questionnaire, such as erroneous skip patterns, unclear layout, and other elements, that do not clearly involve the cognitive processes of the respondent. This problem category also includes more subtle types of logical problems in survey questions. For example, given the question: "<u>How long have you owned your house</u>?" the subject may simply respond that he is a renter. Here, it should not be strictly necessary to study cognitive processes to make the discovery that the question is flawed, because simple knowledge of the appropriate logical relationships ("some people own, some people rent") should have been sufficient to avoid such a problem. However, survey designers often fail to take into account all of these logical truths when constructing a questionnaire, and the laboratory-based interview allows the subject to spontaneously point out flaws (or for the interviewer to notice them, independently of the subject's behavior).

Of course, many structural problems could be detected by either a careful expert review, or in the field pretest, rather than through the cognitive interview. However, from a practical point of view, the expert review may never get done, or it can be imperfect. The field pretest generally occurs late in the process; it is much better to detect the problems earlier rather than later, and the cognitive interview serves this purpose well. Therefore, the cognitive interview often becomes the means for "fixing the logic of the questionnaire." Note that it takes no special "techniques" to detect the types of problems mentioned above, beyond simply attending to the possibility that they can occur.

6. THE SEQUENCE OF COGNITIVE INTERVIEWING ACTIVITIES

The following sections place these techniques described above into the broader context of conducting this testing within a real-life survey development process. To appreciate this overall process, it is useful to first consider an overview of the general sequence of events that may occur after a questionnaire is designed. Below is a schematic diagram of one such sequence that incorporates cognitive interviewing techniques, as well as other pretesting techniques, into the developmental and testing sequence:

FIRST DRAFT OF QUESTIONNAIRE IS AVAILABLE FOR TESTING

PREPARATION FOR INTERVIEWING	**RECRUITMENT**
EXPERT APPRAISAL: Review questionnaire, and make suggestions for modifications prior to testing	Determine the types of subjects to be interviewed
Develop basic probes to use in first round of interviewing	Develop and implement recruitment strategy (advertisements, flyers, ...)
	Recruit and schedule subjects


```
        >     > Individual interviews are conducted with 5-10 subjects
        |         (In early rounds, emphasis is on general concepts)
        |         (In later rounds, emphasis is on question wording, format)
        |                          |
COGNITIVE   |     Interviewers write up comments on each interview done
INTERVIEWING|                     |
ROUND       |     Interviewers meet (as a group, or with sponsors/clients)
        |          to discuss results, propose potential modifications
        |                          |
        |         The questionnaire is revised by designers
        |                          |
        <    <  Is another round of cognitive testing to be conducted?
    YES                            |
                                  NO
                                   |
                  FIELD PRETESTING PHASE
                                   |
            SURVEY IS ADMINISTERED IN THE FIELD
```

7. PRACTICAL ASPECTS OF COGNITIVE INTERVIEWING

How long should a cognitive interview be?

Although interviews of up to two hours are possible, a common view is that one-hour interviews are optimal; longer periods make excessive demands on subjects. In general, the interview process should be as flexible as possible, and not require interviewers to cover a certain number of pages of a questionnaire. Questionnaires often have skip patterns that result in widely varying actual questionnaire lengths for different individuals, and subjects vary in their overall speed and the degree to which they respond in detailed ways to either the survey questions, or to probe questions.

Note that even though the interview itself may take only an hour, the interviewing process requires considerably more time. In all, preparation, interviewing, and writing up results of the interview usually take at least three hours, and sometimes considerably more. Because of this, and because cognitive interviewing can be a taxing activity, it is recommended that any individual do no more than three interviews in a single day, if possible.

What types of individuals make effective interviewers?

It is unnecessary to have an advanced degree in psychology to be a good cognitive interviewer (although a behavioral sciences background appears to be helpful). We have found that good interviewers are those people who:

a) Have experience in questionnaire design, and are knowledgeable about both survey practice and about the purpose of the questionnaire to be tested. These skills are essential when the time comes to apply the results of the interviews in revising the questionnaire.

b) Have learned the basic premises of cognitive interviewing, and are familiar with the ways in which fundamental cognitive processes may influence the survey response.

c) Have been exposed to social science research concepts such as bias, context effects, measurement and scale effects, and so on.

d) Perhaps most importantly, have good inter-personal skills, are capable of putting a subject at ease, and remaining non-judgmental in approach. There is no common agreement concerning how "professional" versus "friendly" the interviewer should be during the interview itself, in order to obtain the best quality data (this may in part depend on the personality of the interviewer, as well as the philosophy of the organization).

A common question is whether field interviewers can be taught to perform laboratory cognitive interviews. This may be possible, if interviewers can be induced to "unlearn" some habits that are

very valuable for field interviewing, but that may be counterproductive for cognitive interviewing. In particular:

a) Field interviewers have learned over time "to make a question work", for example, by re-wording it, so that a confused respondent will ultimately provide a codeable response. It must be emphasized that our task in the lab is different; to <u>find</u>, rather than to adjust for, flaws in the questions.

b) Interviewers tend to work as fast as possible in the field, usually in order to complete a very long interview before the respondent becomes uncooperative. Interviewers must be reminded to work at an unhurried pace in the lab.

c) Field interviewers often focus their attention on very detailed formatting and other structural features such as skip pattern errors and spelling errors. They must be instructed that the format of the questionnaire may be very rough, and that it is the questionnaire <u>content</u> that is of primary concern in lab testing.

d) Field interviewers are taught not to deviate from the instructions contained in the instrument. In contrast, cognitive interviewers must be comfortable departing from the questionnaire flow when this appears to be called for. They also must be able to adjust to a situation in which sequencing instructions are incorrect or missing, which often occurs in the testing of a draft questionnaire.

Cognitive interviewer training

Cognitive interviewing is an acquired skill, consisting of a number of separate sub-skills. Optimally, good interviewers can serve as "detectives" who can find problems in survey questions, and as "engineers" who can work toward developing workable solutions to the problems defined. The former skill is generally obtained more quickly than the latter, and that the attainment of mastery is very gradual. Interviewers can be taught in an incremental, step-wise fashion, consisting of as many of the following steps as possible:

a) Trainee interviewers should conduct expert reviews or appraisals of questionnaires to make determinations of structural and potential cognitive problems. They also attend early questionnaire design meetings, as well as meetings where cognitive interviewers discuss the results of cognitive testing.

b) Trainees familiarize themselves with material on the philosophy and purposes of the cognitive aspects of survey methodology and cognitive interviewing techniques.

c) They are taught the specific probing methods for use in the interview, in a lecture-based training program.

d) They are shown examples of the way that probing is used to detect problems in survey questions. This can be in both written form, and through the use of audio- and video-taped recordings of previous interviews.

e) Trainees observe experienced interviewers performing actual interviews. Unless a topic is very sensitive, subjects generally have no objection to being observed by an individual who is described as "in training."

f) Trainees perform one or more interviews while being observed by a practiced interviewer, or compile tape recording of the interviews for review by other staff. The trainee can then be given feedback.

g) Trainees attends questionnaire review meetings, subsequent to the interviews, and attempt to make specific recommendations for solution of the observed problems.

This guide is intended to serve as "training." There is no substitute for experience, however, and interviewers should begin interviewing as soon as they have a fairly good idea of what is involved.

Other considerations for interviewing

There are several features of laboratory interviewing that are important for cognitive interviewers to understand, and that are useful to express to the subject, before beginning a cognitive interview:

a) The interviewer should stress to the subject that he/she is not primarily collecting survey data on them, but rather testing a questionnaire that has questions that may be difficult to understand, hard to answer, or that make little sense.

b) Make clear that although we are asking the subject to answer the survey questions as carefully as possible, *we are primarily interested in the ways that they arrived at those answers, and the problems they encountered*. Therefore, any detailed help they can give us is of interest, even if it seems irrelevant or trivial.

c) If think-aloud responding is desired, tell subjects, at the least, to "think out loud to the extent possible, so we can tell what you are thinking about when you answer the questions." Be warned that this introduction generally does not produce a great amount of think-aloud, however. Eliciting a spontaneous flow of verbalization often requires subject practice with the technique.

d) It also is somewhat helpful to add: "I didn't write these questions, so don't worry about hurting my feelings if you criticize them- my job is to find out what's wrong with them". This helps to "bring out" subjects who may otherwise be sensitive about being overly

critical.

8. INTERVIEWING LOGISTICS

See Willis (1994) for a very detailed description of the operation of a permanent Federal-level cognitive laboratory (the Questionnaire Design Research Laboratory at NCHS).

Recruitment

In order to test a questionnaire, recruitment of the appropriate subjects is vitally important. One initially needs to identify and recruit volunteers from appropriate sub-populations for testing the survey questionnaire, taking into account several considerations:

> a) Subjects either have characteristics of interest for the survey (a particular status with respect to health, work, age, sex characteristics), or they may be "general" subjects, for questionnaires that are asked of the general population. However, even for a questionnaire that is intended for special populations, it is worth testing the initial screening sections on people who <u>do not</u> exhibit the characteristic(s) of interest. This practice allows the interviewers to ensure that the questions do not create problems in the majority of cases in which the questionnaire will be administered (where the respondent does <u>not</u> have the characteristic). As an example, a questionnaire that is intended to identify individuals with Pediatric conditions might be tested only on individuals who answer an advertisement for "people with foot problems." However, failure to test the screening questions on individuals without foot problems could be catastrophic. If, for example, virtually everyone answers initial screening questions (in effect asking: "Do you have any foot problems") in the affirmative, a large number of inappropriate respondents might wind up "passing" the screener and be subjected to a series of completely irrelevant follow-up questions. As a general rule, questionnaires that seek to identify a particular population should be tested to determine that they adequately 1) screen in people having the characteristic of interest (that is, they exhibit sensitivity), and also 2) screen out those who do not (they also exhibit specificity).

> b) Subjects are recruited through newspapers, fliers, service agencies, and support groups. If payment will be involved, flyers and newspaper ads should clearly emphasize this feature (monetary incentives tend to be very effective).

> c) Statistical sampling methods are *not* normally used in obtaining laboratory subjects. At most, we use a "quota" sample, in which one attempts to obtain a range of ages, genders, and socio-economic levels, if possible.

Payment