Developing and Testing Survey Questions

Gordon Willis, Ph.D.
Applied Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute
Willisg@mail.nih.gov
Always remember that questionnaire data are limited by error due to self-reporting.
Always remember that questionnaire data are limited by error due to self-reporting.
Questionnaire development approach


I. Determine Analytic Objectives
   - *What types of data will answer the research question?*

II. Develop general concepts to be covered
   - *List areas to be covered by questions*

III. Translate concepts into questions

IV. “Appraise” questions for common pitfalls

V. Evaluate questions empirically
I. Determine Analytic Objectives

- *What types of data will answer the research question?*

<table>
<thead>
<tr>
<th>% of respondents with a preventive care visit the past 12 months at a -</th>
<th>% of respondents asked about smoking status in past 12 months at a -</th>
<th>% of respondents checked for oral cancer in past 12 months at a -</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dental office (X%)</td>
<td>(X%)</td>
<td>(X%)</td>
</tr>
<tr>
<td>Physician office (X%)</td>
<td>(X%)</td>
<td>(X%)</td>
</tr>
</tbody>
</table>
I. Determine Analytic Objectives

- What types of data will answer the research question?

II. Develop general concepts to be covered

- List areas to be covered by questions
  - Whether visit in past 12 months to dentist, doctor
  - Whether smoking status was asked at any visit
  - Whether oral cancer check done at any visit
  - (Smokers) Whether advice to stop smoking was given at any visit
  - etc.
I. Determine Analytic Objectives
   • *What types of data will answer the research question?*

II. Develop general concepts to be covered
   • *List areas to be covered by questions*

III. Translate concepts into questions
I. Determine Analytic Objectives
   • What types of data will answer the research question?

II. Develop general concepts to be covered
   • List areas to be covered by questions

III. Translate concepts into questions

IV. “Appraise” questions for common pitfalls

(See Aday, L., & Cornelius, L. (2006). Designing and Conducting Health Surveys, Wiley)
I. Determine Analytic Objectives
   • What types of data will answer the research question?

II. Develop general concepts to be covered
   • List areas to be covered by questions

III. Translate concepts into questions

IV. “Appraise” questions for common pitfalls

V. Evaluate questions empirically
First, consider **Administration Mode**:

Problems tend to be specific to choice of:

- Interviewer administration
  - Telephone
  - In-person

- Self-Administration
  - Mailed paper
  - Internet
  - Personal Digital Assistant (PDA)/Smartphone
Sources of Response Error: Tourangeau (1984) cognitive model

- Encoding of question (understanding it)
  - Have you ever received care from an oral surgeon?
- Retrieval of information (knowing/remembering)
  - How many times have you ridden in a passenger airplane?
- Decision and judgment processes (truth, adequacy)
  - How many sex partners have you had in the past 12 months?
- Response (matching internal representation to given categories)
  - Would you say your health is excellent, very good, good, fair, or poor?
Evaluation: How do we find questionnaire problems?

- Systematic “Expert Review” (aka “Appraisal”, “Technical Review”) ->

  Question Appraisal System (Willis & Lessler, 1999):
  
Question Appraisal System (QAS-2009)  
Simplified Coding Form

<table>
<thead>
<tr>
<th>A.</th>
<th>INSTRUCTIONS: Problems with any introductions, instructions, or explanations from the respondent's point of view</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.</td>
<td>CLARITY: Problems related to communicating the intent or meaning of the question</td>
</tr>
<tr>
<td>WORDING:</td>
<td>Question is lengthy, awkward, ungrammatical, or contains complicated syntax.</td>
</tr>
<tr>
<td>TECHNICAL TERM(S) are undefined, unclear, or complex</td>
<td></td>
</tr>
<tr>
<td>VAGUE:</td>
<td>There are multiple ways to interpret the question or to decide what is to be included or excluded.</td>
</tr>
<tr>
<td>REFERENCE PERIODS are missing, not well specified, or in conflict.</td>
<td></td>
</tr>
<tr>
<td>C.</td>
<td>ASSUMPTIONS: Problems with assumptions made/underlying logic</td>
</tr>
<tr>
<td>INAPPROPRIATE ASSUMPTIONS:</td>
<td>Made about the respondent or about his/her living situation</td>
</tr>
<tr>
<td>ASSUMES CONSTANT BEHAVIOR:</td>
<td>or experience for situations that vary.</td>
</tr>
<tr>
<td>DOUBLE-BARRELED:</td>
<td>Contains more than one implicit question</td>
</tr>
<tr>
<td>D.</td>
<td>KNOWLEDGE/MEMORY: Respondents are likely to not know or have trouble remembering information</td>
</tr>
<tr>
<td>KNOWLEDGE or ATTITUDE:</td>
<td>May not exist.</td>
</tr>
<tr>
<td>RECALL failure:</td>
<td>Respondent may not remember the information asked for.</td>
</tr>
<tr>
<td>COMPUTATION problem:</td>
<td>The question requires a difficult mental calculation.</td>
</tr>
<tr>
<td>E.</td>
<td>SENSITIVITY/BIAS: Sensitive nature or wording/bias</td>
</tr>
<tr>
<td>SENSITIVE CONTENT (general):</td>
<td>The question asks about a topic that is embarrassing, very private, or that involves illegal behavior.</td>
</tr>
<tr>
<td>SENSITIVE WORDING (specific):</td>
<td>Given that the general topic is sensitive, the wording should be improved to minimize sensitivity.</td>
</tr>
<tr>
<td>SOCIALY ACCEPTABLE response is implied by the question.</td>
<td></td>
</tr>
<tr>
<td>F.</td>
<td>RESPONSE CATEGORY PROBLEMS</td>
</tr>
<tr>
<td>OPEN-ENDED QUESTION:</td>
<td>That is inappropriate or difficult.</td>
</tr>
<tr>
<td>MISMATCH between question and response categories.</td>
<td></td>
</tr>
<tr>
<td>TECHNICAL TERM(S) are undefined, unclear, or complex.</td>
<td></td>
</tr>
<tr>
<td>VAGUE:</td>
<td>Response categories are subject to multiple interpretations.</td>
</tr>
<tr>
<td>OVERLAPPING response categories</td>
<td></td>
</tr>
<tr>
<td>MISSING eligible responses in response categories.</td>
<td></td>
</tr>
<tr>
<td>ILLOGICAL ORDER:</td>
<td>Of response categories.</td>
</tr>
<tr>
<td>G.</td>
<td>FORMATTING/ORDERING PROBLEMS: Problems with visual layout, or question ordering is inappropriate</td>
</tr>
</tbody>
</table>

National Cancer Institute
Long/Convoluted Phrasing:

• The last time that you were seen by a doctor, nurse, or other health professional, as part of a regular medical check-up, did you receive any tests specifically designed to diagnose the presence of certain types of cancer?

• Typical response = “What?”

⇒ Especially for interviewer-administration: DECOMPOSE question into concepts -- ask more, but simpler questions, with use of skips
Lack of Clarity: Difficult-to-Understand Questions

“Decomposition” into simpler phrasing

- **When did you last see a doctor, nurse, or other health professional, to get a regular medical check-up?**
- **During that visit, did you receive any tests that check for cancer?**
- **What types of cancer were you checked for?**

⇒ Doesn’t solve problem of respondents not knowing the answer, but makes the question more understandable.

⇒ A VERY common problem -- and a very common “fix”
Lack of clarity: Terms/phrases are difficult to understand

Complex/Unfamiliar Terminology:

• Were you seen on an inpatient or outpatient basis?
• Have you ever had a colonoscopy or sigmoidoscopy?

Better to use “simple language”:

• Did you stay overnight at the hospital?
• (Use explanation of what the medical test entails)
Many questions that use “simple language” are variably interpreted:

- Have you ever been a regular smoker?
- Does anyone in your family now have a car?
- Do you think that headaches can be effectively treated?
Lack of Clarity/Vagueness

Have you had your blood tested for the AIDS virus?

• Unclear: a) “Did I take the initiative in deciding to have my blood tested?” versus b) “Was it tested as part of any type of blood test?”

• If the issue of interest is the act of testing, simply ask “As far as you know, has your blood ever been tested for the AIDS virus?”
**Retrieval problem:**
Respondent doesn’t know the answer

*Estimate the number of your women patients with whom you discussed enrollment in a cancer TREATMENT trial in the **LAST 12 MONTHS:***

<table>
<thead>
<tr>
<th>ALL WOMEN</th>
<th>ASIAN AMERICAN WOMEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Cancer treatment trials</td>
<td>____</td>
</tr>
<tr>
<td>Breast Cancer Treatment Trials</td>
<td>____</td>
</tr>
</tbody>
</table>
Logical Problems

- Questions that simply don’t make sense in case of at least some respondents
  - Sometimes due to cultural issues--
    “Have you ever switched from a stronger to a lighter cigarette”

- Problems exist separately from respondent’s interpretation, recall, decision processes

- Avoidance, resolution requires clear understanding of environment, culture, ‘respondent reality’
• **Self Administered:**
  
  – Especially for mail survey, DON’T CROWD QUESTIONS ONTO PAGES (better to have more pages, “open” layout)

  – If paper (not computer): MINIMIZE use of skip patterns - respondents get confused very easily, make errors
1. Over the past 12 months, how often did you drink tomato juice or vegetable juice?
   - NEVER (GO TO QUESTION 2)
     - 1 time per month or less
     - 2–3 times per month
     - 1–2 times per week
     - 3–4 times per week
     - 5–6 times per week

1a. Each time you drank tomato juice or vegetable juice, how much did you usually drink?
   - Less than ¼ cup (6 ounces)
   - ¼ to ⅛ cups (6 to 10 ounces)
   - More than ⅛ cups (10 ounces)

2. Over the past 12 months, how often did you drink orange juice or grapefruit juice?
   - NEVER (GO TO QUESTION 3)
     - 1 time per month or less
     - 2–3 times per month
     - 1–2 times per week
     - 3–4 times per week
     - 5–6 times per week

2a. Each time you drank orange juice or grapefruit juice, how much did you usually drink?
   - Less than ¼ cup (6 ounces)
   - ¼ to ⅛ cups (6 to 10 ounces)
   - More than ⅛ cups (10 ounces)

3. Over the past 12 months, how often did you drink other 100% fruit juice or 100% fruit juice mixtures (such as apple, grape, pineapple, or others)?
   - NEVER (GO TO QUESTION 6)
     - 1 time per month or less
     - 2–3 times per month
     - 1–2 times per week
     - 3–4 times per week
     - 5–6 times per week

4. Over the past 12 months...
   - How often did you drink other fruit drinks (such as cranberry cocktail, Hi-C, lemonade, or Kool-Aid, diet or regular)?
     - NEVER (GO TO QUESTION 5)
       - 1 time per month or less
       - 2–3 times per month
       - 1–2 times per week
       - 3–4 times per week
       - 5–6 times per week

4a. Each time you drank fruit drinks, how much did you usually drink?
   - Less than ¼ cup (6 ounces)
   - ¼ to ⅛ cups (6 to 10 ounces)
   - More than ⅛ cups (10 ounces)

4b. How often did you drink sugar-free fruit drinks?
   - Almost never
   - About once a month
   - About once a week
   - Almost every day

5. How often did you drink milk as a beverage, or coffee or chocolate milk (NOT in coffee or chocolate milk beverages)?
   - NEVER (GO TO QUESTION 6)
     - 1 time per month or less
     - 2–3 times per month
     - 1–2 times per week
     - 3–4 times per week
     - 5–6 times per week

5a. Each time you drank milk as a beverage,
• Interviewer Administered:
  – Remember that the respondent doesn’t read along—CANNOT be designed like self-administered form
    
    How many glasses (8 oz) of milk (whole, 2%, or skim milk) did you drink yesterday?
  – Need to use HAND/SHOW CARDS for long lists
  – Response categories implied should match those on form:
How do you feel about your present weight? (Do not read):

___ Overweight
___ Underweight
___ About right

In the past ten years, how many times have you had a headache severe enough to cause you to stay in bed for more than half a day? (Do not read)

___ Never
___ 1-5 times
___ 6-10 times
___ more than 10 times
PROBABLY THE SINGLE MOST SERIOUS PROBLEM IN FEDERAL SURVEYS

- Length will increase survey costs, decrease response rate

- Length/burden will lead to response error in unmeasurable ways
  - Interviewers read faster, less carefully (get it done!)
  - Respondents answer less carefully as time goes on

- If possible, don’t go over (average) 30 minutes for face-face, 15 minutes for phone survey, 15 minutes for Web survey
Evaluation: How do we find questionnaire problems?

- **Cognitive interviewing:** Search for underlying problems

**Manual available at:**


**Book:**

The cognitive testing process in a nutshell

- **Develop** a questionnaire or material (advance letter, etc.) to be evaluated
- **Recruit** (paid) members of the targeted population (e.g., recipients of home loans, people without employment, cancer survivors)
- Conduct **one-on-one interviews**, in “laboratory” or other location:
  - Home
  - Homeless shelter
  - Health clinic
  - Elderly center
- Use both **Think-Aloud** and **Verbal Probing** techniques
“Classic” verbal probes

Comprehension probe: What does the term “dental sealant” mean to you?

Paraphrase: Can you repeat the question in your own words?

Confidence judgment: How sure are you that your health insurance covers...

Recall probe: How do you know that you went to the dentist 3 times...

‘General’ probe: How did you arrive at that answer?
Tested ("classic") question: Pain in the abdomen

“In the last year have you been bothered by pain in the abdomen?”

What probes make sense here?

- What time period are you thinking about, exactly?
- What does “bothered by pain” mean to you?
- Where is your “abdomen?”
Please look at this diagram. During the past 12 months, have you had pain in this area (the area shaded on the diagram)?
Using cognitive interviews to detect question wording problems

**VERSION 1 (No filter)**

On a typical day, how much time do you spend doing strenuous physical activities such as lifting, pushing, or pulling?

___ None ___ Less than 1 hour ___ 1-4 hours ___ 5+ hours

**VERSION 2 (Filtered)**

On a typical day, do you spend *any* time doing strenuous physical activities such as lifting, pushing, or pulling?

IF YES: Read Version 1

---

On a typical day, how much time do you spend doing strenuous physical activities such as lifting, pushing, or pulling?

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>&lt;1</th>
<th>1-4</th>
<th>5+</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FIELD PRETEST (N=78)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No-filter version</td>
<td>32%</td>
<td>32%</td>
<td>35%</td>
<td>0%</td>
</tr>
<tr>
<td>Filtered version</td>
<td>72%</td>
<td>18%</td>
<td>10%</td>
<td>0%</td>
</tr>
<tr>
<td><strong>WOMEN’S HEALTH (n=191)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No-filter version</td>
<td>4%</td>
<td>42%</td>
<td>50%</td>
<td>4%</td>
</tr>
<tr>
<td>Filtered version</td>
<td>49%</td>
<td>16%</td>
<td>27%</td>
<td>8%</td>
</tr>
</tbody>
</table>
**Verbal probing techniques: Concurrent versus retrospective**

**Concurrent probing:** Probe immediately after the subject has answered each survey question

**Advantage:** Probing when the memory still exists

**Disadvantage:** Measurement process interrupts normal flow

**Retrospective probing:** Wait until after the interview, and then go back to probe

**Advantage:** Mirrors “field” procedures
- Is particularly useful for self-administered Q’s

**Disadvantage:** Subject may have forgotten key information

Gordon Willis NCI 3/2012
Adapting C.I. to Web surveys

• Challenge: Integrating C.I. and Usability testing
  – We need to consider both (a) the **questionnaire** and (b) **Usability** of the computer system
  – Makes sense to (a) test questionnaire prior to programming, but to (b) include “regular” cognitive testing during Usability test (combine these so we don’t forget about the questions!)
  – Example of a complex Web questionnaire ->
Diet History Questionnaire

Please check the box next to each food that you ate at least once in the **past 12 months**.

- [ ] Applesauce  
- [ ] Apples  
- [ ] Pears (fresh, canned, or frozen)  
- [ ] Bananas  
- [ ] Dried fruit, such as prunes or raisins (not including dried apricots)  
- [ ] Peaches, nectarines, or plums  
- [ ] Grapes  
- [ ] Cantaloupe  
- [ ] Melon, other than cantaloupe (such as watermelon or honeydew)  
- [ ] Strawberries  
- [ ] Oranges, tangerines, or tangelos  
- [ ] Grapefruit  
- [ ] Other kinds of fruit (not listed above)

Next Question

1) **Where do users look?**
2) **Can they tell what the panel on the left is for?**
3) **Do they go directly to the question...?**
4) **Can they tell what the colors signify?**
Examples of Cognitive Testing Reports

Many cognitive testing reports are retrievable from the Web, through the **Q-BANK database**

Miller, et al. (National Center for Health Statistics, Centers for Disease Control and Prevention):

**Q-Bank Home:**  [http://www.cdc.gov/qbank/](http://www.cdc.gov/qbank/)

-> Consult these reports for examples of how cognitive testing results are written up in U.S. Federal Agencies
Other methods: How do we find questionnaire problems?

- Field pretesting (respondent debriefing, behavior coding)
- Analysis (item missing, strange looking data distribution, psychometric analysis of attitudinal measures)
Useful Internet Resources

1) US Census Bureau guide to self-administration:

2) NCI resources:
   a) The Question Appraisal System (Full version):
   b) Cognitive Interviewing: A How-To Guide:
   c) An Introduction to Modern Measurement Theory:
In closing...

“The uncreative mind can spot wrong answers, but it takes a very creative mind to spot wrong questions.”

- Anthony Jay