Questionnaire Design

Theresa A Scott, MS

Department of Biostatistics
theresa.scott@vanderbilt.edu
D-2217 MCN; 3-1713
http://biostat.mc.vanderbilt.edu/TheresaScott

Questionnaires in Clinical Research

- Much of the data in clinical research is gathered using questionnaires or interviews.
- The validity of the results depends on the quality of these instruments.
- Poorly designed questions can result in poor data quality.
- Critical to improve our understanding of the inherent flaws of survey questions.

▷ *Objective of this lecture*:

- Describe the components of good questionnaires and outline the procedures for developing them.

Designing a Questionnaire [1]

- Requires development of a set of questions used to obtain clinically and statistically useful information from an individual.
- Difficult for several reasons:
  - Each question must provide a valid and reliable measure.
  - The questions must clearly communicate the research intention to the respondent.
  - The questions must be assembled into a logical, clear instrument that flows naturally and will keep the respondent sufficiently interested to continue cooperation.
- Good questionnaires are difficult to construct; bad questionnaires are difficult to analyze.

Designing a Questionnaire [2]

▷ Start early and plan for plenty of time.

- More challenging and time-consuming than you think.
- Time spent $\Leftrightarrow$ Quality of questionnaire.
- Wrong approach:
  - A questionnaire is finished when time runs out, not necessarily when it is the best it can be.

▷ Three distinct phases:

- Initial questionnaire planning.
- Development of specific questions.
- Final construction of the data collection instrument as a whole.

Phase I:

Initial Questionnaire Planning

Initial Questionnaire Planning [1]

▷ **Prior** to writing any questions:

- Define the problem and specific aim(s) of the study, including the population of interest.
- Make a detailed list of the information to be collected and concepts to be measured.
  - Don't forget about demographics and possible inclusion/exclusion criteria to define the target population.
- Formulate a statistical analysis plan that outlines how every item will be analyzed.
  - Helpful to list the role of each item (predictor, outcome, or confounder) in addressing each specific aim.
  - Useful to think ahead to the reporting of results (i.e., sketch out the final results tables).

Initial Questionnaire Planning [2]

- **Prior** to writing any questions, *cont'd*:
  - Review the literature and collect any existing measures, related surveys, and/or data collection instruments that might have measured similar concepts.
    - Saves development time and allows comparison with other studies if used appropriately.
    - Ideal to use existing instruments without modification.
    - Existing instruments may not be entirely appropriate for the question or the population, or may be too long; may be necessary to delete, change, or add a few items.
    - Direct comparison with other studies may no longer be possible if original instrument has been modified.
      - No longer comparing apples to apples.

Initial Questionnaire Planning [3]

- Compose a draft.
  - Questions can ask for reports of
    - *Knowledge* – How often does the National Cancer Institute (NCI) recommend that women age 50 and older should get mammograms?
    - *Behavior* – Since turning 50, how many mammograms have you had?
    - *Attitude* – Do you agree with the mammogram recommendations of the NCI for women age 50 and older?
    - *Opinion* – How often do you think that women age 50 and older should get mammograms?
    - *Proxy* (ask respondent regarding others) – Since turning 50, how many mammograms has your wife had?
      - Discouraged by the IRB

Phase II:

Development of Specific Questions

Development of Specific Questions [1]

▷ First goal: Shorten the set of questions.

- Questions not essential to addressing the specific aim(s) increase the amount of effort involved in entering, cleaning, and analyzing the data.
  - Decrease the overall quality and productivity of the study.
- Every item in the questionnaire must be a meaningful contribution to the intended analyses.
  - Compare the draft questions to the survey objectives to ensure that the right types of questions (e.g., knowledge) are being asked for a given topic.
- Resist the temptation to include additional questions or measures "just in case" they might produce interesting data.

Development of Specific Questions [2]

▷ Second goal: Refine the remaining questions.

- Every word in a question can influence the validity and reproducibility of the responses.
- Iterative cycles of review and revision.
  - Refine and clarify the research objectives.
  - Focus the concepts included in the survey.
- Target:
  - Terms and concepts should be familiar and easy to understand.
  - Cues and ordering of questions should serve to stimulate recall.
  - Ordering and format of questions should be unbiased and balanced.

Development of Specific Questions [3]

▷ *Terms and concepts should be familiar and easy to understand.*

- Questions should be simple, be free of ambiguity, and encourage accurate and honest responses without embarrassing or offending the respondent.
  - *Clarity*: specific and concrete wording.
    - "How much exercise do you usually get?" vs. "During a typical week, how many hours do you spend exercising (e.g., vigorous walking or sports)?".
  - *Simplicity*: short non-technical words and simple grammar.
    - "Over-the-counter medications" vs. "Drugs you can buy without a doctor's prescription".
  - *Neutrality*: avoid "loaded" words and stereotypes.
    - "During the last month, how often did you drink too much alcohol?" vs. "During the last month, how often did you drink more than five drinks in one day?".

Development of Specific Questions [4]

▷ *Cues and ordering of questions should serve to stimulate recall.*

- Respondents often asked to recall and access information from memory.
  - Problems: asked to recall too much information or asked to recall information from too far in memory.
  - Regarding behavior, interested in the average or the extremes?
- Steps that can help the respondent's memory search:
  - Ask a short series of related questions.
  - Provide an anchor for the reference period or time frame.
    - Goal: To ask about the shortest recent segment of time that accurately represents the characteristic over the whole period of interest for the research question.
    - Example: "During the last 7 days, how many beers did you have?"
  - Keep recall to a minimum and focus on the recent past.

Development of Specific Questions [5a]

▷ *Ordering and format of questions should be unbiased and balanced.*

- Ordinarily group together questions on the same subject matter.
  - The groups can encourage context effects.
    - Occur when two or more questions deal with aspects of the same issues or with closely related issues.
  - The order within the groups can encourage order effects.
  - General questions seem to be more sensitive.
    - General questions are so broad that their frame of reference is more open to interpretation and, therefore, respondents are more likely to use the context from another question in their interpretation.

Development of Specific Questions [5b]

▷ *Ordering and format of questions should be unbiased and balanced.*

- Open format → Open-ended questions.
  - Respondent free to answer; verbatim response recorded.
    - "What habits do you believe increase a person's chance of having a heart attack?"
  - Advantages/Disadvantages:
    - Fewer limits imposed by researcher regarding the response.
    - More information reported than is possible with a discrete list of responses, but responses may be less complete.
    - Usually require qualitative methods or special systems to code and analyze, which takes more time.
  - Often require subjective judgments.
    - Often used in exploratory phases of question design.
    - Phrases and words used by respondents can form the basis for more structured items in a later phase of the survey.

Development of Specific Questions [5c]

▷ *Ordering and format of questions should be unbiased and balanced.*

- Closed format → Closed-ended questions.
  - Respondent instructed to choose from two or more pre-selected responses provided.
    - "Which one of the following do you think increases a person's chance of having a heart attack the most? (Check one.)
      [ ] Smoking [ ] Being overweight [ ] Stress"
  - Advantages/Disadvantages:
    - Quicker/easier to respond to; responses easier to analyze.
    - Provided pre-selected responses often help clarify the meaning of the question.
    - Do not allow respondents to express their own, potentially more accurate, answers.

Development of Specific Questions [5d]

▷ *Ordering and format of questions should be unbiased and balanced.*

- Closed-ended questions, *cont'd*.
  - Important to include "Other (please specify)", "Don't know", or "None of the above" options (where appropriate) to ensure the pre-selected provided responses are as exhaustive as possible.
  - Pre-selected provided responses should be mutually exclusive (i.e., the categories should not overlap) and contain no missing intervals.
    - Age categories of 20-25, 25-30, and 30-35 vs. 20-24, 25-29, and 30-35.
  - Includes Likert scales (e.g., "Disagree", "Neutral", "Agree"), quality-of-life (QOL) scales with scores that range from (e.g.) 1 to 100, and visual analog scales that instructs the respondent to mark a line at a spot along a continuum.

Development of Specific Questions [5e]

▷ *Ordering and format of questions should be unbiased and balanced.*

- Closed-ended questions, *cont'd*.
  - When the question allows more than one answer, instructing the respondent to mark "all that apply" is not ideal.
    - Does not force the respondent to consider each possible response, and a missing item may represent either an answer that does not apply or an overlooked item.
    - Better, therefore, to ask respondents to mark each possible response as either "yes" or "no".
    - "Which of the following increases the chance of having a heart attack?

      Smoking: [ ] Yes [ ] No [ ] Don't know
      Being overweight: [ ] Yes [ ] No [ ] Don't know
      Stress: [ ] Yes [ ] No [ ] Don't know"

Pitfalls to Avoid [1]

- Double-barreled questions.
  - Questions that use the words "or" or "and" can lead to unsatisfactory responses – each question should contain only one concept.
  - "How many cups of coffee or tea do you drink during a day?" vs. "How many cups of the following do you drink during a typical day? Coffee: ___ cups/day Tea: ___ cups/day"
- Hidden assumptions.
  - May not apply to all respondents.
  - "I felt that I could not shake off the blues even with help from my family" assumes that respondents have families and ask for emotional support.

Pitfalls to Avoid [2]

- The question and answer options don't match.
  - "Have you had pain in the last week?
    [ ] Never [ ] Seldom [ ] Often [ ] Very often"
    - Grammatically incorrect and can be confusing.
    - Question should be changed to "How often have you had pain in the last week?" or the response options should be changed to No/Yes.
  - "I am sometimes depressed. [ ] Agree [ ] Disagree"
    - Question about intensity is given agree/disagree options.
    - Disagreeing with this statement could mean that the person is often depressed or never depressed.
    - Usually clearer to use a simple question about how often the person feels depressed matched with options about frequency ("never", "sometimes", "often").

Other things to consider

- ▷ Branching questions:
  - Respondents' answer to initial question determines whether they are directed to answer additional questions or skip ahead to later questions (e.g., "Go to question 11").
  - Saves time; avoids irrelevant or redundant questions.
  - Can be used to "screen" respondents to target population.
    - "Please discontinue the survey if you answered 'No' to either of the two previous questions."
- ▷ (Prospective) Diaries:
  - Keeps track of frequency, intensity, duration, etc.
  - Approach allows average scores (daily, weekly, etc) of the measure being assessed to be calculated.
  - Can be time consuming for respondents; can lead to more missing data than retrospective questions.

Phase III:

Assembling the Final Questionnaire

Assembling the Final Questionnaire

▷ *Objective*: Fit the items together in a meaningful way so that the entire questionnaire is unified.

- Order of sections of questions and order of questions within sections.
- Question and response formats.
- Skip patterns/Branching questions.

▷ Also need to consider *mode of administration*.

- Self-administered questionnaire, face-to-face interview, telephone interview, or computer-assisted approaches?
- For self-administered questionnaires, give to respondents in person or administer through the mail, by email, or via a Web site?

Formatting [1]

- Describe purpose of study and how data will be used in a brief statement.
- Mention confidentiality of respondent and their right to refuse to answer any question.
- Instruct respondent/provide example(s) of how to complete specific questions (e.g., "Check one").
- Indicate unit of measurement (e.g., cups/day) for clarification.
- Group questions concerning major subject areas together and introduce them by headings or short descriptive statements.
- If the instructions include different time frames, repeat the time frame at the top of each set of new questions.
- Want plenty of white space.

Formatting [2]

- Space pre-selected provided answers wide enough apart so easy to circle/check the intended response without the mark accidentally including other responses.
- Open-ended questions: response space should be big enough to allow respondents with large handwriting to write comfortably.
- Closed-ended questions: line up answers vertically and precede them with boxes or brackets to check, or by numbers to circle, rather than open blanks.
- Place identifier (e.g., random ID) on each page of a multi-page, paper-based questionnaire in case pages separate.
- Use larger font size (e.g., 14) and high contrast (black on white).

Mode of Administration [1]

▷ Self-administered questionnaires:

- More economical, more readily standardized, and the added privacy can enhance the validity of responses.
- No middle-man bias (no verbal or visual clues from an interviewer to influence the respondent); more uniform.

▷ Interviews:

- Can ensure more complete responses and enhance validity through improved understanding.
- May be necessary when participants will have variable ability to read and understand questions.
- Requires substantial training and practice of interviewers.

Mode of Administration [2]

- Self-administered questionnaires vs. interviews, *cont'd*.
  - Both susceptible to errors caused by imperfect memory.
  - Both affected by the respondent's tendency to give socially acceptable answers, although not necessarily to the same degree.
- Another decision to make: *software*.
  - Software to aid creation/formatting, administration (e.g., create Web site), and/or data collection/entry.
  - An option: *REDCap Survey*.
    - Go to `www.mc.vanderbilt.edu`.
    - Click on "StarBRITE" link under "For Employees" area.
    - Login with your VUnetID and password.
    - Click on the "Data Management" tab.

Other things to consider [1]

- Pre-testing:
  - Done to clarify, refine, and time the instrument.
  - Large pilot studies valuable to find out whether each question produces an adequate range of responses and to test the validity and reproducibility of the instrument.
- Translation:
  - Use professional or certified translator.
  - Initially translate instrument to language of interest; follow with a back-translation.
  - Consider cross-language equivalence/cultural differences.
    - Example: Highest level of education – school system in Mexico and United States very different.
  - Submission of translated instrument required by IRB.

Other things to consider [2a]

▷ Creating Scales and Summary Scores:

- Abstract variables (e.g., quality of life, QOL) commonly measured by generating a summary score from the responses to a series of questions that are converted to a numeric scale.
- Advantages/Disadvantages:
  - Can increase the range of possible responses (e.g., multi-item QOL scale that generates scores ranging from 0 to 100 vs. single QOL questions with "poor" to "excellent" responses).
  - Produce results (QOL = 46.2) that can be difficult to understand intuitively.

Other things to consider [2b]

▷ Creating Scales and Summary Scores, *cont'd*:

- Difficult to develop a *new* multi-item scale and/or summary score.
  - What items are needed to define the scale? What numeric value should each item response be assigned to? How should the numeric values be summarized across the items? How do you know you made the right decisions?
  - Systematic approach needed can take several years from initial draft to final product; involves *validation*.
  - Should generally only be undertaken for variables that are central to a study, and when existing measures are inadequate or inappropriate.

Other things to consider [3]

▷ Validation:

- Involves assessing questionnaire for
  - *Validity*: an aspect of accuracy; how well the measurement represents the phenomenon of interest; includes
    - *face validity*: subjective judgment that the items assess the characteristics of interest.
    - *content validity*: how well the assessment represents all aspects of the phenomena under study.
    - *construct validity*: how well a measurement conforms to the theoretical constructs.
  - *Reproducibility*: precision.
- If possible, new instrument compared with established gold standard.
- Process is time consuming and expensive; worthwhile only if existing instrument is inadequate.

Other things to consider [4]

▷ Response rate:

- Most important indicator of how much confidence can be placed in your results; directly related to sample size.
  - Traditionally, only 10-60%.
- Poor response rate is potential source of bias.
  - Non-respondents likely to differ from respondents.
  - Can also have response bias (when the answers provided by respondents are invalid).
- Follow-ups or reminders can help increase; incentives too.

▷ Incomplete questionnaires/Missing data:

- Related to questions, length of instrument, respondent.
- Suggestions: place the most important items in the first half to increase response on the important measures; make the instrument the best you can.

Conclusion:

▷ You need plenty of **time**!

- The goal is to gather valid, reliable, unbiased, and discriminatory data from a representative sample of respondents.
- A good questionnaire grows from research hypotheses that have been carefully studied and thought out.
- Discussion of the research problem with colleagues and subject matter experts is critical to developing good questions.
- Questions should be reviewed, revised, and tested on an iterative basis.
- Examining the questionnaire as a whole is an essential element of good questionnaire design.

References & Resources

- *Designing Clinical Research* (3rd ed), Hulley, et al.
- *Encyclopedia of Biostatistics* (© 1998).
- "Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients", Health Technology Assessment, 2001. Vol 5, No 31.
- "What is a survey", Fritz Scheuren (`www.whatisasurvey.info`).
- "How to design a survey" tutorial (`www.statpac.com/surveys`).
- Questionnaire Design lectures by Cathy Jenkins and Mario Davidson.

- Survey Research Shared Resource:
  - Provides services related to the development & implementation of mail, telephone, web & mixed mode surveys.
  - For more info, contact Sara Hollis at sara.hollis@vanderbilt.edu.