Survey Design Workshop

Yale Center for Science and Social Science

Sherlock Campbell, Oriana Aragon, Sara Burke

(with much credit to Betsy Levy Paluck, Melinda Manley and Pia Raffler)

January 31, 2014

This workshop will help you learn the fundamentals on designing a survey. We will cover:

Overview of the survey process
1. Why survey?
2. The preparation phase
3. Design and conceptualization phase
4. Sampling
5. Designing your instrument
6. Piloting & translation
7. Data collection and logistics

How to write questions
8. Writing questions
9. Constructing response scales
10. Formatting
11. Take aways

At the end of this handout are a number of references for you to follow up with specific issues within this enormous subject area.

Please ask questions along the way.
1. **WHY SURVEY?**

Collect data from individuals to answer research questions

- Measure the effect of a program or treatment
- Attitudes—direct measures v. indirect measures; vignettes
- Behavior
- Survey experiments

**Units:** Usually survey individuals or households, but sometimes larger groups can be surveyed as well

**Types of data you can collect through surveys:** Hard data, attitudes, behavior, and knowledge

**Settings vary:** Field versus lab; over phone, face-to-face, mail, internet, etc.

2. **PREPARATION PHASE**

Sampling: collecting data from a subsample of the population

- Target population. Probability and non-probability samples.

**Think about what types of survey:** cross-sectional, longitudinal, experimental—your research question will differ.

**Model your research question:** Really hammer out what variables you are interested in and how you will measure them.

**Human subjects approval:** You *must* get approval from Yale’s institutional review board for any research you do that involves human subjects (even undergraduates must get it). For a summary and how-to, see: [http://www.yale.edu/hsc/Investigator/Student_projects.pdf](http://www.yale.edu/hsc/Investigator/Student_projects.pdf). If you are time pressed, it is often a good idea to submit a preliminary protocol for approval that you can later amend once you have the final version of your survey. The approval process can take 2-3 months, at times longer, so don’t wait until you are ready to hit the field. To make this process easy both for you and the human subjects committee, it is a good idea to be painstakingly organized in your application and do your homework beforehand (for example, special rules apply for working with minors).
**Make a plan.** Steps involved in primary data collection:

![Diagram of steps]

(Wolnock, 2004)

### 3. DESIGN AND CONCEPTUALIZATION

Before you get started, spend time thinking about the exact hypotheses you want to test, and how you want to test them. This is crucial (a) to avoid missing important variables, (b) to avoid gathering data you won’t use, (c) to ensure that your measurements are compatible with each other, and (d) to reduce the impetus for data mining later on.

- A flow chart outlining the mechanisms you are positing (or a formal model)
  What are all the possible mechanisms through which the independent variable “causes” changes in the dependent variable? What are alternative explanations?

- A list of your hypotheses
  What is your dependent variable (outcome variable)? What are the independent variables? How might you measure each of these variables? How can you test your hypotheses and falsify alternative explanations?

- A detailed analysis plan
  This includes the specifications you want to use, the construction of indices, and any heterogeneous treatment effect / subgroup analysis you may want to run. Coming up with a detailed analysis plan prior to data collection
(a) Ensures you gather the data you need in the form you need it, and nothing else;

- Define and clarify variables and figure out what kinds of measurement tools (response scales, categorical answers, etc.) you want. Generally, continuous variables are easiest to work with once you get to data analysis.

(b) Saves you time later on (you have to do it anyways), and

(c) Helps protect you from accusations of data mining.

- **Don’t data-mine:** Be careful about asking a million questions. You will find a statistically significant effect 1 out of 20 times by chance alone (at alpha = .05), even when the true effect is zero. When reporting results, you must report results for all questions you ask. Making analysis plans publicly available prior to receiving data is increasingly becoming a best practice.

---

**Example: Outline your survey content and question objectives**

*Purpose of survey: Survey investigates positive community relations. We think positive community relations are likely to be a function of the following:*

- Cooperation
- Open communication
- Trust

*Within these categories, measurement objectives include:*

- Cooperation
  - Assistance with taking care of children in village
  - Helping out with fieldwork when there is a need
  - Contributing to public goods like schools and roads
  - Protecting community goods like wells

- Open communication
  - Frequency of community meetings
  - Attendance at community meetings
  - Discussions in public areas
  - Visitations to neighbor houses (social network)

---

**Identify method of data collection:** Specify mail vs. person vs. phone vs. internet forms. If person, decide on paper & pen versus digital data collection (PDA, smart phone, netbook, tab) and whether to go through a survey company or hire enumerators directly.
4. SAMPLING

Choosing a sample design
1. Determine your target population

2. Determine your sampling frame
   o The sampling frame is the list of the target population from which the sample will be drawn
     (E.g. a list of seniors enrolled in city high schools at the end of the school year)
     • Might not be the same as the target population: not all elements in the population may be
       represented in the sample frame (telephone numbers of voters versus door-to-door canvassing
       of voters)
     • If possible it should be representative—makes your survey generalizable
     • Representativeness is not always an option: field experiments

⇒ What to do if no reliable sampling frame exists?
   a. Make one: Household listing exercise
   b. Geographic sampling based on GPS coordinates: requires knowing the geographic boundaries
      of the sample area
   c. Improve existing unreliable ones, for example by sitting with village elders
   d. “Random walk”: Draw or obtain a detailed map of the settlement of interest
      (e.g. village), assign routes to enumerators, and instruct them to interview every
      xth household. This method has a number of drawbacks, including giving
      enumerators discretion about where to stop.

   o Inclusion/Exclusion criteria based on who you are hoping to sample. These are key for you to
     know about which population you can draw inferences about. For example a sample randomly
     drawn from (1) all heads of households in villages XY&Z with no access to piped water within
     a 1 km radius, or (2) seniors enrolled at the city high schools who have a transcript available
     and are enrolled in a certain number of classes.

3. Choose your design.
4 basic types of probability sample designs:
   o Simple random sample (SRS): selected by a procedure that gives every element in the population
     an equal chance of being included in the sample.
     • Simple to analyze, minimizes bias
     • If sampling is done in the field, can use a PDA to generate random numbers or pre-printed
       sheets with randomly generated numbers.

   o Systematic random sample: researcher picks a random starting point and systematically selects
     cases from the sampling frame at a specified interval (E.g. list of households, start with the 3rd
     household on the list and sample every 5th household).
     • Sometimes more simple than randomization, representative geographically
     • *We don’t recommend it*: Can be biased if there is a pattern in the ordering of the
       sampling elements, same problems as SRS.


- **Stratified sample**: Entire sampling frame is divided into groups of interest (e.g. racial/ethnic groups) and then use SRS or systematic random sampling within each strata
  - Can be more efficient, more representative, answer more specific questions about subgroups in population.
  - Requires knowledge of population in advance and may introduce more complexity in data analysis.

- **Clustered sample**: Divide sample into clusters (e.g. schools) and randomly sample the clusters. Then you can use SRS, systematic, or stratified sampling within the sampled clusters. Note that if your unit of analysis is the individual but your sampling frame consists of households you need a strategy to randomly sample individuals from within the household.
  - Logistically more feasible in large surveys
  - Need to consider in data analysis.

**Sample Size Estimation**

You need to make sure you have enough people to detect a significant difference between your groups of interest, if a difference really exists. This is known as statistical power. You can use your statistical program to calculate the needed sample size (STATA, SPSS, R…) or one of the calculators available online that take into account your study design and analyses. A very good (free) one is Optimal Design from the University of Michigan: [http://sitemaker.umich.edu/group-based/optimal_design_software](http://sitemaker.umich.edu/group-based/optimal_design_software). Take a look at the extensive manual and documentation. Other calculators are available here: [http://calculators.stat.ucla.edu](http://calculators.stat.ucla.edu).

In order to be able to do your power calculations, you need to know

(a) the statistical significance level at which you want to be able to test your hypotheses,
(b) the number of your treatment arms (if an experiment; note that OD assumes two treatment arms),
(c) the minimum standardized effect size you want to be able to detect – a review of similar studies can greatly help you in figuring out which effect size is reasonable to detect,
(d) if you are conducting a clustered randomized control trial, your expected intra-cluster correlation, and of course
(e) your research and sampling design – panel or cross-sectional data; clustered, stratified or simple sampling?

### 5. DESIGNING YOUR INSTRUMENT

**Check for existing surveys and survey items**

- Validity and reliability of the existing surveys strengthens your own data collection
- What are other people in your small area of research finding? Check for what to expect in terms of response rates, effect sizes, and sample sizes.
- This is also useful information for doing power calculations

**Think about participant burden**: How long can you make the questionnaire and reasonably expect them to fill it out?
We will talk about the details of designing an instrument in the second part.

6. PILOTING

Repeated pilot tests are crucial. Start with interviews, then focus-groups, then individual surveys.
In reality, you will be going back and forth between survey design and piloting many times.

- Open-ended interviews prior to the survey can help you better understand what to measure
- Logistics for interviewers: number of interviewees per day, per interviewer, interviewer fatigue and mischief
- How long does the survey take?
- Check for variation in survey response, consistently skipped questions
- Cognitive interview: ensure understanding, find out how people go about answering the question, put questions in the vernacular
- Compare alternatively worded questions (95% employment v. 5% unemployment)
- Compare against other data sources (ask villagers for number of households in village, compare against census data)
- Ensure you have comprehensive pre-coded response options – having to code many ‘others’ after the survey is unreliable and a pain
- Ensure the answer options make sense to the enumerators / respondents.
- Use piloting to get qualitative feedback on your survey instrument – ask respondents why they answered the way they did, how they interpreted the question etc.

Translation. If your survey is implemented in other languages than it was designed, a good translation is crucial. In general, you always want to implement your survey in a language that both the enumerator (if there is one) and the respondent are absolutely comfortable. If this means translating your survey into several languages, do it. Translate your survey instrument once it is finalized to avoid having to make simultaneous edits in multiple languages.

Phases of a good translation include:
(1) First translation, either by a professional translator or – if your survey team is fluent in both languages, translation in a group can also yield very good results (as people tend to discuss over every formulation and double check each other)
(2) Back translation into the original language by a good translator who has no access to the original survey document
(3) Comparing original and back-translation until all subtle mistakes are corrected
(4) Piloting the translated survey in the field
7. DATA COLLECTION LOGISTICS

Implementing a survey well is a massive, often mundane undertaking. But getting it right is absolutely crucial for ensuring the quality of your data.

Who should conduct the interview?

- Who interviews and potential biases this introduces in response to questions and sampling
  - E.g. race/ethnic group of interviewer, urban interviewer in rural settings, male interviewers unable to interview women in conservative settings, etc.
- You, a team you hire, or a survey firm?
  - Unless you know of a very reliable firm hiring your own enumerators has many advantages
  - If you do go through a survey firm, (i) make sure you are involved in the planning process and ideally reserve the right of firing enumerators -- that’ll give you leverage if things go awry; (ii) be involved in designing the back check and spot check plan and ideally have your own independent auditors as well; (iii) sign a very detailed contract including back check rates, auditing procedures, acceptable error rates, the time frame etc. and make payment contingent on compliance with the contract.
  - A field team often consists of
    - Enumerators, who do the actual interviews
    - Team leaders (1 per 3-6 enumerators) who oversee enumerators, scrutinize their surveys, conduct quality checks, handle logistics, enforce discipline etc.
    - Auditors who conduct back checks and spot checks
    - For large teams a field manager, who handles hiring, logistics etc.
    - If finding respondents is a concern: mobilizers who identify and mobilize respondents 1-3 days in advance.
- Training
  - Be present for or conduct the training sessions.
  - Interviewers should practice on real, non-sample individuals.
  - Invite more enumerators to the training than you will eventually need – people drop out or do not satisfy your criteria.
  - Be very clear about the rules and your expectations. Design and communicate a bonus and repercussion scheme.
  - Depending on context and survey complexity, this process can take 2 days or 2 months. Allow adequate time!

Back and spot checks. Not optional.

- Used to ensure interviewers are following sampling and interviewing protocol.
- Back checks: Re-interview a small random sample of individuals that were interviewed. Select a set of factual questions from the survey to do this where responses are unlikely to change. Part of back checks is also to ensure that enumerators were polite, interviewed the right respondent, read the informed consent and abided by survey protocols.
- Spot checks: In field settings, send auditors out to small samples of interview areas to observe interviewing in real time – unannounced.
- Desk scrutiny: When a survey is submitted by the enumerator, someone (often the team leader) should ensure the survey is complete and answers add up. Ideally this is done the same day and on site in case corrections are necessary. For digital data collection this sanity checks should be programmed into your survey.
- It is often a good idea to make enumerator payment contingent on error rates.

**Quality checks for self-administered surveys**
If you conduct your survey online or have respondents fill mailed surveys, consider including questions that test whether the respondent pays attention.

Examples:
*While watching the television, have you ever had a fatal heart attack? Answer options: Six point scale, anchored on “Never” and “Often” (Paolacci et al., 2010)*

*Instruction: Please respond to the following question with answer choice “B” regardless of the true answer. During the past week, how many times have you eaten red meat?*

**Response rate issues:** What to expect, how to shoot for it. Based on expectations, remuneration, and recruitment methods.

**Replacement rules:** Have clear rules when enumerators are allowed to give up on reaching a respondent in order to avoid winding up with a sample of folks who are easy to reach. For example: Call / visit three times unless refusal to participate or another reason for not participating in the survey (moved, sick, unknown etc.). Have enumerators document each attempt.

**General tips**
- Have a unique ID for each survey.
- Label questionnaires. Do not rely on interviewers to fill in identifying information such as individual name or village. These should be prefilled to eliminate headache later. (e.g. need to later match village name on surveys to state voting data, but interviewer spells the village name differently than it is spelled on the voting data). OR: Have a sheet with village, state ID etc. and have someone scrutinize correct filling of these.
- Data-entry of paper surveys. All data should be double-entered by two different data entry operators in order to reduce entry errors. The two entries are then reconciled and all discrepancies checked and corrected based on the paper survey. Iterate this process until you reach an acceptable error rate (a standard is 0.2%).
- If you hire an entry firm, do your own reentry of a random subset of surveys and calculate the error rate. For long surveys, it is a good idea to break them down into sections. Acceptable error rates should be specified in the contract. For example: A deduction of 5% of payment for a section of the survey (5% of the entire survey) if the error rate in that section is between 0.2 and 0.5%, a discount of 10% of payment for a section and free reentry of said section if the error rate is between 0.5 and 1%, and a discount of 20% on said sections if error rates exceed 1%, as well as the right to terminate the contract immediately.
- You may also want to include discount rates in response to long delays in submitting the data in the contract with the data entry firm.
8. WRITING QUESTIONS

Preliminary Considerations:

- Write questions with your respondent sample in mind. Use plain language and avoid technical terminology or unnecessary complexity and jargon.
- Consider a logical order of questions by subject matter as well as degree of difficulty or sensitivity.

Theory, Method of Analysis, Sources of Measurement Error

These three aspects of research are interrelated, and you should consider each factor before you design your survey.

- Theory: What are you really trying to measure? What information would you like to have if there were no constraints placed upon your inquiry? What aspects of the desired information can be measured or approximated through survey responses? Which information cannot directly be obtained through a survey, and what is the next best survey approximation of such items?

- Method of Analysis: What kind of information are you going to want available for analysis? Consider the analysis you will eventually run.
  - Responses as dependent variables: If you plan on using regression analysis to analyze your data, consider the implications of using responses as dependent variables.
  - Consider variables through latent variable framework for the purpose of theory and analysis. Is your response measure a good proxy for the latent variable?
  - Anticipate your full model: Remember to get information not just on variables that immediately interest you, but also on confounders, variables that are known to be associated with your independent and dependent variables. Look to the literature to find out what other people control or adjust for and operationalize those variables in your survey.
  - Several types of response measures, variables:
    - Open-ended: respondent is not presented with options.
      - Can be useful as starting points in new areas of research or may be desirable in extended format surveys.
      - Are used in some instances to avoid priming subjects or recall cues.
      - Offer content for qualitative analysis and can be coded by relevant attributes for the purposes of quantitative analysis.
    - Closed-ended: Respondent is offered a series of options.
      1. Continuous
        E.g. what is your age? Respondent enters number.
      2. Categorical: discrete groups.
        E.g. what is your age? Respondent chooses category 18-24, 25-34, 35-44
      3. Dichotomous: yes/no or presence/absence
        E.g. do you own a pet?
      4. Ordinal: Type of categorical variable where there is a natural ordering to the
response categories.
  E.g. Political ideology - most to least liberal
5. Nominal: No natural ordering to categories
  E.g. Race/ethnicity
6. Rated: respondents indicate how closely their attitudes correspond to statement.
  E.g. Likert scale – Global warming is a serious problem facing the country.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td>Strongly agree</td>
</tr>
</tbody>
</table>

- Sources of Measurement Error: Anticipate possibility of bias stemming context effects, order effects, framing, priming, etc.

Matching questions to objectives

- Make the objective specific

  **Objective: Alcohol consumption**

  Q: How many alcoholic drinks did you have yesterday?
  Q: How many alcoholic drinks did you in the last seven days?
  Q: How many alcoholic drinks do you have on an average weekday?
  Q: How many alcoholic drinks do you have on an average weekend?
  Q: On days when you drink alcohol at all, how many drinks do you have on average?
  Q: At a meal when you drink alcohol, how many drinks do you have?

  **Objective: Distance from hospital**

  Q: How many miles are you from the hospital?
  Q: How much time does it take you to travel to the hospital? (Think of bus transportation, vs. people who own cars, vs. city dwellers, etc.)

  *(Objective needs to specify what kind of information)*

  All questions above are associated with different kinds of uncertainties and will be understood and answered in different ways by individuals (why we must pilot, see above)

What’s a good question? Some principles and pitfalls to look out for.

1. Design questions that mean the same thing to all respondents. All the terms should be understandable or defined, time periods specified, complex questions asked in multiple stages.

  E.g., Including your visits to psychiatrists, ophthalmologists, and any other professional with a
medical degree, how many times have you seen or talked with a medical doctor in the past two months?

*It is important to specify the meaning of “doctor” because some people may exclude doctors like psychiatrists.

E.g., How often do you feel tired during the day—always, usually, sometimes, rarely, or never?

*To what day are you referring? My answer will drastically change depending on whether I’m thinking of a workday, vacation day, or I’m doing some weird calculus to try to mash all these days together.

* Note that when answer options are scaled, you want to include the answer options in the question.

E.g., “In your household, how are the responsibilities divided up?”

*This is a pretty complex question. Not all respondents will remember to touch upon finances, cooking, inside chores, outside chores, shopping, etc etc. Best to split up the responsibilities and ask about them separately. You also want to define what you mean by household first – and possibly fix ideas having the respondent list all the members in their household.

Or the classic: “What is your income?”

*Over which time span? Monthly, yearly? Formal income? From all jobs? Including stocks, mutual funds, income from other family members? Best to specify and give a rule for rounding, specifying, etc.

How many times have you eaten butter this week?

*Even the term “butter” can be too broad:
When it is: “How many times have you eaten butter, excluding margarine, this week?” the answers changed significantly (Fowler, 1992)

“How many times were you hospitalized this year?”

*The term is not understandable or is confusing (you’d be surprised at the number of words that confuse people):
Correction: “How many times were you admitted to the hospital as a patient overnight or longer this year?”

Piloting your questions ahead of time helps to identify what is clear and to test alternative wordings (see below, piloting), but you should strive to be extremely clear and focused from the start. Always, if you can, use multiple questions for one complex concept.
2. Design all aspects of data collection to minimize possibility that any respondent will feel his or her interests are best served by giving an inaccurate answer to a question. (More on social desirability bias below.)

3. Make clear, and standardize the response task. Clearly define the dimension or continuum respondents are to use in their rating task, give them a reasonable way to place themselves on a continuum, or make clear how long or short of an answer is desired or expected in an open ended question. (More on response scale construction below.)

\[ Q: \text{When did you move to New Haven?} \quad \text{________________________} \]
Potential answers:
A: 1978
A: When I became a grad student
A: After I graduated from college

Instead:
\[ Q: \text{In what year did you move to New Haven?} \]
19\_\_\_
200 \_

For self-administered surveys: Note that some people won’t respect the number of lines you give them for an open ended question (they’ll use the back, ask for another page) but some will respect them too much (they’ll cut their answer short if the line on the page is short.) Give as many lines as you expect length of ideal answer to be. Know that some will go over and under, but the lines communicate your expectation for how much to write.

4. Asking several questions to get at one concept always increases the “validity” of measurement (actually measuring the concept in question.)

\[ \text{E.g., asking questions about the same thing, just in different ways.} \]

\[ \text{Overall, how satisfied or dissatisfied are you with the way democracy works in} \]
\[ \text{Ghana -- very dissatisfied, somewhat dissatisfied, neutral, somewhat satisfied, or very satisfied?} \]
1=very dissatisfied;
2=somewhat dissatisfied;
3=neutral;
4=somewhat satisfied;
5=very satisfied

\[ \text{Choose the statement you agree with more:} \]
A. In Ghana today, we enjoy a real choice among different political parties and candidates.
B. This country is well on its way to becoming a single-party state without real political choice
But beware if you use different kinds of response categories; it is difficult if you want to combine answers into an index. See more below on scales.

5. Questions should ask for firsthand experiences (if accuracy is the objective): Beware of asking for secondhand knowledge, hypothetical questions, asking about causality, or asking about solutions to complex problems.

E.g.: Hypothetical questions. Women who have delivered a baby are better at estimating their probability of using anesthesia than women who haven’t delivered a baby.

E.g.: Causality. We are notoriously bad at explaining our own behavior. See Ross & Nisbett, 1977. Or see the entire field of social psychology circa 1960-2007.

6. Ask one question at a time! Avoid asking 2 questions, imposing unwarranted assumptions, or hidden contingencies.

“Would you like to be rich and famous?”

*This question confuses people who would like one thing and not the other.

“With the way the Iraq war is going, do you think it’s a good idea to send more troops?”

*These questions are “double barreled” or impose an assumption: It asks people to accept its premise that the war is going badly before they respond vis-à-vis the troops. This question is unfair to people like Dick Cheney and Karl Rove.

“Are you unhappy with the way you budget your expenses?”

*You are assuming I have a budget. I may be unhappy with the way I spend my money, but if I don’t have a budget, I don’t know how to answer this question.

“In the past month, did you agree to do something your husband asked because you were afraid of him physically?”

*This question has a hidden contingency: She may have refused to do what he asked, but still been afraid. Or he may not have asked her to do anything.

7. Ask the ultimate question LAST, so it sticks in the mind. Define terms first, and give the response choices last.

NOT:

“Would you say you are not likely, somewhat likely, or very likely to get a job after you graduate?”

“If there was an election today, do you think you would be more likely to vote for Senator Obama or Senator Clinton, considering what’s happened with their campaigns up to this point?”
INSTEAD:
"Which of these categories describe how likely you think you are to get a job after you graduate: not likely, somewhat likely, or very likely?"

"Considering all that has happened with the 2008 presidential campaign up to this point, if there was an election today, do you think you would be more likely to vote for Senator Obama or Senator Clinton?"

People should have the possible answers presented to them right at the end of the question.

8. Measurement is better if respondents are oriented to the task in a consistent way.

Generally—the introduction to the survey is delivered in the same manner. Specifically, an introduction to a question is clear and consistent, e.g., CHOOSE ONE or CHOOSE ALL THAT APPLY. Note that people often ignore directions to "choose all that apply" and think that they are supposed to choose only one. People are bad at reading directions. Make directions stick out or make the choices obvious from the way you format. See formatting below.

8. RESPONSE SCALES

- Response categories
  - Make them logical and meaningful:
    NOT: Many......Some......A Few......Very Few.....None

  - Arrange them well in space:
    NOT:
    Very helpful Not very helpful
    Somewhat helpful Not at all helpful

- Likert scale
  - Strong scales are symmetric, clear and provide meaningful gradations:
    1 Strongly disagree
    2 Moderately disagree
    3 Mildly disagree
    4 Mildly agree
    5 Moderately agree
    6 Strongly agree

OR
  1 Completely true
  2 Mostly true
3  Equally true and untrue
4  Mostly untrue
5  Completely untrue

To have a neutral midpoint or not? An ongoing debate between people who say that some people are too likely to give “non-opinions” or otherwise lazily fill out a questionnaire, and those on the other side who say it is not good to “force” people onto one side or the other when they are truly neutral. You should choose which side you agree with based on your research question (e.g., is neutral a meaningful and important category of response?)

- **Semantic differentials**
  E.g., Rate economists: *(mark an “x” your place on the scale)*
  
  Kind ______ ______ ______ ______ ______ ______ ______ ______ Mean
  Honest ______ ______ ______ ______ ______ ______ ______ ______ Dishonest

  *Rate lawyers:* *(mark an “x” your place on the scale)*
  
  Kind ______ ______ ______ ______ ______ ______ ______ ______ Mean
  Honest ______ ______ ______ ______ ______ ______ ______ ______ Dishonest

- **Thermometers and visual analogues**
  How warm do you feel toward women?
  Very warm ____________________________________ Very cold

  How bad was your pain?
  No pain at all ____________________________________ Worst pain I ever Experienced

**Combining questions into scales**
- Ask multiple questions to get at one “complex concept”—e.g., “empowerment” or “reconciliation” (again, check to make sure that a scale for this complex topic does not already exist—that is, one that is reliable and valid according to previous research.)
- This allows for factor analysis, scale development following the survey.

**Other considerations**
- Ask intimate and demographic questions at the end, after the respondent has “warmed up” to the questionnaire and answered most of the questions. The interview should flow as much as a conversation as possible.
- **Link your questionnaire to other datasets** (GIS, and census, archival datasets)

- Think about what “linking data” you need for each individual or community, and collect it: e.g., zip code, address, GIS parameters. Check with Human Ethics committee first to see if you can retain identifying data.

### 9. POSSIBLE SOURCES OF BIAS

**Biasing effect: memory.** Thumbnail rules for asking questions that draw on memory:

- Memory for daily, mundane events: Recall deteriorates even after 24 hours. When asked for mundane behaviors over longer period of time, (1 week, 1 month, etc) people average, estimate, guesstimate. *Ask for recall of very short period, or ask to keep diary*

- Memory for salient events: e.g., visits to doctor: recall deteriorates within six months.

- Some tips for stimulating memory:
  - Ask multiple questions. *A way of saying “try again”*
  - Ask over the course of two interviews; preview questions for next interview to stimulate reflection
  - Ask related questions. *For asking about doctor’s visit, also ask “Have you bought any medications recently? Have you had to file insurance?”*
  - Provide memory aids, or place them in time. *“Think of last spring. Were you at the hospital when it was warm outside? Here is a calendar you can look at to help you.”*

**Biasing effect: Avoiding looking bad, trying to look good, preserving our self image: In short, social desirability.**

E.g.: voting and having a library card is overreported; drunk driving is underreported.

- Assure confidentiality of responses, communicate this to respondent that protection is in place.

- Interview respondents alone.

- Communicate the priority of response accuracy. Some people have asked respondents to verbally or in writing make a commitment to giving accurate answers; interviewers stress “there are no right answers” “we need to know this because of X goal of research”

- Reduce role of interviewer in data collection process (e.g., use an online data collection process or hand respondents a tablet to complete the survey on privately)

- Use a preamble that minimizes the sense that certain answers will be negatively viewed: “We have found that many people did not vote in this past local election, for many reasons. Did you vote in this past election?” *But be careful; do not create a reverse bias against a positive answer!*
- Allow the respondent to provide perspective on the meaning of the answer using ordering of questions

  Q: Did you go to the doctor in the past two weeks?
  Q: Did you go to the doctor in the past month? (Loftus, et al., 1991)

  People are likely to overreport visits “in the past two weeks” because they want to be seen as “a person who goes to the doctor.” When you allow them a larger time frame first: did you go to the doctor in the past month and then ask the question of interest: did you go to the doctor in the past two weeks you are likely to get a more correct answer.

  Q: In general, would you say you drink more than your friends, less than your friends, or about the same amount as your friends?

  Q: Think about the friend you know who drinks the most. About how many drinks would you say that person has?

  Q: And how about you? On days when you have any alcoholic drinks, about how many drinks do you usually have? (Sudman & Bradburn, 1982)

  Here you’ve allowed the person to give context for their answer (e.g., I run with a heavy/non-drinking crowd)

- Provide a scale that shows acceptance of any answer that is potentially viewed negatively:

  Q: How many sexual partners would you say you have had? None, one, two, three, four, five to ten, eleven to twenty, twenty one to thirty, more than thirty?

  Not, obviously, “one, two, three, more than three?” Or, leave the line blank, make it open-ended. But note that while broader categories (e.g., thirty to fifty partners) are less informative, they are also less stressful for respondents. More on scale below.

**Indirect questions.** Ask questions in a way that keep the individual responses anonymous. Obviously, these approaches only gather population rate, but depending on your analysis an unbiased population rate can be better than a biased individual response. For more details, see Corstange, 2008.

- Indirect “risk” questions. Note this technique only works for a representative sample for which you know population rates for (a) and (b). National Health Interview Survey:

  Q: Is any of these statements true for you?
  a. You have hemophilia and have received clotting function concentrates since 1977.
  b. You are a native of Haiti or central East Africa who has entered the U.S. since 1977.
  c. You are a man who has had sex with another man at some time since 1977, even one time.(etc.)
○ List randomization
Randomly split your sample into two. 50% are asked “How many of the following statements are true for you?” followed by three innocuous statements. The innocuous statements should be chosen such that there are no (very few) people in the sample for whom all three are true and such that you do not know the answers from other parts of the survey. For example:

Group A is asked: How many of the following statements are true for you? Do not tell me which ones.

1) You like apples.
2) You regularly bike to work.
3) Your mother’s birthday is in July.

The other 50% of the sample (Group B) are asked the same question plus a sensitive statement of interest: How many of the following statements are true for you? Do not tell me which ones.

1) You like apples.
2) You regularly bike to work.
3) Your mother’s birthday is in July.
4) The last time you had sex you used a condom.

Since who received which set of questions was randomly assigned, the average number of true innocuous statements is in expectation the same for sufficiently large samples. Subtracting the average number of true statements in group A from the average number of true statements in group B you can infer the average probability of having used a condom during last intercourse in your population.

By also asking the sensitive question directly in Group A and comparing averages between Group A and B you can measure social desirability bias.

○ “Random response” method:

Q. I want you to think – without telling me – of the number of days in the past week in which you have used any marijuana at all. Keep this number in your head. Now please roll this die and add the number you got on the die to the number of days you have used marijuana. What is that sum?

Make sure the interviewer cannot see the die roll. During analysis, subtract the average number of eyes on a die, 3.5, from the responses.

**Biasing effect: Positive and negative bias**
Some people see the questionnaire half full and some see the questionnaire half empty. To avoid measuring people’s personal tendencies to answer “yes” or “no” to questions, try to pose roughly half of your attitudinal questions with a positive slant and half with a negative slant.
E.g., Please rate your agreement with each statement:
“There is a great deal of trust in my community”
“I don’t trust many people.”

If the respondent is going to answer consistently, she will have to agree AND disagree using these two questions. The benefit of this method is that it can also pick up on random answerers—if she agrees to both, you know she is answering randomly (i.e., all yes, all c’s, etc.)

**Biasing effect: Ordering bias**
Sometimes there are distinct sections to your questionnaire, and you think that one may affect the way answers are given to another. For example, one part of my survey asks questions about your mother, and the other asks questions about your romantic partner. Thinking about your mother first may change your responses toward your romantic partner.

*Advice*: Switch ordering for half of respondent sample; test them to see if the responses vary with the order. There are tons of ways to bias a question. See Schwarz & Sudman, *Context Effects in Social and Psychological Research*, 1992; and Sudman & Bradburn, *Asking Questions*, 1982 for a fuller discussion.

10. **FORMATTING**

(from Aday and Cornelius, 2006)

<table>
<thead>
<tr>
<th>Do</th>
<th>Don’t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a. Assign numbers to each question.</td>
<td>Don’t leave off the question number.</td>
</tr>
<tr>
<td>1b. Use letters to indicate subparts of a question when it has more than one part.</td>
<td>Don’t leave off the letter for subparts of a question.</td>
</tr>
<tr>
<td>2a. Do you own a pet?</td>
<td></td>
</tr>
<tr>
<td>Yes……………………………………….1</td>
<td></td>
</tr>
<tr>
<td>No……………………………………….2</td>
<td></td>
</tr>
<tr>
<td>(If yes):</td>
<td></td>
</tr>
<tr>
<td>1b. How many pets do you own?</td>
<td></td>
</tr>
<tr>
<td>Number:</td>
<td></td>
</tr>
<tr>
<td>2. Use a vertical response format for closed-end responses.</td>
<td>Don’t list them horizontally:</td>
</tr>
<tr>
<td>White……………………………………….1</td>
<td>White …1 Black …2 Other …3</td>
</tr>
<tr>
<td>Black……………………………………….2</td>
<td></td>
</tr>
<tr>
<td>Other……………………………………….3</td>
<td></td>
</tr>
<tr>
<td>3. Use numerical codes for closed-end responses</td>
<td>Don’t use alphabetic codes or blank lines to place X or check on, for closed-end responses.</td>
</tr>
<tr>
<td>4. Use consistent numerical codes and formats - e.g. 1 always indicates ‘yes’ on yes/no items, 2 indicates ‘no’, and 8 indicates ‘don’t know’.</td>
<td>Don’t use different codes and formats for comparable responses to different questions; keep it the same!</td>
</tr>
</tbody>
</table>
5. Align response codes:

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>Don’t Know</td>
</tr>
<tr>
<td>1</td>
<td>White</td>
</tr>
<tr>
<td>2</td>
<td>Black</td>
</tr>
<tr>
<td>3</td>
<td>Other</td>
</tr>
</tbody>
</table>

Don’t vary alignment of response codes on a page.

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>Don’t Know</td>
</tr>
<tr>
<td>1</td>
<td>White</td>
</tr>
<tr>
<td>2</td>
<td>Black</td>
</tr>
<tr>
<td>3</td>
<td>Other</td>
</tr>
</tbody>
</table>

6. Provide clear instructions for open-ended items:

What was your blood pressure the last time you had it checked?

**RECORD HIGH VALUE:** ________  
(systolic reading)

**RECORD LOW VALUE:** ________  
(diastolic reading)

Don’t just leave a space with no instructions for the answer.

What was your blood pressure the last time you had it checked?

7. Provide clear special instructions

**(Ask males only):**

Did you use a condom?

Don’t have instructions in the same typeface and format as the question.

Ask males only:

Did you use a condom?

8. Provide clear skip instructions.

8a. Do you smoke cigarettes?

**Yes (Ask Q. 8b) ** 1  
**No (Skip to Q 9)** 2

8b. How many cigarettes do you smoke per day on average?

**RECORD NUMBER OF CIGARETTES:** ________

Don’t leave out explicit skip instructions.

Skip patterns reduce respondent burden!

9. Phrase full and complete questions.

What is your age? _____

Don’t simply use words or headings.

Age? _____

10. Use a forced-choice format for a list.

Should an employer be allowed to require job applicants to be medically tested for … (**circle answer for yes or no to each**.)

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Sexually transmitted diseases (STD’s)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>b. Using illegal drugs?</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>c. High blood pressure?</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

DON’T use a “check all that apply” or “circle all that apply” format.
11. Use a column format for a series with the same response categories.

12. Use a column format for a series with comparable skip patterns. Don’t fail to clearly link a series of questions to subsequent dependent items.

13. Put all parts of the question on the same page. Don’t split a question between pages.

14. Allow plenty of space on the questionnaire. Don’t crowd the questions and space for recording the answers.

15. Carefully consider the overall appearance of the questionnaire. Don’t just start the questions on page 1 without introducing the study, identifying the sponsoring organization, and so on.

16. End the questionnaire with a thank you to the respondent.

17. Consider how the data will be processed

11. TAKE AWAYS

- Gathering high quality primary data is a major undertaking.
- Work backwards: Write your analysis plan before the survey.
- No good inference without a good sample.
- There exist lots of best practices and rules of thumbs for writing and formatting surveys. Use them. Use questions that have been shown to work.
- Get out there. A good survey is rarely written at a desk. Pilot, tweak, pilot, tweak, pilot…
- Translation matters.
- Enumerators, survey companies and data entrants: Trust is good, control and the right incentives are better. Interviewing is hard work. Have tight oversight and checks in place to ensure high data quality. Think about setting the right incentives.

References


Morgan, David L. 1997. Introduction; Focus Groups as a Qualitative Method; and The Uses of Focus Groups, Ch. 1-3 in *Focus Groups as Qualitative Research*. Thousand Oaks, CA: Sage Publications.


