

*turning knowledge into practice*

# Survey Design and Methodology

Rachel A. Caspar

August 30, 2005



*RTI International is a trade name of Research Triangle Institute*

3040 Cornwallis Road  
Phone 919-541-6376

■ P.O. Box 12194

■ Research Triangle Park, North Carolina, USA 27709  
Fax 919-316-3166

e-mail: [caspar@rti.org](mailto:caspar@rti.org)

# What is a Survey?

- A systematic method for gathering information from (a sample) of entities for the purpose of constructing quantitative descriptors of the attributes of the larger population of which the entities are members.
  - The entities may be individuals, businesses, or other types of establishments
- Survey methodology seeks to identify principles about the design, collection, processing, and analysis of surveys that are linked to the cost and quality of the survey estimates.

# Error

- Definition

The act, assertion, or belief that unintentionally deviates from what is correct, right, or true; a mistake

- “Error” is a negative term

- Errors on a test
- Errors on your tax form
- Fatal error on the computer
- Baseball error

- Fact of life in survey research

- Survey Error: Discrepancy between quantitative statements made from the survey data and reality

# Steps in the Survey Process

- Frame Construction
- Sample Design
- Questionnaire Design
- Data Collection
- Data Processing
- Analysis and Report Preparation

# Frame Construction

- Sampling frame: the list of units that comprise a population of interest from which a sample is drawn
- The representativeness of sample survey findings is dependent on the completeness of the initial sampling frame
- But, there are rarely perfect frames...

# Problems with Frames

- Error sources:
  - Omissions
  - Duplicates
  - Inclusion of non-population elements
- All elements in a sampling frame must have a known probability of selection (and preferably an equal probability)

# Sample Design

## Selecting a Representative Subset

- Probability sampling
  - Simple random
  - Systematic
  - Stratified
  - Multistage cluster
  - Probability proportionate to size
- Non-Probability Sampling
  - Purposive
  - Quota
  - Street corner/intercept surveys
  - BOPs

# Sampling Error

- Error due to measuring a subset of the target population.
- Function of the natural variability
  - Sample design including sample size, stratification, clustering, selection methods, etc.



# Questionnaire Design

- Questionnaires are the tools used to collect information to answer the research questions that interest us
- At a basic level questionnaire design involves:
  - Concept development
  - Operationalization of the concept in the form of measurable indicators
  - Construction of text wording that will collect the information in a valid and reliable manner

# The Design Path

- Delineate the scientific objectives
- Gather background literature, existing questionnaires, input from experts
- Prepare an analysis plan relating the concepts of interest
- Helpful to put together table shells, important correlations, comparisons
  - Questions must be designed to meet analytic needs
- Translate concepts into questions (measures)
- Evaluate questions for basic questionnaire design errors
- Evaluate questions for fit to concepts

# Characteristics of “Good” Questions

- Questions need to be consistently understood; both across respondents and in a manner consistent with what the researcher intended
- Questions need to be consistently administered or communicated to respondents
- What constitutes an adequate answer should be consistently communicated
- Unless measuring knowledge is the goal of the question, all respondents should have access to the information needed to answer the question accurately
- Respondents must be willing to provide the answers called for in the question

# The Role of Standardization

In survey research,

*“The goal of standardization is that each respondent be exposed to the same question experience, and that the recording of the answer be the same, too, so that any differences in the answers can be correctly interpreted as reflecting differences between respondents rather than differences in the process that produced the answer.” (p. 14)*

Fowler, F.J., and Mangione, T.W., (1990), Standardized Survey Interviewing: Minimizing Interviewer-Related Error, Thousand Oaks, CA: Sage.

# Implicit Assumptions

- That interviewers adequately fulfil their role.
- That the wording of questions constitute a complete and adequate script such that when respondents hear or read the question, they will be fully prepared to answer it.

# Components of the Response Process

- Comprehension
- Retrieval
- Judgment
- Response

# Comprehension

- What does the question mean?
  - Literal meaning
    - Example: During the past two weeks have you felt blue or down?
  - Pragmatic meaning
    - Example: When did you begin work at RTI?
- Does the respondent's understanding coincide with what the researcher intended?

# What Makes a Question Difficult to Comprehend?

- Unfamiliar or poorly defined terms
- Complex grammar
- Ambiguous clauses
- Vague or ambiguous terms
- Intended use of the question



# Retrieval

- Generate retrieval strategy
- Retrieve specific, generic memories
- Fill in missing details

# What Affects the Ease with Which Information Can be Retrieved?

- Distinctiveness and/or importance of the event
- Number and utility of cues provided by the question
- Degree of fit between event(s) stored in memory and the question asked
- Source of the memory
- Length of time since events occurred

# Judgment

- Assess completeness and relevance of retrieved memories
- Draw inferences based on accessibility of memories
- Integrate material
- Make estimate based on partial retrieval

# Response

- Mapping the retrieved information onto the response options provided
  - Open vs. closed question
  - Degree of specificity needed
- Tweak the retrieved information based on:
  - Desire to please researcher/interviewer
  - Social desirability
  - Desire to appear like “others”
  - Opinion of researcher’s right to know
  - Concerns about confidentiality

# “Opportunities” for Error

- Comprehension
  - Respondent’s understanding is not consistent with that intended by researcher
  - Intent of question is understood differentially across respondents
  - Lack of attention to important qualifiers
- Retrieval
  - Request for information from too long ago causes details to be fuzzy or incomplete
  - Events that are somewhat similar get lumped together

# “Opportunities” for Error (cont’d)

- Judgment
  - Estimation strategies rather than exact counts
  - Events telescoped into the reference period
  - Satisficing

# “Opportunities” for Error (cont’d)

- Response
  - Respondent alters retrieved data to meet some other goal
  - Embarrassing events may be excluded
  - More positive events may be double counted
  - Available response categories are not sufficient or are not easily linked to the retrieved information
  - Example: How do you feel about your present weight?
    - Overweight
    - Underweight
    - About right
  - Response categories provide information to respondents

# Example of Informative Response Categories

0

1-2

3-5

6-10

10 or more

0

1-20

21-40

41-60

61 or more



# Translation

- Consider the need for translating your instrument into other languages
  - The nonresponse due to language barriers may not be inconsequential
- Involve translators early in the design process so that problematic concepts/questions are identified early

# Evaluating Survey Questions

- Expert review
- Focus groups
- Cognitive laboratory methods
- Field pretests
- Behavior coding
- Experimental field tests
- Debriefings
- Analysis of the data

# Modes of Data Collection

- Self-administered questionnaires (SAQs)
  - Mail surveys
  - Self-administered answer sheets
  - Computer-assisted self interviewing (CASI, ACASI, TACASI)
  - Web surveys
- Questionnaires administered by an interviewer
  - Face-to-face using either paper-and-pencil (PAPI) or computer (CAPI)
  - Telephone interviewing (CATI)
- Mixed mode approaches

# Additional Data Collection Modes

- Interactive Voice Response
  - Touch tone data entry
  - Voice recognition entry
- Hand-held devices (Palm, pocketPC, etc.)
- Computer-assisted recorded interview (CARI)
- Web-enabled panels

# Considerations in Choosing a Mode

- Cost
- Quality of data
- Response rates
  - Both at the unit and item level
- Timeliness
- Questionnaire content, length, and complexity
- Population of interest
- Frame availability for sampling
- There is no one “best mode”

# Reasons for Utilizing a Mixed Mode Design

- May reduce costs if more expensive modes are only utilized for those cases that could not be interviewed via a less expensive mode
- To maximize response rates by allowing respondents a choice of how they provide their data
- Reduce costs in longitudinal surveys
- The goal is to exploit the advantages of one mode while neutralizing the disadvantages through a combination of methods

# A Caution for Mixed Mode Studies

- Instruments and procedures need to be designed to ensure equivalence across modes
  - This may mean that all of the benefits of a particular mode will not be realized
  - Example: data collection using both a mail survey and web survey
- The goal is to keep the essential survey conditions as similar as possible across modes

# Different Modes...Different Results

- National Survey of Drug Use and Health
  - Face-to-face
  - Interviews conducted in the home
  - CAPI/ACASI
- Monitoring the Future
  - Group administration
  - School-based survey
  - Paper-and-pencil answer sheets



# Data Processing

- Manual Editing – process of preparing paper questionnaires for keying; resolution of inconsistent data and/or inappropriately answered questions
- Coding
  - Open-ended responses
  - “Other, specify”
  - Medical diagnoses
  - Industry and Occupation
- Keying
- Machine Edits

# The Challenge for Survey Research

- Rather than discrete steps, survey research is a process
  - Decisions made early have implications for the remainder of the process
- Balancing the trade-off between survey error and survey cost
  - No single solution will be right every time

# To Conclude: The Full Process

- Specify the scientific objectives
- Choose the population
- Delineate the required information
- Determine the number of observations
- Design and select a sample
- Develop measurement methods
- Develop measurement instrument
- Devise the measurement protocol
- Pretest the procedures
- Hire and train staff
- Locate sample members
- Secure participation
- Make and record measurements
- Transmit data
- Code and edit data
- Create analysis files
- Calculate sampling weights
- Adjust for missing data
- Analyze data
- Draw conclusions
- Prepare reports / documentation