

Class 3. Surveys

Objectives:

At the end of this class the students will know

- ◆ what are surveys and why they are conducted
- ◆ steps included in conducting a survey
- ◆ main types of study designs
- ◆ types of data collection methods
- ◆ probability and non-probability sampling methods
- ◆ identifying the main products of surveys

As we have learned from the previous class, in order to plan effectively, we should have information about the current situation, problems of interest, etc. Unfortunately in real life the information is often not available, inaccurate, out-of-date, and/or incomplete. This is particularly true for developing countries because of underdeveloped data collection infrastructures and poor surveillance systems. When data is absent, in order to be able to make informed decisions and considering prioritizing, planners need to gather and analyze necessary data prior to planning.

Depending on the type of the data needed, *qualitative* or *quantitative* research methods or both can be used. Quantitative research answers to the question “how many” or “how often” (whereas qualitative research provides data, answering to the question “why”). Surveys are widely used in quantitative research. They are important sources of information for health care policy-makers, public health professionals, private providers, and health care consumers concerned with the planning, implementation and/or evaluation of health-related programs and policies.

Surveys systematically collect information on a topic by asking selected individuals questions about a variety of factors that influence, measure, or are affected by people's health in order to generate statistics on the group(s) represented by those individuals.

Any survey has the following key dimensions:

- ◆ a research topic or problem of interest is clearly delineated
- ◆ information on the issue is gathered by asking individuals questions
- ◆ the data collection process itself is systematic and well defined
- ◆ the purpose of the study is to generate group-level summary statistics
- ◆ the results are generalizable to the group(s) represented by the individuals included in the study

A properly conducted survey can give planners the information they need at moderate cost, and known precision.

For the reasons of time, resources, efficiency and precision, measuring the variable using a sample of the population rather than measuring the whole population (i.e. census) is usually preferred. Survey data are often preferred over census data because a complete enumeration (census) will often suffer from underreporting, the inability to obtain data of the uniformly high quality from a vast army of data gatherers, and errors, delays in processing masses of data.

Thus, surveys

- ◆ are useful for describing the characteristics of a large population
- ◆ are comprehensive (can ask numerous questions, which increases the flexibility of analysis)
- ◆ can be customized to fit your needs
- ◆ use the standardized approach which strengthens analysis
- ◆ have versatile forms (personal interview, telephone interview, direct mail, etc)

Surveys are done to

- ◆ identify existing problems
- ◆ establish baseline levels of health indicators (to know where we are now)
- ◆ define objectives and measure their achievements

In general, any survey consists of the following steps:

1. Choosing a topic for health survey
2. Matching the survey design to survey objectives
3. Defining and clarifying the survey variables
4. Planning the analysis of the survey data
5. Choosing the methods of data collection
6. Drawing the sample
7. Formulating the survey questions
8. Formatting the questionnaire
9. Monitoring and carrying out the survey
10. Preparing the data for analysis
11. Implementing the analysis of the survey data
12. Writing the research report

Each of these steps is a topic for separate extensive discussion, which is out of the scope of this introductory course. Therefore we are going to provide you with general information about some key points related to surveys, which we think are important to know for anyone who is involved in the health care planning process.

1. Types of the study design

In general, epidemiological studies can have *experimental* or *observational* designs. They are distinguished with respect to whether the major factor of interest in the study is under the control of the investigator or not.

In *experimental* studies, the investigator actually introduces a factor or intervenes in the environment of the study subjects to see what impact the intervention has on the study subjects compared to a group of subjects that didn't have the intervention. In *observational* studies the investigators do not directly intervene, but instead, develop methods for describing events that occur naturally without direct intervention and effect on study subjects.

There are three major types of observational study designs - *cross-sectional*, *group-comparison*, and *longitudinal*. These designs differ principally with respect to (1) the number of groups included in the study

and the criteria for choosing them and (2) the number of points in time and reference period(s) for gathering the data.

The distinguishing features of the three types of observational designs are presented in the table below.

Table 1. Types of study designs (L. A. Aday "Designing and conducting health surveys")

Characteristics				
Types of study design	Groups		Time periods	
	# of groups	Criteria for the selection of groups	# of periods of data collection	Reference periods for data collection
<i>Observational</i>				
Cross-sectional	1	Population of interest	1	Present (and recall of past)
Group-comparison (case-control)	2+	Population subgroups with and without characteristics of interest	1	Present and recall of past
Longitudinal (Prospective)	1 or 2+	Population or subgroups that are and are not likely to develop characteristics of interest	2+	Present and Future

Cross-sectional studies are conducted at a given point in time, and are used to characterize generally a single group representative of some population of interest. The reference period for characteristics that study subjects are asked to report may, however, be either for that point in time or for some reasonable period of time that they can recall in the past. They, for example, are appropriate to determine the prevalence of specified morbid conditions, or to reveal current patterns of health services utilization.

Cross-comparison studies explicitly focus on two or more groups. The groups are chosen based on the criterion that one has a characteristic of interest and the other does not. As is the case with the cross-sectional design, data are collected at one point in time, and the reference period for asking study subjects questions could be either the present or some period of time in the past. With analytical group-comparison designs (called "case-control or retrospective studies" in epidemiological studies) there is an effort to explicitly look back in time at the factors that could have given rise to one group having the characteristic (for example, a particular disease) and the other not having the characteristic.

Longitudinal studies, in contrast, assess changes over time. They employ procedures for repeated sampling of a population. Longitudinal studies focus on a population or subgroups, some members of which will be exposed to or experience certain events over time, while others will not; and data are collected at more than one point in time, the reference period is prospective rather than retrospective. Longitudinal designs can be *cohort studies* or *panel studies*. In the cohort approach a new sample of individuals is selected randomly for each set of measurements (cohort studies of given age groups are commonly used by demographers). In the

case of the panel approach the sample selected initially (same individuals) is followed over time. A panel study allows maximum explanation of change over time, but they are expensive and highly time-consuming as well as possibly produce serious problems of non-response through attrition. Panel studies require complex statistical analysis. To avoid these problems and still capture many of the mentioned benefits, surveyors frequently attempt in a single respondent contact to *reconstruct retrospectively* the history of interest, for example through a pregnancy history. In this instance the validity of the data will be subject to recall problems.

2. Types of data collection methods

There are three principal methods of gathering data in surveys.

- ◆ self-administered surveys
- ◆ personal interviews
- ◆ telephone interviews

Self-administered surveys/questionnaires provide inexpensive means of broad geographical coverage in a survey, provided the literacy of the population and an effective postal service or other means of questionnaire dissemination is available. However, care needs to be taken to make completing and returning questionnaires easier for respondents, i.e., enclose a self-addressed stamped envelope, be aware of postal regulations—size, closure, etc; prepare clear instructions, understandable for respondents; follow-up on non-returned questionnaires, mailing second time, phone to remind if necessary. It is also necessary to keep track of response rates, track trends in information based on the time of a response – e.g. it is likely that later respondents are in poorer health (assume non-responders are in poorer health too), whereas early respondents are those most interested in a topic. The response rate is a guide to representativeness; this is necessary to achieve at least 50 % response rate; 70% response rate is considered very good (but usually response is lower). In general, a self-administered survey is easy to administer; is good for sensitive topics (since it is more anonymous than some other types of surveys), and is less expensive than face-to-face interview.

Personal (face to face interviews) guarantee, or at least maximize, the likelihood of response (if done correctly they provide an 80-85% response rate) and minimize the number of incomplete questionnaires. An interviewer can use visual aids, face-to-face clarification of questions helps with confusing questions, and decrease “don't know” responses. This method is better for dealing with complicated (but not sensitive) issues. Also the interviewer can make observations that might be pertinent and increase the quality of data. On the other hand, personal contact may inhibit honesty and objectivity in response. Further approach requires research team (well-trained interviewers capable of producing reproducible responses), and field supervision. Interviews are expensive also due to time spent on travel.

Telephone interviews save money and time. The respondent may find it easier to answer sensitive questions anonymously. Telephone interviews are also safer for both an interviewer and respondent. However, telephone interviews are limited to those who have phones. There is less rapport between an interviewer and a respondent on telephone interviews and this possibly may increase suspicion as to how the data will be used. In addition, telephone interviews are affected by the time of a call, respondents ending interview by hanging up, and answering machines used to screen calls.

3. Sampling

A highly important question related to health surveys is how to draw a sample for them. Before discussing this, we would like to provide you with some key definitions.

Sample universe

The population that the results of the survey are intended to describe.

Sample frame

Actual list from which elements are selected.

Sample element (unit)

The ultimate individual or unit from which data are collected [for example, persons: residents of the defined area, health personnel or inanimate objects (e.g., hospitals or medical records)].

Sample

A subset/group of elements drawn from a sample frame for the purpose of making statements about the population that the frame represents.

Probability sampling

A system for drawing samples from a population such that every possible sample has a known (but doesn't have to be equal) probability of being drawn.

Parameter

A summary constant that describes one characteristic of a population.

How do we sample?

In general sampling approaches can be divided into two large groups - *probability* and *non-probability sampling*.

a) *Probability sampling* approaches include:

1. Simple Random Sampling (SRS)
2. Systematic Sampling
3. Stratified Sampling
4. Multi-stage sampling (i.e. cluster sample)

Simple Random Sample: offered by drawing elements out of a "hat" randomly or by using a random number generator (one can use a currency bill for picking a random number). But for SRS we need a complete list including all elements! This may be impossible. For example, as in a national survey where a huge list will be needed.

Systematic Sample: every element should be listed; the first element should be drawn randomly, then every k^{th} (for example, fourth) element should be chosen from the list. Make sure that a list is not periodic!

Stratified Sample: this technique implies sampling from a predetermined grouping (i.e. strata) that, as suspected, will vary for the measure of interest (the population can be stratified by age/level of education,

etc); within the strata SRS or systematic sampling is used. Stratified sampling ensures heterogeneity of a sample and allows for comparisons between strata, if sample size (n) for each strata is sufficient.

Multi-stage sampling (i.e. cluster sampling) is used when above mentioned methods aren't feasible (i.e. lists of all elements aren't available; when we are dealing with wide area). It includes the following steps: (1) select a "cluster" of elements (e.g. hospitals or city blocks), then (2) select elements (e.g. physicians or households). This sampling method is cost-effective and easy to administer.

b) Non-probability sampling approaches include

- ◆ Purposive or judgmental approach, based on an educated guess of representative units (we are interested in studying the attitude of teenagers toward smoking, and draw a sample, so we visit the school or a club for teenagers);
- ◆ Quota sampling (select elements as you want following a pre-set quota pattern - for example, 5 males, 10 females);
- ◆ Available subject sampling (whoever is available included, asking anyone at a bus stop to answer to your questions).

It is highly important to remember that the samples we get as a result of non-probability sampling method are not truly representative. There is no objective way to use the data gathered through these techniques, to estimate population parameters these results cannot be generalized to the study population.

Considerations of sample size

One of the most important questions any planner faces is "How large of a sample is needed/how many units need to be sampled?" This depends on many different things – what we intend to estimate (the prevalence (p) of the particular condition in a defined population or to make a judgment as to whether two population means or proportions differ by a stipulated amount); how precise an estimate we are interested to obtaining and the type of data that we have (discrete, continuous, etc), and so on.

The detailed explanation of how to calculate sample size is out of the scope of the current course, however, if necessary, you can consult corresponding literature. In real life, a survey statistician should be included in planners' group to give a professional advice regarding sampling issues.

4. Survey products

The main products of survey include:

1. Data
2. Staff development (broad understanding of the health process, new technical skills gained by the organization staff, government health officials, and community members, new sensitivity to the concerns and needs of the communities)
3. Community relations (non-clinical point of contact between providers and clients)

In this class we have touched many important issues related to survey methodology. Certainly, many of them are complex and are topics that need separate attention, beyond the objectives of today's lessons. Nevertheless, the importance of survey information for planning suggests that anyone involved in the planning process should have a general understanding of the survey concepts and principles discussed above.

Questions for Self Assessment

1. Please choose the corresponding definition for each term (drag and drop).

1. Sample universe
 2. Sample element (unit)
 3. Probability sampling
 4. Sample
 5. Sample frame
-
- a) A system for drawing samples from a population such that every possible sample has a known (but doesn't have to be equal) probability of being drawn
 - b) The ultimate individual or unit from which data are collected (for example, individuals, households, families, institutions, etc.)
 - c) Actual list from which elements are selected.
 - d) A subset/group of elements drawn from a sample frame for the purpose of making statements about the population that the frame represents.
 - e) The population that the results of the survey are intended to describe.

Answer: 1 e, 2 b, 3 a, 4 d, 5 c. See page 5.

2. Is the following statement correct?

“Self-administered questionnaires are better for dealing with complicated (but not sensitive) issues”.

True False

Answer: False. See page 4.

Face-to-face interview is better for dealing with complicated (but not sensitive) issues.

3. Is the following statement correct?

“In observational studies investigators do not directly intervene, but instead develop methods for describing events that occur naturally without direct intervention and their effect on study subjects”.

True False

Answer: True. See page 2-3.

4. What type of study is defined below?

“It is conducted at the point in time and used to characterize a population in certain respects at the time.”

- a) longitudinal
- b) cross-sectional

Answer: b) See page 3.

5. Which sampling approach requires availability/ existence of a complete list including all elements?

- a) Multi-stage sample (i.e. cluster sample)
- b) Systematic Sample,
- c) Simple Random Sample (SRS),
- d) Stratified Sample

Answer: c) See page 6.

6. Which sampling approach is described below?

“...Every element should be listed; the first element should be drawn randomly, then every “Kth” (for example, fourth) element should be selected from the list (make sure that list isn’t periodic!)”

- a) Multi-stage sample (i.e. cluster sample),
- b) Systematic Sample,
- c) Simple Random Sample (SRS),
- d) Stratified Sample.

Answer: b) See page 6.