DEM 2110
SOURCES AND MEASURES OF DEMOGRAPHIC DATA

Sample Surveys

2020/2021

Outline

- Definition and historical background of sample surveys
- Types of surveys
- Planning of sample surveys
- Advantages and limitations of surveys
- Errors in surveys and their sources

Introduction

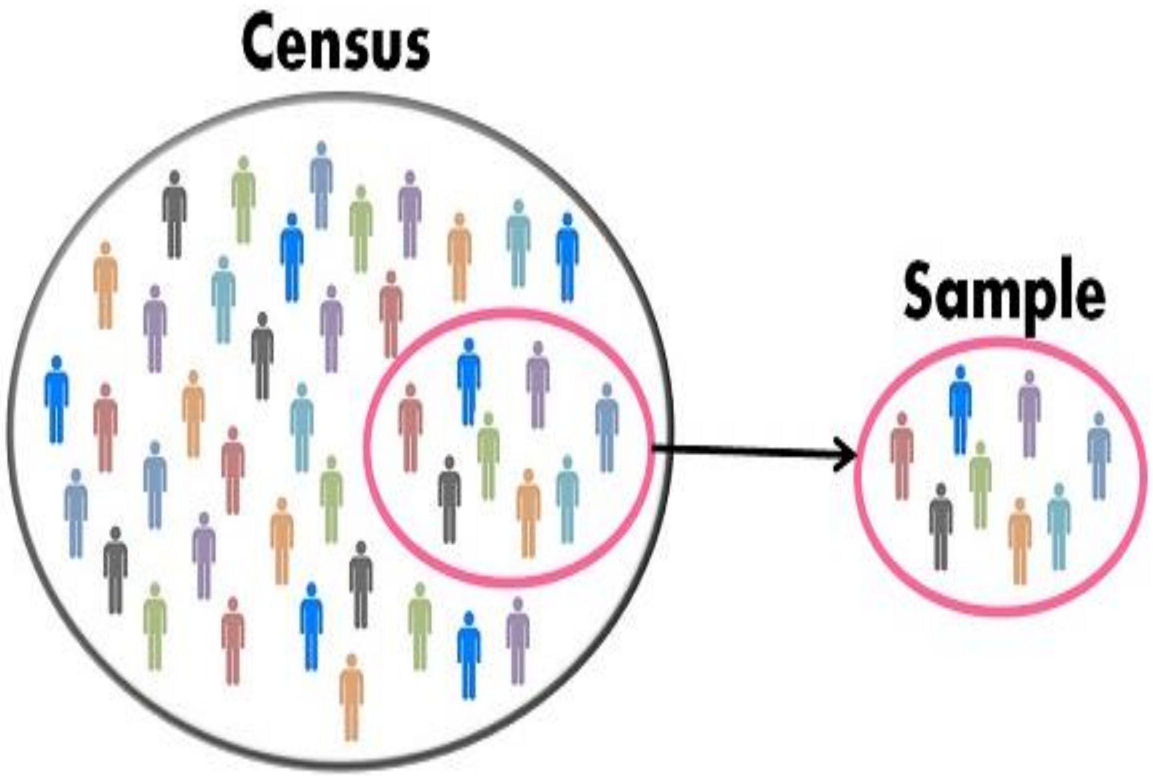
- Sample survey is one of the main sources of demographic and socio-economic data.

- It is not always that we need to collect data about the entire population, i.e. through a census.

- Sometimes we just need to collect data from a subset of the population in order to understand the entire population, and to meet the **additional** and **emerging data needs** on a continuous basis.

- In a sample survey, **part of the population is selected** from which **observations are made or data are collected** and then **inferences are made** to the **whole population**.

Census vs sample



Early developments of survey sampling

ÅTo fully understand the historical development of survey sampling, there is need to understand: history of official statistics; history of sampling; and history of the representative method

ÅIntuitive application of the principles of sampling in science has been taking place for a long time.

ÅGraunt (1662) and Laplace (1812) applied the principles of sampling in an intuitive way, but there was no proper scientific foundation.

ÅIt was not called sampling but inductive reasoning.

John Graunt (1620-1674)

- In 1662, Graunt made the first known attempt to make statements about the population using only information about part of it

- In his writing, Graunt describes a method to **estimate the population of London based on partial information**

- He surveyed families in a sample of parishes where the registers were well kept

- Graunt found that on **average** there were **3 burials per year in 11 families**
  - He **assumed** this ratio was constant for all parishes

John Graunt (1620-1674)

- Knowing the **total number of burials per year** in London to be about **13 000**, Graunt concluded that the total number of families was approximately **48 000**

- Putting the average family size at **8**, he estimated the population of London to be **384 000**

John Graunt (1620-1674)

Graunt invented **two concepts** that are even today important for survey sampling

1) He observed and then applied the fact that some social and demographic indicators remained stable in time and space

ÅThis stability is an essential assumption, without which making inferences based on data collected by social surveys would not be justified.

Å***Example***: *He discovered that nearly the same proportion of boys and girls were born, though slightly more boys*

2) He used averages to estimate total values

ÅEssential to his method was the observation that the **proportion of burials** in a year remained around **3 to 11 families**, and that the **average** family size was **8 persons**

Pierre Simon Laplace (1749-1827)

- The second time a survey-like method was applied was more than a century later.
- Laplace realized that it was important to have some indication of the accuracy of the estimate of the French population
- He implemented an approach that was similar to that of John Graunt.
- He selected 30 departments distributed over the area of France

Pierre Simon Laplace (1749-1827)

ÅTwo criteria controlled the selection process;

1)He saw to it that all types of climate were represented

ÅIn this way, he could compensate for climate effects

2)He selected departments for which the mayors of the communes could provide accurate information

ÅHowever, Laplace used a cluster sample instead of a simple random sample, and communes were selected purposively and not at random

Historical background of sample survey

- Many scientific results are based on observations in just a few experiments. Apparently, it was possible to generalize these experimental results

- Although inductive reasoning has been commonly applied both in everyday life and in science for a long time, sampling as a well-defined statistical method is fairly young
  - It was invented together with the development of other statistical methods
  - Its history, as a well defined statistical method, started in the year 1895 and it took some 50 years before the principles of probability sampling were widely accepted
  - Credit is given to one man, Anders Kiaer who was director of the Norwegian Statistical Bureau.

Anders Kiaer (1838-1919)

- Kiaer started the process that ended in the development of modern survey sampling theory and methods

- He was the first to use a sampling survey on its own

- The survey was an investigation carried out **throughout Norway** on a proposed retirement and sickness insurance scheme

Anders Kiaer (1838-1919)

- He used enumerators to fill out 120 000 forms about the adult population **according to rules** he had laid out
  - **80 000 forms** were collected by the **representative method** and **40 000 forms** by a **special** method areas where the working class people lived
  - For the 80 000 respondents, households in Norway were divided in two strata based on the 1891 census
  - 20 000 were selected from cities and the rest from rural areas
  - Further selections were made following laid down rules

Anders Kiaer (1838-1919)

- Kiaer's representative method resembles the methods of modern sampling theory
    - The main difference is that there was no explicit random selection mechanism but some kind of purposive sampling
    - The major innovation in Kiaer's method, was that the variation in population was considered an essential characteristic

Anders Kiaer (1838-1919)

ÅThe three important principles in Kiaerõs approach were:

1) Importance of the **representativeness** of the sample;

2) The **selection of elements** for observation should, insofar as possible, be made **objectively** (subjective judgment of enumerators should not be allowed); and

3) For every survey, the **reliability** of the results should be assessed.

Historical background of sample survey

- Contributions by different scholars led to development of survey sampling as we know it today.

- The UN Statistical Commission established the Sub-commission on Statistical Sampling in 1947.

- The work of the Commission consisted of drawing up standards that would assist national statistical institutions in improving their statistics by using modern sampling procedures.

- The first publication was a paper on òThe Preparation of Sampling Survey Reportsó, see UN Statistical Commission (1950).
  - This paper promotes probability sampling and still provide a useful guide for achieving clarity, comprehensiveness, and international comparability in sample survey reporting.

What do you observe in each table?

Table A.1 Distribution of residential households by provinces and type of residence

Province	Residential households			Percentage	
	Urban	Rural	Total	Provinces	Urban
Central	76,002	198,744	274,746	9.8	27.7
Copperbelt	336,672	90,217	426,889	15.2	78.9
Eastern	47,371	295,534	342,905	12.2	13.8
Luapula	44,254	199,656	243,910	8.7	18.1
Lusaka	422,029	92,051	514,080	18.3	82.1
Muchinga	26,585	127,665	154,250	5.5	17.2
Northern	44,296	196,260	240,556	8.5	18.4
North Western	31,460	110,464	141,924	5.0	22.2
Southern	79,551	206,791	286,342	10.2	27.8
Western	27,196	163,099	190,295	6.8	14.3
Zambia	**1,135,416**	**1,680,481**	**2,815,897**	**100.0**	**40.3**

Source: The 2010 CPH conducted by the Zambia Statistics Agency.

Table A.3 The 2018 ZDHS sample allocation of SEAs and households by provinces and type of residence

Province	Number of SEAs allocated			Number of households allocated		
	Urban	Rural	Total	Urban	Rural	Total
Central	18	37	55	450	925	1,375
Copperbelt	37	24	61	925	600	1,525
Eastern	15	47	62	375	1,175	1,550
Luapula	15	39	54	375	975	1,350
Lusaka	42	24	66	1,050	600	1,650
Muchinga	12	33	45	300	825	1,125
Northern	15	38	53	375	950	1,325
North Western	13	30	43	325	750	1,075
Southern	19	38	57	475	950	1,425
Western	12	37	49	300	925	1,225
Zambia	**198**	**347**	**545**	**4,950**	**8,675**	**13,625**

Table A.4 The 2018 ZDHS sample allocation of expected completed women and men interviews by province and type of residence

Province	Expected number of interviews with women age 15-49			Expected number of interviews with men age 15-59		
	Urban	Rural	Total	Urban	Rural	Total
Central	464	758	1,222	402	705	1,107
Copperbelt	954	492	1,446	827	457	1,284
Eastern	387	963	1,350	336	896	1,232
Luapula	387	800	1,187	336	743	1,079
Lusaka	1,083	492	1,575	940	457	1,397
Muchinga	309	676	985	268	629	897
Northern	387	780	1,167	336	725	1,061
North Western	335	615	950	291	572	863
Southern	490	780	1,270	425	725	1,150
Western	309	758	1,067	268	705	973
Zambia	**5,105**	**7,114**	**12,219**	**4,429**	**6,614**	**11,043**

TYPES OF SURVEY

ÅIntroduction

ÅMany countries have in place both ***periodic*** and ***ad hoc*** surveys.

ÅThe household survey programme can be part of an integrated statistical data collection system of a country.

ÅIf this is the case, then the social and demographic statistics intercensal household surveys can constitute part of this system.

ÅThere are different types of household surveys that can be conducted to collect data on social, economic and demographic statistics such as specialized surveys, multi-phase surveys, multi-subject surveys and longitudinal surveys.

Types of Survey

- The selection of the appropriate type of survey is dependent on a number of factors including **subject matter requirements**, **resources and logistical considerations**.

- Surveys can be classified in a number of ways
  - By their **design** and **method**, **approach**
  - By **size** (how many, national, province, district etc.) and **type** (who-male or female, what-fertility, adult mortality, infant mortality etc.) of sample

Types of Survey- design and methods

1) Descriptive surveys - describe the phenomenon of interest, test hypotheses and generate hypotheses

ÅBut cannot provide robust evidence about the direction of cause and effect relationship

ÅCross-sectional studies are an example of descriptive surveys

ÅMost cross-sectional studies are retrospective - i.e., they involve questioning respondents about past events

Types of Survey

Retrospective

Å**Advantages**

ÅIt is fast to administer and generate results, since events already occurred, we just have to ask and tabulate all the results

ÅRelatively inexpensive process to undertake

Å**Limitations**

ÅSome of the information that is needed may be hard to get due to recall lapse or memory bias

Types of Survey

- **2) Analytic surveys** - suggest *the* ***direction*** *and* ***cause*** *and* ***effect association***
- Data is collected and analyzed at more than one point in time
- Example: longitudinal surveys - ***analyze events at more than one point in time***
- In longitudinal surveys, **data is collected from the same sample units over a period of time** (panel) or **different samples at each data collection period** (trend) or **a follow-up study of people with the common characteristics** (cohort)

Types of survey

Longitudinal studies are **prospective**

- Data is collected prospectively - **over the passage of time**
- The interval can be **monthly**, **quarterly** or **annually**.
- The **purpose** for conducting such surveys is to **measure changes** in some characteristics **over a period of time**
- Also involve questions about the past events

Types of survey

- One of the **advantages** of prospective studies is that the collection of information is at regular time intervals, so recall error is minimized

Problems

- Conditioning effect
- High attrition rate of respondents
- Takes time to design and implement
- Very costly to undertake

Types of survey - scope and content

Specialized surveys

- Specialized surveys cover single subjects in one round of a survey

- Since the 1970s, countries have been conducting specialized surveys concerned with very specific topics, such as fertility, infant and child mortality, reproductive health and morbidity

- E.g. individual in-depth single-round retrospective surveys that use extended individual questionnaires, including a birth history and/or pregnancy history (maternity history) such as the World Fertility Survey (WFS) programme

Multi-subject surveys

ÅIn this survey, **different subjects** are covered in a **single survey**.

ÅThis approach is generally **more cost-effective** than conducting **a series of single subject surveys**.

ÅGive examples of subjects that could be investigated using this type of survey.

Multi-phase surveys

- These surveys entail collecting statistical information in **succeeding phases** with one phase serving as a forerunner to the next.

- The **initial phase** usually constitutes a **larger sample than subsequent phases**.

- It is used to screen sample units based on certain characteristics to ascertain the eligibility of such units to be used in the subsequent phases.

- Give two examples of topics that could be investigated using this type of survey. Explain how each phase will be conducted.

Exercise

1) What is a sample survey?
   a) Explain your understanding of this definition.
   b) Why is probability sampling used in such surveys?
2) What question(s) does each of the following surveys aim to answer:
   a) Descriptive
   b) Analytic
3) Explain why each of the following survey would be selected over the other types of surveys
   a) Specialised survey
   b) Multi-subject survey
   c) Multi-phase survey
4) Compare and contrast cross-sectional and longitudinal surveys.
5) What type of survey is the DHS?

ÅGive reasons for your answer

PLANNING OF SAMPLE SURVEY

- **Introduction**
- In planning of surveys, there is need to pay particular attention to the preparations that precede the field work
- Amount of planning will vary depending on the type of survey, materials and information required
- The development of an adequate survey plan requires sufficient time and resources
  - A planning cycle of two years is not uncommon for a complex survey

1) PLANNING OF A SURVEY

i) General planning of the survey programme

- This is usually carried out by a relatively small group of senior staff members representing *subject-matter, technical* and *administrative backgrounds* (budget, personnel, survey design, calendar, etc.).

ii) Objectives of a survey

ÅThe first step in planning is to lay out the objectives of the investigation.

ÅThis is generally the function of the sponsor, who maybe a government agency, a university researcher, etc.

ÅThe objectives of the survey should be as specific, clear-cut, and unambiguous as possible.

ÅClearly stated objectives is the first step in forming the basis of what questions to ask in the survey for which statistical answers are required.

Examples: Objectives of a survey

- The primary objective of the 2018 ZDHS was to provide up-to-date estimates of basic demographic and health indicators.
- Specific objectives included:
  - To provide up-to-date estimates on fertility levels and preferences
  - To estimate anaemia prevalence among women age 15-49 and children age 6-59 months
  - To assess the nutritional status of children under age 5 (via weight and height measurements)
  - To estimate HIV prevalence among men age 15-59 and women age 15-49 and behavioural risk factors related to HIV
  - To assess of situation regarding violence against women

iii) Survey universe

ÅDefine the geographical areas to be covered and the target population.

ÅIn defining the universe, the exact population to be sampled should be identified.

ÅThe first stage would be enumeration areas, e.g. in urban and/or rural areas and the second stage would be the households in selected enumeration areas.

ÅThe target population, however, is somewhat smaller than the population forming the universe.

ÅOnce target population is defined, sample size is then determined.

Exercise

- Teenage childbearing in Zambia
- **Survey objectives**
  - To determine prevalence of teenage childbearing in Zambia
  - To determine the demographic and socio-economic factors related to teenage childbearing in Zambia
    - Geographical area?
    - Target population?

iv) Subject matter/information to be collected

- Need to clearly specify the "*subject matter*" or information to be collected.
- This must be adequate to answer the questions the survey is intended to answer and in line with the objectives.
- Manage a tension between desire to collect more information and need not to overburden the respondent.
- In determining subject matter, these should be organized by core items, e.g. ??
- What topics are in the Demographic and Health Survey (DHS)?
  - https://www.dhsprogram.com/Topics/more-topics.cfm

The DHS Program - More Topics

dhsprogram.com/Topics/more-topics.cfm

The DHS Program
Demographic and Health Surveys

SEARCH | LOGIN | Select Language

USAID FROM THE AMERICAN PEOPLE

COUNTRIES | DATA | PUBLICATIONS | METHODOLOGY | RESEARCH | TOPICS

The DHS Program > Topics > More Topics

FEATURED

- Childhood Mortality
- COVID-19
- Family Planning
- Gender
- Malaria
- Maternal Mortality
- Nutrition
- SDGs
- Wealth Index

MORE TOPICS

- Alcohol and Tobacco
- Anemia
- Child Health
- Education
- Fertility and Fertility Preferences
- HIV
- Household and Respondent Characteristics
- Male Circumcision
- Maternal Health
- Tuberculosis
- Unmet Need
- Youth

More Topics

The DHS Program surveys and publications provide data on a wide range of health and population topics as listed below for use in research, advocacy, and programs.

Each topics page serves as an introduction to the topic survey indicators, DHS publications and relevant DHS resources and tools. It also highlights other resources and tools that are available in the global health and development field.

- Alcohol and Tobacco
- Anemia
- Child Health
- Education
- Fertility and Fertility Preferences
- HIV
- Household and Respondent Characteristics
- Male Circumcision
- Maternal Health
- Tuberculosis
- Unmet Need
- Youth

v) Budget Estimates and Calendar

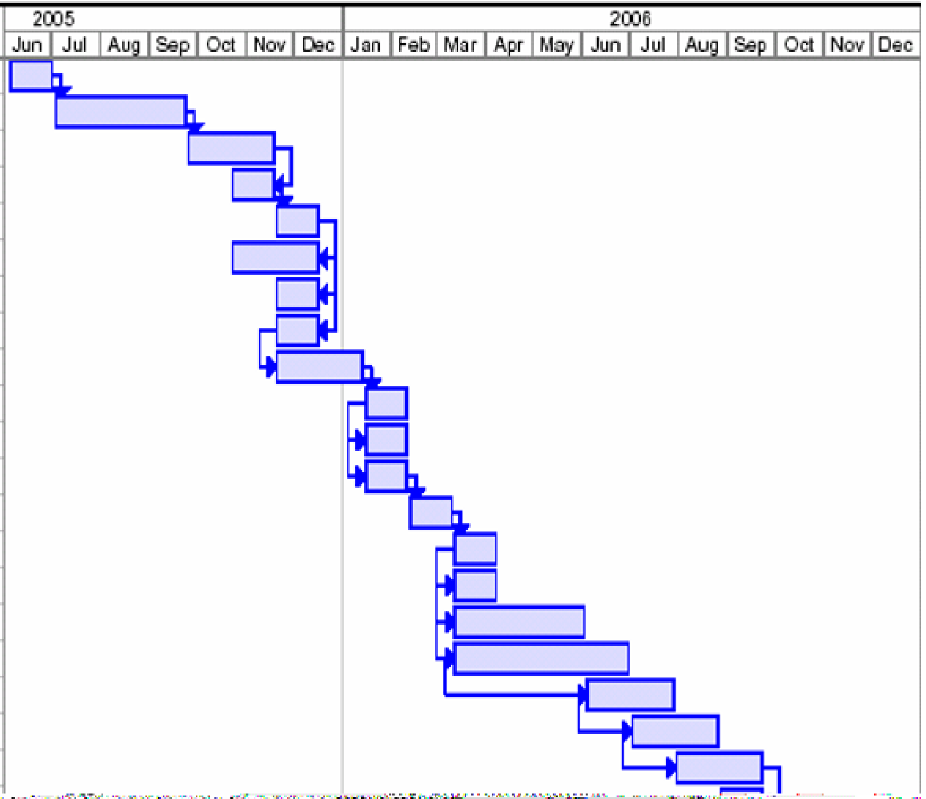
- There is need to establish a general calendar/timetable of operations, which is logically coordinated.

- The budget indicates the financial requirements of the survey which is to be conducted.
- Cost estimates must be as detailed as possible.
- The budget shows cost of personnel, equipment and all other items of expense.
- The budget will depend largely on the survey design, and geographical coverage.
- ***A good survey does not come cheap***

Example

Time-table of Household Survey Activities for Country X

ID	Task Name	Duration
1	Meeting with stakeholders	1 mon
2	Preparatory activities	3 mons
3	Initial questionnaire design	2 mons
4	Send questionnaires to user committee members	1 mon
5	Include in questionnaire agreed suggestions	1 mon
6	Draft interviewer's & supervisor's manuals	2 mons
7	Print questionnaires & manuals	1 mon
8	Make plans for pre-test	1 mon
9	Train interviewers and supervisors	2 mons
10	Conduct the pre-test	1 mon
11	Revise questionnaire (if need be)	1 mon
12	Revise manuals (if need be)	1 mon
13	Sample design	1 mon
14	Design & test data entry programme(s)	1 mon
15	Design & test data cleaning programme(s)	1 mon
16	Data collection	3 mons
17	Data entry	4 mons
18	Design & test tabulation plan	2 mons
19	Data cleaning (eg. range and consistency checks)	2 mons
20	Assignment of chapters to authors	2 mons



iv) Tabulation and Analysis Plan

- A **tabulation plan** should be produced at the time of planning the survey.
- **Design output tables** in accordance with **survey objectives**
- **Dummy tables** are draft tabulations which include everything **<u>except</u>** the **<u>actual data</u>**.
- The blank tables should be circulated for comments and improvement.
- The way in which the data collected in the survey will be used to answer the questions (attain the objectives) can be referred to as the data analysis plan.

Table 5.1 Current fertility

Age-specific and total fertility rates, the general fertility rate, and the crude birth rate for the 3 years preceding the survey, according to residence, Zambia DHS 2018

	Residence		
Age group	Urban	Rural	Total
<15			
15-19			
20-24			
25-29			
30-34			
35-39			
40-44			
45-49			
TFR (15-49)			
GFR			
CBR			

Note: Age-specific fertility rates are per 1,000 women. Rates are for the period 1-36 months preceding the interview. Rates for the 10-14 age group are based on retrospective data from women age 15-17.

TFR: Total fertility rate, expressed per woman

GFR: General fertility rate, expressed per 1,000 women age 15-44

CBR: Crude birth rate expressed per 1,000 population

Table 5.11 Teenage pregnancy and motherhood

Percentage of women age 15-19 who have had a live birth or who are pregnant with their first child, and percentage who have begun childbearing, according to background characteristics, Zambia DHS 2018

	Percentage of women age 15-19 who:		Percentage who have begun childbearing	Number of women
Background characteristic	Have had a live birth	Are pregnant with first child		
Age				
15-17				
15				
16				
17				
18				
19				
Residence				
Urban				
Rural				
Province				
Province 1				
Province 2				
Province 3				
Province 4				
Education				
No education				
Primary				
Secondary				
Higher				
Wealth quintile				
Lowest				
Second				
Middle				
Fourth				
Highest				
Total				

v) Implementation of field work

- For survey operations to be successfully carried out, field work should be properly organized and implemented.

a) ***Equipment and materials*** –Vehicles, bicycles, boats, etc., should be available and in working condition

- Materials, like folders, pencils, pencil sharpeners, notebooks, etc., should be available in adequate supplies for use during the survey operations

b) ***Management of survey operations***–there must be a clear and well defined line of command from the survey manager to the interviewer.

c) Publicity

- Some surveys have had limited success partly due to high non-response owing to refusals.
  - Publicity plays an important role in soliciting cooperation from respondents.
  - Different approaches can be adopted depending on prevailing circumstances (rural/urban).
    - *E.g. local/traditional leaders could play a vital role especially in rural areas*

d) Selection of interviewers

- Interviewers are the interface with the respondents.
- They are the representatives of the survey organization.
- Therefore, the selection of an interviewer should be given great consideration.
- An interviewer should be capable of effectively communicating with the respondent
  - *-should have adequate level of education*
  - *-should be able to record information honestly, without "**cooking figures**"*

- *What other attributes should be considered when selecting interviewers?*

e) Training of interviewers

ÅThe main purpose of the training programme is to brings about uniformity in the interviewing procedures.

ÅThis is necessary to avoid different interpretations of the definitions, concepts, and objectives of the survey by the interviewers and hence avoiding interviewer bias.

ÅInvolves training on how to present the questions, fill out the responses, how to attend to non-responses or invalid responses, etc.

e) *Training of interviewers cont..*

- The purpose of the survey and how the results are going to be used should be explained to the interviewers.

- **Qualified trainers** should be **responsible** for the **training**.

- During the training process, the **interviewers**, in the presence of the trainers, should take turns in explaining to others the various items in the questionnaire and also **do role play**.

f) *Pretests*

- This is **involves testing** of the questionnaire and survey operations to learn if the questionnaire and field methods are right for the survey objectives.
- Whether any of the questions:
  - *are too sensitive?*
  - *unduly invade the respondent's privacy?*
  - *are too difficult even for the willing respondent to answer?*
- Pretest report needs to be reviewed and adjustments are made where necessary.

Data processing

- How is the data going to be analyzed?
- Plan for the software to be used in data entry (if hard copy questionnaires are used)
- Software to be used in data analysis

3) TECHNICAL REPORT WRITING AND DISSEMINATION

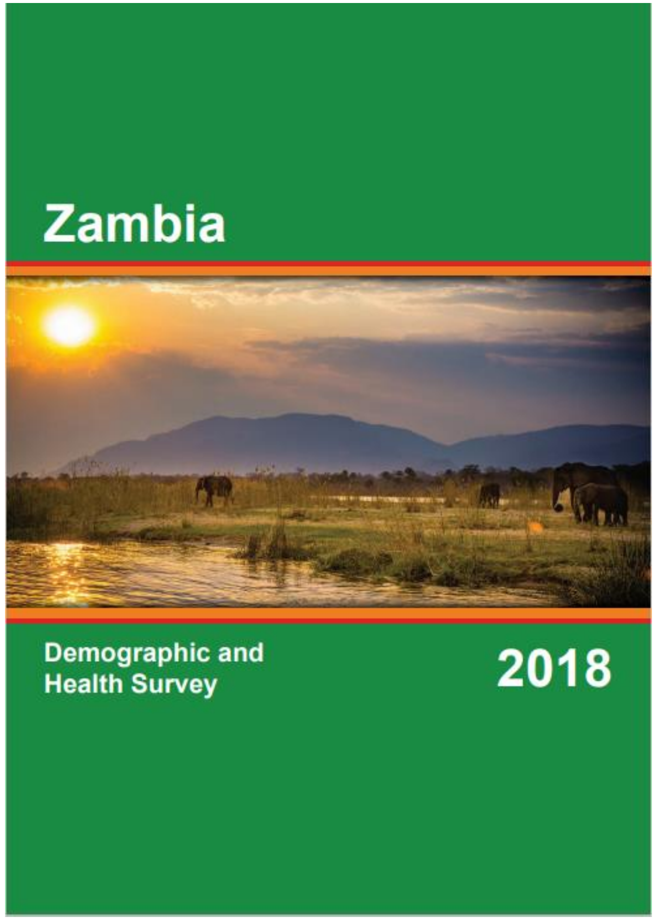
- There is need to plan for report writing and dissemination of results.

- The analytic results are often presented in the form of tables and charts (examples are in the two next slides).

- Authors are assigned sections/chapters and interpretation of the results is then done.

- Report is compiled, validated and finalized for dissemination.

Example



Examples: Results from the Zambia DHS 2018

Table 2.1.1 Household drinking water

Percent distribution of households and de jure population by source of drinking water and by time to obtain drinking water, percentage of households and de jure population with basic drinking water service, and percentage with limited drinking water service, according to residence, Zambia DHS 2018

Characteristic	Households			Population		
	Urban	Rural	Total	Urban	Rural	Total
Source of drinking water						
Improved source	**91.8**	**58.0**	**72.3**	**91.1**	**57.4**	**70.7**
Piped into dwelling/yard/plot	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Piped to neighbour						