International Association for Statistical Computing
African Members Group

THEME:

STATISTICAL COMPUTING USING EMERGENCY NUTRITION ASSESSMENT SOFTWARE: SMART Methodology

TOPIC:

CROSS SECTIONAL STUDIES USING CLUSTER SAMPLING TECHNIQUES

By

Jonathan Atsua Ikughur (Ph.D)

Associate Professor (Sampling Theory/ Methods)

Department of Statistics

Joseph Sarwuan Tarka University Makurdi

Preamble

- The generality of Man's efforts in Science, Technology, Engineering and Management to better the existence of Humankinds.
- Because of the complexity in needs of the humankinds, there is need for emergency interventions especially in this dispensation of wars, terrorism, instabilities that destroy lives and property and also destabilizes people from their abode.
- Important demographic issues of concern under such situations are: Malnutrition & Mortality

Preamble (Continues)

- To achieve effective intervention and ensuring sustainability of relief interventions in order to alter the trends of these challenges, appropriate MODEL must be utilized.
- Scientific inference, whether deductive or inductive targets at reaching decision that minimizes risks of action. This, DATA is required.

Preamble (Continues)

- Several organizations are responsible for DATA collection: In Nigeria, NPC, National/States Bureau of Statistics, International and Local NGOs, CBN, UNSD etc.
  - ❖ collect data to investigate development outcomes for attaining the Sustainable Development Goals and the pluses.
- In Population studies, there are two basic methods of data collection namely, **Census and Sample Survey**
- Our interest in this discourse is **in Sampling or Sample Survey** rather than in Census.

Preamble (Continues)

- In SMART (Standard Monitoring and Assessment for Relief Transition) Methodologies, any of the simple sampling or Complex schemes can be used depending on some considerations that will be presented later in this discourse.
- Here, we present
  - the sampling scheme called Cluster Sampling (Single and Multi-stage)
  - the implementation of Cross Sectional Studies in cluster population;
  - some applications in relations to emergency surveys
- Part B will focus on the use of ENA for SMART Surveys.

? Sampling

- Originally, we seek to make decision about unknown population characteristics. Census brings about high cost in terms of Money, Time, Convenience and associated errors etc.
- ***So we revert to sampling.***
- **Definition 1:** sampling is a scientific method of collecting and using representative sample to seek the truth about the whole.
- **Definition 2:** Sample Survey is the collection of data from a sample in order to make inference about the whole.

- **Remark 1:**
- it involves (i) sample design; (ii) data collection (iii) Estimate (point or interval) of population characteristics of interest (eg. Population mean, variance, ratio, proportion, correlation etc) (iv) determination of accuracy and suitability of the estimates (unbiasness, efficiency, consistence, sufficiency etc)

(v) Conclusions are made to guide policy actions

Population Structure & Sampling Schemes

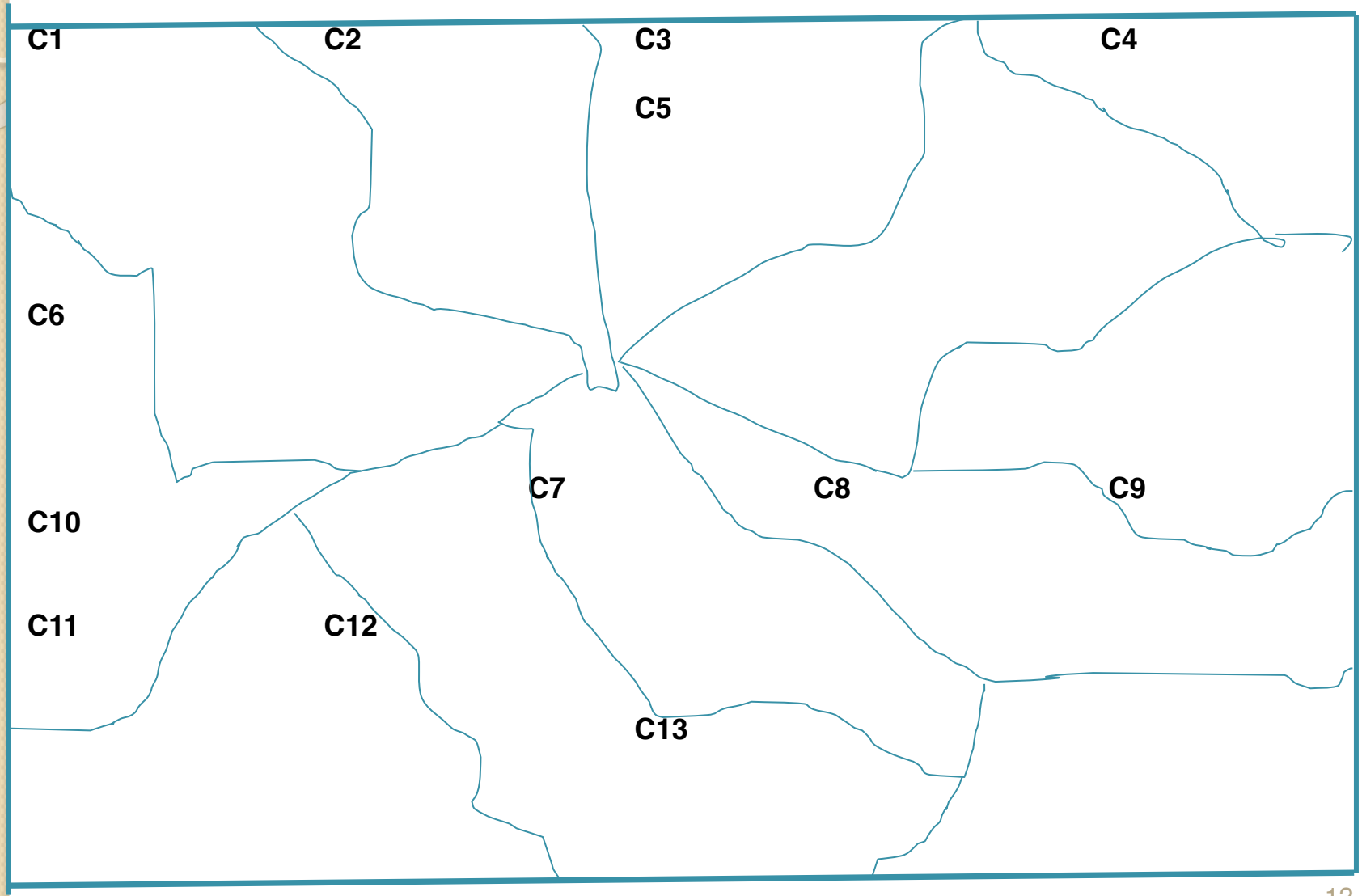
- Sampling scheme refers to the method of selecting a sample from the population.
- it is determined by several factors including:
  - The population structure (systematic, stratified, cluster etc.)
  - Sampling Design: Simple Random Sampling or Probability Proportional to Size (PPS);
  - Availability of sampling frame
  - Cost of sampling with respect to money or time and efforts.

- To venture into sampling, a plan otherwise known as survey protocol has to be made addressing the followings:
  - the survey objectives ?
  - the study population ?
  - availability of sampling frame ?
  - sampling scheme?
  - Specification of expected characteristics of interest
  - Cost of survey
  - Training of enumerators/supervisors etc.

What is a Cluster?

- **Definition 3:** It consists of forming suitable clusters of contiguous population units and surveying all the units in a sample of clusters selected according to an appropriate probability sampling scheme.
- Examples of cluster:
  - Electoral ward;
  - Schools
  - Districts
  - IDP Settlements
  - Refugees' settlements

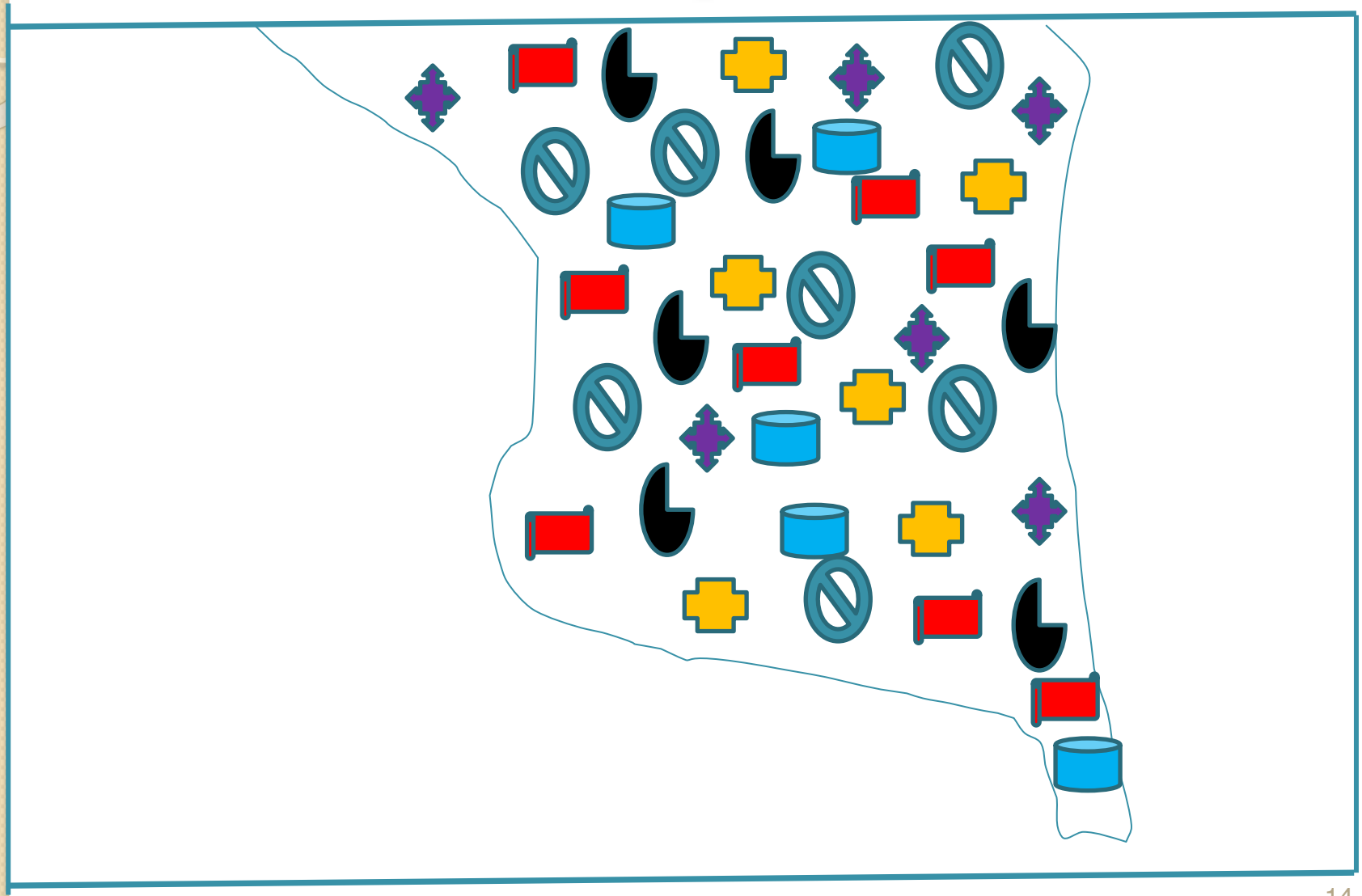
Clusters (N =13)



- Here, the population denoted as Ω is represented as a disjoint union of subsets, each subset being called a cluster.
- Denote the clusters as $C_1, C_2, \ldots, C_N$ (N=15) so that $\Omega = \bigcup_{i=1}^{15} C_i$.
- It is assumed that one knows which units are in C_i.
- Let any cluster, say C_i be a reflection of all clusters consisting of elements as follows:

The Scheme For units in the i^{th} Cluster,

(i=2, Say) , Cluster Size = M_2



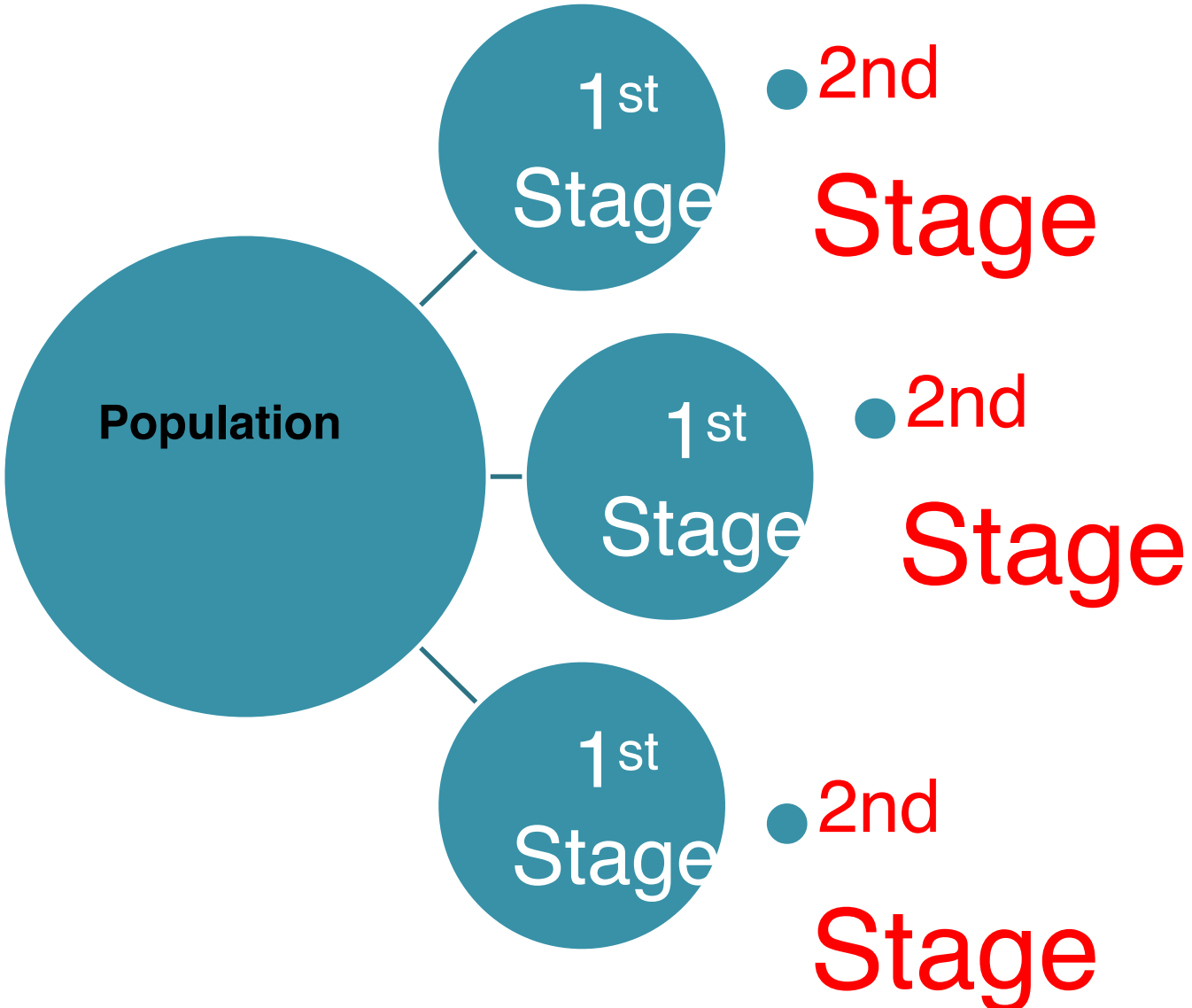
- Here, a random sample of size n < N (ie. ***n-groups < N-Groups*** are selected according to probability sampling scheme.
- If ***all the primary units*** in the selected clusters are investigated upon at minimal cost/time, it is referred to as single stage cluster sampling.

Multistage Sampling?

- In Multistage sampling, we first select a sample of clusters and then select a subsample of the units in each cluster. It is sometimes referred to as "sampling of samples".
- The selected cluster is called a Primary Sampling Unit (PSU) and considered as *first-stage sample.*
- The units selected in the subsample are called Secondary Sampling Unit (SSU) considered as second-stage sample.
- This can be done in *several stages depending on the structure of the population*.

- For instance, In a Survey of Mortality in Nigeria, the hierarchy may be as follows:
  - ❖States constitute Primary Sampling Units
  - ❖Local Governments constitute Two-stage units
  - ❖Ward or district constitute Three-stage units
  - ❖Households constitute the Fourth-Stage units
  - ❖Individual in the fifth-stage constitute Fifth-stage unit (perhaps, the unit of interest upon which observations are made.

A Scheme of Two-Stage Sampling



Advantages and Disadvantages

- Multistage sampling is
- flexible,
- cost effective
- and easy to implement.

- However, it has a subjective component,
- it has problems with external validity.
- It is also less accurate than simple random sampling.

Estimators of Interest

- Population Total
- Population Mean
- Population Proportion
- Population Ratio
- Population Variance etc
- Upon estimation, we may place confidence intervals and investigate the suitability of the estimates realized.

Case Studies involving Multistage sampling

- Johnston *et. al's* **survey on drug use in high schools** used three stage sampling: **geographic areas**, followed by **high schools within those areas**, followed by **senior students** in the schools.
- The **Australian Bureau of Statistics** divides cities into "**collection districts**", then **blocks**, then **households**.

- Ekpo *et al.* Survey to determine the magnitude and severity of malnutrition rates among the under-five children in IDP Camps. Here, the clusters are:
  - the LGA's (Stage 1)
  - IDP settlements (Stage 2)

A sample of children under-5 years in each selected cluster are observed in a Two-Stage Sampling

Design Effect in Cluster Sampling

- It is the ratio of the variance of an estimator under a sample design to that of the estimator under SRS

- It is used in survey sampling for
  (i) planning a sample design and
  (ii) report the effect of the sample design in estimation and analysis.
- It shows the efficiency of a complex sampling designs as the ratio of the variance of a statistic under srswor of size n to the variance of the statistic under the complex design with the same sample size.

- It is given by the relation
- $Deff = 1 + \rho(n-1)$

where

- $\rho = (MSB - MSW)/(MSB + MSW(n-1))$

and

- $$MSB = 1/(k-1)\sum_{i=1_k}^{k}\left(y_i - n_i p\right)^2/n \quad ,$$

$$MSW = 1/(N-k)\sum_{i=1}\left(n_i - y_i\right)^2/n,$$

Depending on the sampling scheme being implemented.

The sample size considering DEFF is given as

$n = n_{srs} * DEFF$

CROSS SECTIONAL STUDY

- Cross sectional is the study of human population by means of a sample that includes representative of all sections of the society.
- The use of the term Cross Sectional Survey is a misnomer in the sampling world
- Scholars see it as epidemiological/public health design used to assess exposure (cause) and a disease (effect) and compare the rates of diseases and symptoms of an exposed with an unexposed group

- Cross Sectional studies

i. can assess how frequently, widely, or severely a specific variable occurs throughout a specific demographic.

The principle is that the study is

- conducted at a single point in time (Snapshot)
- Conducted over a short period of time.
- Aimed at describing a variable but not to measure it
- researchers are able to look at numerous characteristics at once.

- For example, a cross-sectional study could be used to investigate whether exposure to certain factors, such as overeating, might correlate to particular outcomes, such as obesity.

Cross-sectional study can be

- **i. Analytical Cross-sectional Studies:** Here, researchers investigate an association between two parameters.
- data on exposures and outcomes at one specific point in time are collected in order to measure an association between an exposure and a condition within a defined population.
- The aim is to compare health outcome differences between exposed and unexposed individuals.

- **ii. Descriptive Cross-sectional Studies:**
- **T**hese studies are used to characterize and assess the prevalence and distribution of one or many health outcomes in a defined population.
- The indicators of interest could be

i. Prevalence Ratio (PR) when the prevalence of the outcome is more or equal to 10%;

ii. Prevalence Odds Ratio (OR) when that prevalence is lower.

- Different regression models like Binomial log or Poisson log, or still, generalized lineal models.
- If the association measure is OR, then, the most common model used is the logistic regression.

Prevalence

- Prevalence is the proportion of people in a population who have a disease or condition of interest at a particular time and its value range between 0 to 1.0.
- **Prevalence ratio:** is the prevalence of disease in the exposed group divided by the prevalence of disease in the non-exposed group.
- Prevalence ratio is close to the risk ratio when the disease is rare (<5%),
- **Prevalence Odd Ratio**

Steps of Conducting a Cross-Sectional Study

- To successfully conduct a cross-sectional study, The steps from designing to implementing are:

- 1 Definition and analysis of the problem;
- 2 Review the literature
- 3 Formulation of objectives of study

- 4. Determine the research methodology:
  - i. sampling method
  - ii. variables of interest
  - iii. Plan for data collection/analysis.
  - iv. Ethical consideration
- 5. Work plan: Researcher should determine the timetable of the study, who is going to do what and when.

CASE STUDY: SURVEY ON ACUTE MALNUTRITION

- **General Objective**
- The overall objective of the survey was to determine the magnitude and severity of malnutrition rates currently among the under-five children in XXX, Metropolitan Council in YYY State, Nigeria.

- **Specific Objectives**

- To estimate the prevalence of acute malnutrition among children 6 to 59 months of age using WHZ, MUAC and bilateral oedema.
- To estimate the prevalence of chronic malnutrition using HAZ, underweight and overweight using WAZ among children 6 to 59 months of age.

Methodology

- **Study Design**
- A **cross sectional survey procedure, using <u>two-stage cluster sampling</u> technique based on <u>Probability Proportional to Size (PPS)</u>**, which is useful in a Nutrition and Mortality surveys as it allows for more representation from villages/wards with larger population sizes.

- The sample sizes would be sufficiently minimum to ensure minimal acceptable precision as per the SMART guidelines.

- The survey follows the usual methods for measuring MUAC, weight, height and age using well trained measurers according to the SMART guideline.

First Stage Sampling Procedure: Selection of Clusters

- The study population will comprise all individuals identified in the X settlements (Sub-wards) in XXX, with their housing pattern traditionally arranged in clusters .

Second Stage Sampling Procedure: Household Selection

- Since the arrangement of houses/ settlements in most parts of XXX are **systematic** in nature, the second stage of sampling which consist of selecting households within each cluster would be based on the systematic random selection procedure.

Parameters for Anthropometry Data Collection Plan

- To be discussed by the ENA/SMART Expert

Thanks for Listening