

CHAPTER 10

SAMPLING FOR MEAL

Designing Representative and Purposeful Samples for Data Collection

Presented by

Dr Abenet Yohannes, Ph.D

Jigjiga University

Learning Objectives

By the end of this chapter, participants will be able to:

- Define population, sample and sampling frame
- Distinguish a census from a sample
- Explain probability and non-probability sampling
- Apply simple random, systematic, stratified, cluster and multistage sampling
- Apply purposive, convenience, snowball and quota sampling
- Identify factors affecting sample size
- Recognize representation and selection bias
- Design qualitative sampling strategies
- Manage non-response and replacement
- Develop sampling strategies for quantitative and qualitative assessments

What is Sampling?

Sampling is the process of selecting a smaller group of units from a larger population for data collection.

SAMPLING UNITS MAY INCLUDE

- Individuals
- Households
- Children
- Schools
- Health facilities
- Communities
- Organizations
- Project sites

Why Use Sampling?

Sampling can:

- Reduce cost
- Save time
- Reduce respondent burden
- Improve field supervision
- Enable deeper data collection
- Produce timely findings
- Make large studies manageable

Sampling and Data Quality

A credible sample should be:

- Relevant to the research question
- Selected using a clear method
- Large enough for the required analysis
- Inclusive of important groups
- Protected from avoidable bias
- Documented and reproducible
- Feasible within available resources

What is a Population?

A population is the complete group about which the project wants information or conclusions.

EXAMPLES

- All project-supported households
- All children enrolled in target schools
- All health workers in supported facilities
- All communities in the intervention area
- All feedback cases received during the year

Target Population

The target population is the complete group to which the study findings are intended to apply.

EXAMPLE

All displaced households living in the five target districts.

THE DEFINITION MAY SPECIFY

- Who
- Where
- Eligibility criteria
- Time period
- Project exposure

Study Population

The study population is the group that is practically accessible for sampling.

EXAMPLE

Target households appearing on verified registration lists and currently residing in accessible project locations.

What is a Sample?

A sample is the subset of the population selected to participate in the study.

EXAMPLE

From 12,000 project-supported households, 600 households are selected for a survey.

What is a Sampling Unit?

A sampling unit is the unit selected at a particular stage of sampling.

POSSIBLE UNITS

- District
- Community
- Household
- Individual
- School
- Classroom
- Health facility

Unit of Analysis

The unit of analysis is the unit about which conclusions are made.

EXAMPLE

- Household survey: household
- Student assessment: student
- Facility assessment: health facility
- School observation: school or classroom

What is a Sampling Frame?

A sampling frame is a specific list or structure used to select the sample.

EXAMPLES

- Household registration list
- School enrolment register
- List of project participants
- Facility database
- Community census
- Employee list

Characteristics of a Good Sampling Frame

A good frame should:

- Cover the full study population
- Include eligible units
- Exclude ineligible units
- Avoid duplicate entries
- Include accurate contact or location information
- Be updated before selection
- Allow disaggregation where needed

Sampling Frame Problems

Common problems include:

- Missing households
- Outdated lists
- Duplicate participants
- Incorrect locations
- Exclusion of mobile populations
- Inclusion of people who have left
- Missing disability or vulnerability information
- Lists controlled by powerful local actors

What is a Census?

A census collects information from every eligible unit in the defined population.

EXAMPLES

- All project employees
- All 25 supported schools
- All safeguarding cases
- All 40 health facilities
- Every household in a very small settlement

Advantages of a Census

A census can provide:

- Complete coverage
- Detailed subgroup information
- No sampling error
- Information for every unit
- Strong operational records

Limitations of a Census

A census may be:

- Expensive
- Time-consuming
- Difficult to supervise
- Burdensome to participants
- Vulnerable to inconsistent collection
- Unnecessary for certain decisions
- Difficult in large or mobile populations

Census or Sample Decision

USE A CENSUS WHEN

- The population is small
- Every case is operationally important
- Data already exist in a complete system
- Missing any unit creates serious risk
- Subgroup-level results are required for every unit

USE A SAMPLE WHEN

- The population is large
- Time and resources are limited
- Statistical estimates are sufficient
- Data collection is intensive
- Rapid evidence is required

What is Probability Sampling?

In probability sampling, every eligible unit has a known and non-zero chance of selection.

PROBABILITY SAMPLING SUPPORTS

- Statistical estimation
- Confidence intervals
- Generalization to the population
- Calculation of sampling error
- Reduced selection bias

Main Probability Sampling Methods

Common methods include:

1. Simple random sampling
2. Systematic sampling
3. Stratified sampling
4. Cluster sampling
5. Multistage sampling

Simple Random Sampling

Simple random sampling gives every unit an equal chance of selection.

EXAMPLE

Selecting 300 households from a verified list of 5,000 households using random numbers.

Steps in Simple Random Sampling

STEPS

1. Define the population
2. Prepare and clean the sampling frame
3. Assign a unique number to every unit
4. Determine the sample size
5. Generate random numbers
6. Select the corresponding units
7. Document the selection process

Advantages of Simple Random Sampling

Advantages include:

- Easy to understand
- Equal selection probability
- Supports statistical analysis
- Reduces intentional selection
- Works well with complete lists

Limitations of Simple Random Sampling

Limitations include:

- Requires a complete list
- Selected units may be widely dispersed
- Small subgroups may be underrepresented
- Travel costs may be high
- Difficult with mobile populations

Systematic Sampling

Systematic sampling selects every n th unit from an ordered list after a random starting point.

EXAMPLE

Population = 2,000 households

Required sample = 200 households

Sampling interval: $2,000 \div 200 = 10$

Select every 10th household.

Steps in Systematic Sampling

STEPS

1. Clean and order the list
2. Calculate the sampling interval
3. Select a random starting point
4. Select every n th unit
5. Continue until the sample is complete

Advantages of Systematic Sampling

Advantages:

- Simple to administer
- Faster than repeated random selection
- Spreads the sample across the list
- Useful for household or participant registers
- Easy to explain to field teams

Risks of Systematic Sampling

Potential risks include:

- Hidden patterns in the ordered list
- Incorrect interval
- Non-random starting point
- Staff changing the route
- Replacement of unavailable households
- Exclusion of households outside the route

Random Route Sampling

Where no complete household list exists, a random-route approach may be used.

TYPICAL STEPS

- Map the community
- Estimate the number of households
- Select a random starting point
- Establish a route
- Visit every n th household

Stratified Sampling

Stratified sampling divides the population into important subgroups, called strata, before sampling.

POSSIBLE STRATA

- Sex
- District
- Urban and rural
- Refugee and host community
- School type
- Vulnerability category
- Project intervention type

Why Use Stratification?

Stratification helps:

- Ensure key groups are represented
- Compare subgroups
- Improve statistical precision
- Prevent small groups from disappearing
- Support disaggregated reporting

Proportionate Stratified Sampling

Each stratum receives a sample proportional to its population size.

Stratum	Population Share	Sample
District A	50%	250
District B	30%	150
District C	20%	100

Disproportionate Stratified Sampling

Some strata receive a larger sample than their population share.

REASONS MAY INCLUDE

- Small but important groups
- Required subgroup comparison
- Vulnerable populations
- Higher expected non-response
- Donor reporting requirements

Stratification Risks

Possible problems include:

- Too many strata
- Very small subgroup samples
- Incorrect population proportions
- Missing stratum information
- Complex weighting
- Higher cost
- Inadequate analysis planning

Cluster Sampling

In cluster sampling, naturally occurring groups are selected first.

POSSIBLE CLUSTERS

- Villages
- Schools
- Camps
- Health facilities
- Districts
- Neighbourhoods

Data are then collected from all or some units within selected clusters.

Example of Cluster Sampling

Population: 20,000 households in 100 villages.

PROCESS

1. Randomly select 20 villages
2. Select 25 households from each village

Total sample = 500 households.

Advantages of Cluster Sampling

Advantages:

- Reduces travel
- Useful when no individual-level frame exists
- Suitable for dispersed populations
- Operationally efficient
- Allows sampling from community lists created after cluster selection

Limitations of Cluster Sampling

Limitations:

- People within a cluster may be similar
- Lower statistical precision
- Larger sample may be required
- Analysis must account for clustering
- Poor cluster selection can create bias

Multistage Sampling

Multistage sampling selects units through two or more stages.

EXAMPLE

Stage 1: Select districts

Stage 2: Select communities

Stage 3: Select households

Stage 4: Select one adult respondent

Multistage Sampling Example

Education survey:

1. Randomly select schools
2. Randomly select classes within schools
3. Randomly select students within classes

Advantages of Multistage Sampling

Advantages:

- Practical for large geographic areas
- Reduces listing requirements
- Controls travel costs
- Allows staged field organization
- Can combine stratification and clustering

Risks in Multistage Sampling

Risks include:

- Complex probabilities
- Higher sampling error
- Loss of representativeness at each stage
- Incorrect household selection
- Failure to account for cluster effects
- Informal respondent substitution

What is Non-Probability Sampling?

In non-probability sampling, selection probabilities are unknown.

UNITS ARE SELECTED BASED ON

- Researcher judgment
- Accessibility
- Referral
- Specific characteristics
- Information needs

Main Non-Probability Methods

Common methods include:

1. Purposive sampling
2. Convenience sampling
3. Snowball sampling
4. Quota sampling

Purposive Sampling

Purposive sampling deliberately selects participants because they have specific experiences, characteristics or knowledge.

EXAMPLES

- Women using remote water points
- School principals
- Participants who dropped out
- Highly successful project sites
- Communities with poor results
- Technical government officials

Purpose of Purposive Sampling

Purposive sampling helps teams:

- Understand experiences
- Explore causes
- Explain quantitative patterns
- Identify barriers
- Compare contrasting cases
- Develop theories or explanations
- Capture marginalized perspectives

Types of Purposive Sampling

Examples include:

- Typical-case sampling
- Critical-case sampling
- Best- and worst-case sampling
- Maximum-variation sampling
- Extreme-case sampling
- Expert sampling
- Criterion sampling

Best- and Worst-Case Sampling

SELECTS CASES WITH

- Strongest results
- Weakest results

PURPOSE

- Compare implementation
- Identify success factors
- Understand barriers
- Generate lessons

Typical-Case Sampling

Selects cases that appear reasonably average or typical.

PURPOSE

- Understand common experiences
- Describe routine implementation
- Provide an illustrative case

Critical-Case Sampling

Selects cases particularly important to understanding the issue.

EXAMPLE

A remote health facility serving a highly vulnerable population.

Convenience Sampling

Convenience sampling selects participants who are easiest to reach.

EXAMPLES

- People attending a community meeting
- Visitors at a service centre
- Staff available in the office
- Respondents near a main road

Limitations of Convenience Sampling

Convenience samples may exclude:

- Remote participants
- People dissatisfied with the project
- People who stopped participating
- Working people
- Persons with mobility limitations
- Less visible groups
- People unable to attend meetings

Appropriate Uses of Convenience Sampling

It may be used for:

- Initial exploration
- Tool pretesting
- Rapid operational feedback
- Small pilot studies
- Training exercises

Snowball Sampling

Snowball sampling begins with known participants, who then refer the team to others.

IT IS USEFUL FOR

- Hidden populations
- People without formal lists
- Sensitive networks
- Hard-to-reach groups
- Specialized informants

Snowball Sampling Process

PROCESS

1. Identify initial participants
2. Conduct interviews
3. Ask for appropriate referrals
4. Screen new participants
5. Continue until sufficient diversity or saturation is reached

Snowball Sampling Risks

Risks include:

- Overrepresentation of connected networks
- Exclusion of isolated people
- Gatekeeper influence
- Confidentiality concerns
- Repeated similar perspectives
- Difficulty knowing the population size

Quota Sampling

Quota sampling establishes the required number of participants from selected categories.

EXAMPLE

50 women

50 men

30 youth

20 persons with disabilities

Participants within the quota may be selected non-randomly.

Advantages of Quota Sampling

Advantages:

- Ensures selected groups are included
- Faster than probability sampling
- Useful without complete lists
- Supports qualitative comparison
- Can address underrepresentation

Quota Sampling Limitations

Limitations:

- Selection bias within quotas
- Unknown sampling error
- Overreliance on accessible participants
- Quotas may use inaccurate population estimates
- Findings cannot be generalized statistically

Probability versus Non-Probability Sampling

Probability Sampling	Non-Probability Sampling
Known selection chance	Unknown selection chance
Supports population estimates	Supports depth and exploration
Usually quantitative	Often qualitative
Requires stronger sampling frame	Can work without complete lists
Sampling error can be estimated	Sampling error cannot be calculated
More generalizable	Context-specific

Sampling Method Selection Questions

Ask:

- Do we need population-level estimates?
- Is a complete sampling frame available?
- Are subgroup comparisons required?
- Is the population geographically dispersed?
- Is the issue qualitative or quantitative?
- Are participants hidden or difficult to reach?
- What resources and time are available?
- What level of precision is needed?

What Determines Sample Size?

Quantitative sample size depends on:

- Population size
- Expected proportion or variability
- Confidence level
- Margin of error
- Sampling design
- Subgroup comparisons
- Expected non-response
- Available resources

Confidence Level

The confidence level reflects how confident the researcher wants to be that the estimate captures the true population value within the stated margin of error.

COMMON LEVELS

- 90%
- 95%
- 99%

Margin of Error

The margin of error expresses the expected range of uncertainty around a sample estimate.

EXAMPLE

**60% of households use safe water, with a $\pm 5\%$ margin of error.
The likely population value is approximately 55% to 65%.**

Confidence and Sample Size

As precision requirements increase:

- Confidence level increases
- Margin of error decreases
- Sample size increases
- Cost increases

Population Size

Population size affects sample size most strongly when the total population is relatively small.

FOR VERY LARGE POPULATIONS

- Sample size does not increase proportionally with population size
- Precision requirements become more important than total population alone

Design Effect

Clustered samples often require a larger sample because respondents within the same cluster may be similar.

The design effect adjusts for reduced precision caused by clustering.

Non-Response Adjustment

If the required completed sample is 400 and expected non-response is 10%:

ADJUSTED SAMPLE

$$400 \div 0.90 = \text{approximately } 445$$

Subgroup Analysis

Larger samples are needed when comparing:

- Women and men
- Age groups
- Districts
- People with and without disabilities
- Intervention groups
- Vulnerability categories

Practical Sample-Size Constraints

Consider:

- Budget
- Field days
- Number of enumerators
- Travel time
- Security
- Access
- Interview length
- Analysis capacity
- Participant burden

What is Representation?

Representation means that the sample appropriately reflects the population characteristics relevant to the study.

EXAMPLES

- Geographic distribution
- Sex
- Age
- Disability
- Project participation
- Vulnerability status

What is Selection Bias?

Selection bias occurs when some types of participants are more likely to be selected than others.

EXAMPLES

- Interviewing only available households
- Surveying only service users
- Excluding remote areas
- Allowing leaders to select respondents
- Replacing unavailable households with nearby households

Common Sources of Selection Bias

Selection bias commonly arises from:

- Incomplete sampling frames
- Convenience selection
- Voluntary response
- High non-response
- Gatekeeper control
- Inaccessible locations
- Unsafe areas excluded
- Poor replacement practices
- Surveys conducted only during working hours

Reducing Selection Bias

Strategies include:

- Clean the sampling frame
- Use random selection
- Include remote locations
- Schedule repeat visits
- Monitor non-response
- Use appropriate stratification
- Train staff not to substitute respondents
- Document exclusions
- Weight data where appropriate

Volunteer Bias

Volunteer bias occurs when people who choose to participate differ from those who do not.

POSSIBLE DIFFERENCES

- Stronger opinions
- Greater satisfaction
- More time
- Higher education
- Greater project engagement

Gatekeeper Bias

Gatekeepers may influence selection by directing teams toward:

- Supportive participants
- Easily accessible households
- Community leaders
- Successful beneficiaries
- Individuals with particular views

Purpose of Qualitative Sampling

Qualitative sampling seeks:

- Depth
- Diversity
- Context
- Explanation
- Information-rich cases
- Contrasting experiences
- Emerging themes

Qualitative Sample Criteria

Participants may be selected based on:

- Experience with the project
- Role or expertise
- Outcome achieved
- Vulnerability
- Location
- Participation status
- Success or failure
- Contrasting perspectives

Data Saturation

Saturation occurs when additional interviews or discussions produce little significant new information.

SIGNS INCLUDE

- Themes repeat
- Explanations become consistent
- No major new categories emerge
- Key subgroups are adequately represented

Sample Size for Interviews

Interview sample size depends on:

- Diversity of participants
- Complexity of the question
- Number of subgroups
- Depth of interviews
- Saturation
- Time and resources
- Triangulation needs

Sample Size for Focus Groups

Plan enough FGDs to:

- Cover required subgroups
- Compare perspectives
- Confirm recurring themes
- Identify important differences

The MEAL DPro guide notes that two to three FGDs may identify many recurring themes within a defined perspective, while broader thematic coverage may require more groups.

Qualitative Subgroup Planning

Example — Research question: How do women experience access to water points?

POSSIBLE GROUPS

- Women near water points
- Women in remote locations
- Women with disabilities
- Adolescent girls
- Host-community women
- Displaced women

Triangulation in Qualitative Sampling

Qualitative findings may be strengthened by comparing:

- Different respondent groups
- Different locations
- Positive and negative cases
- Interviews and FGDs
- Observations and records
- Baseline and endline perspectives

What is Non-Response?

Non-response occurs when a selected unit:

- Cannot be located
- Is unavailable
- Refuses participation
- Is ineligible
- Cannot complete the interview
- Has an incomplete response

Why Non-Response Matters

High or uneven non-response can:

- Reduce sample size
- Create selection bias
- Reduce subgroup representation
- Affect precision
- Distort findings
- Weaken generalizability

Managing Non-Response

Good practices include:

- Make repeat visits
- Vary visit times
- Provide clear study information
- Use trained interviewers
- Protect privacy
- Reduce questionnaire burden
- Track reasons for refusal
- Monitor patterns by location and subgroup

Response Rate

A simplified response rate may be calculated as:

RESPONSE RATE

Completed interviews ÷ Eligible selected units contacted × 100

What is Replacement?

Replacement means selecting another unit when an originally selected unit cannot participate.

REPLACEMENT CAN INTRODUCE BIAS WHEN FIELD STAFF SELECT

- The nearest household
- A cooperative participant
- A friend or relative
- Someone recommended by a leader

When Replacement May Be Used

Replacement should only occur when:

- It is included in the sampling protocol
- Eligibility rules are clear
- The replacement is preselected or randomly selected
- Required contact attempts have been completed
- The reason is documented
- Supervisory approval is obtained

Reserve Sample

A reserve sample is a list of additional randomly selected units prepared before fieldwork.

GOOD PRACTICE

- Keep reserve units in random order
- Use only after defined attempts
- Record the original case outcome
- Obtain supervisor authorization
- Do not allow enumerator choice

Household Respondent Selection

USE A DEFINED METHOD

- Random selection
- Kish grid
- Birthday method
- Predefined eligibility rule

AVOID

- Interviewing whoever answers the door
- Selecting the household head automatically
- Allowing family members to choose

Sampling Plan Components

A complete sampling plan should include:

- Study objective
- Target population
- Study population
- Sampling unit
- Unit of analysis
- Sampling frame
- Sampling method
- Sample-size rationale
- Stratification or clustering
- Non-response adjustment
- Replacement rules
- Weighting requirements
- Limitations

Field Sampling Instructions

Field instructions should explain:

- Selected locations
- Selected units
- Contact procedures
- Eligibility screening
- Respondent selection
- Repeat-visit rules
- Replacement rules
- Refusal documentation
- Daily reporting
- Supervisor authorization

Sampling Tracking Sheet

A tracking sheet may include:

Sample ID	Location	Selected Unit	Contact Status	Eligibility	Outcome	Replacement

Reporting Sampling Limitations

Reports should disclose:

- Incomplete frames
- Excluded locations
- High non-response
- Security restrictions
- Replacement levels
- Underrepresented groups
- Cluster effects
- Non-probability selection
- Limits on generalization

Practical Exercise Overview

Participants will develop:

1. A sampling strategy for a household survey
2. A sampling strategy for a qualitative assessment

Use an existing project or the provided case study.

Household Survey Scenario

A LIVELIHOODS PROJECT SUPPORTS

- 8,000 households
- 40 communities
- Four districts
- Host and displaced populations
- Female- and male-headed households

THE PROJECT WANTS TO ESTIMATE

- Percentage of households with increased income
- Differences by district and displacement status

Household Survey Planning Questions

Groups should define:

- Target population
- Sampling frame
- Sampling unit
- Sampling method
- Stratification
- Cluster selection
- Sample-size considerations
- Non-response allowance
- Household selection
- Replacement protocol

Possible Household Survey Strategy

Example:

- Stratify by district and displacement status
- Select communities using probability methods
- List eligible households in selected communities
- Select households systematically
- Randomly select one eligible adult respondent
- Add allowance for expected non-response

Qualitative Assessment Scenario

The project wants to understand:

- Why some households achieved strong income growth
- Why others did not improve
- Barriers faced by women and persons with disabilities
- Community perceptions of project support

Qualitative Sampling Plan

Groups should identify:

- Key respondent categories
- Purposive sampling method
- Number of KIIs
- Number and composition of FGDs
- Locations
- Inclusion criteria
- Saturation approach
- Triangulation plan
- Ethical risks

Sampling Strategy Template

Element	Household Survey	Qualitative Assessment
Purpose		
Population		
Sampling method		
Sample size		
Subgroups		
Selection process		
Non-response plan		
Limitations		

Peer Review Questions

Review another group's sampling plan:

- Is the population clearly defined?
- Is the method appropriate?
- Is the sampling frame adequate?
- Are key groups represented?
- Is the sample size justified?
- Are non-response and replacement addressed?
- Can the findings be generalized?
- Are qualitative selection criteria clear?
- Are limitations acknowledged?
- Is the plan feasible?

Sampling Quality Checklist

Confirm that:

- Population is clearly defined
- Sampling unit is identified
- Sampling frame is assessed
- Method matches the study purpose
- Sample size is justified
- Important subgroups are included
- Selection bias is addressed
- Non-response procedures are clear
- Replacement is controlled
- Limitations are documented

Common Quantitative Sampling Mistakes

Avoid:

- Using convenience sampling for population estimates
- Incomplete sampling frames
- Ignoring small subgroups
- Selecting easy-to-reach locations
- Incorrect systematic intervals
- Informal replacement
- No non-response adjustment
- Ignoring clustering in analysis
- Reporting percentages from very small samples

Common Qualitative Sampling Mistakes

Avoid:

- Interviewing only leaders
- Including only successful participants
- Conducting one FGD and claiming saturation
- Mixing participants with major power differences
- Ignoring marginalized groups
- Selecting respondents without clear criteria
- Generalizing qualitative findings statistically
- Stopping only because the budget is exhausted

Common Non-Response Mistakes

Avoid:

- Replacing households immediately
- Failing to make repeat visits
- Recording refusals as absent
- Allowing enumerators to select replacements
- Ignoring subgroup response differences
- Hiding high non-response
- Pressuring people to participate

Chapter Summary

In this chapter, we learned that:

- The population is the complete group of interest.
- A sample is the subset selected for study.
- A sampling frame supports sample selection.
- A census includes every eligible unit.
- Probability sampling supports statistical generalization.
- Simple random and systematic methods require clear selection processes.
- Stratification protects important subgroup representation.
- Cluster and multistage sampling improve field feasibility.
- Purposive sampling provides depth and information-rich cases.
- Convenience sampling has high bias risk.
- Snowball sampling helps access hidden populations.
- Quota sampling ensures selected categories are included.
- Sample size depends on precision, design, subgroup needs and resources.
- Non-response and replacement must be carefully managed.
- Qualitative sampling focuses on diversity, depth and saturation.

Knowledge Check

1. What is the difference between a population and a sample?
2. What is a sampling frame?
3. When is a census appropriate?
4. What distinguishes probability sampling?
5. How does systematic sampling work?
6. Why is stratified sampling used?
7. What is the main limitation of cluster sampling?
8. When should purposive sampling be used?
9. Why is convenience sampling risky?
10. What is data saturation?
11. How can non-response create bias?
12. Why should replacements be preselected?

Action Planning

Each participant should identify:

- One project with a weak sampling frame
- One survey using an inappropriate sampling method
- One underrepresented group
- One non-response problem
- One replacement risk
- One qualitative sampling weakness
- One corrective action
- Responsible person
- Completion date

THANK YOU

Questions and Discussion

Dr Abenet Yohannes, Ph.D

Email: abenetyohannes@gmail.com

Jigjiga University

A sample is valuable only when the selection process matches the information need, protects representation and supports honest conclusions about what the data can—and cannot—show.