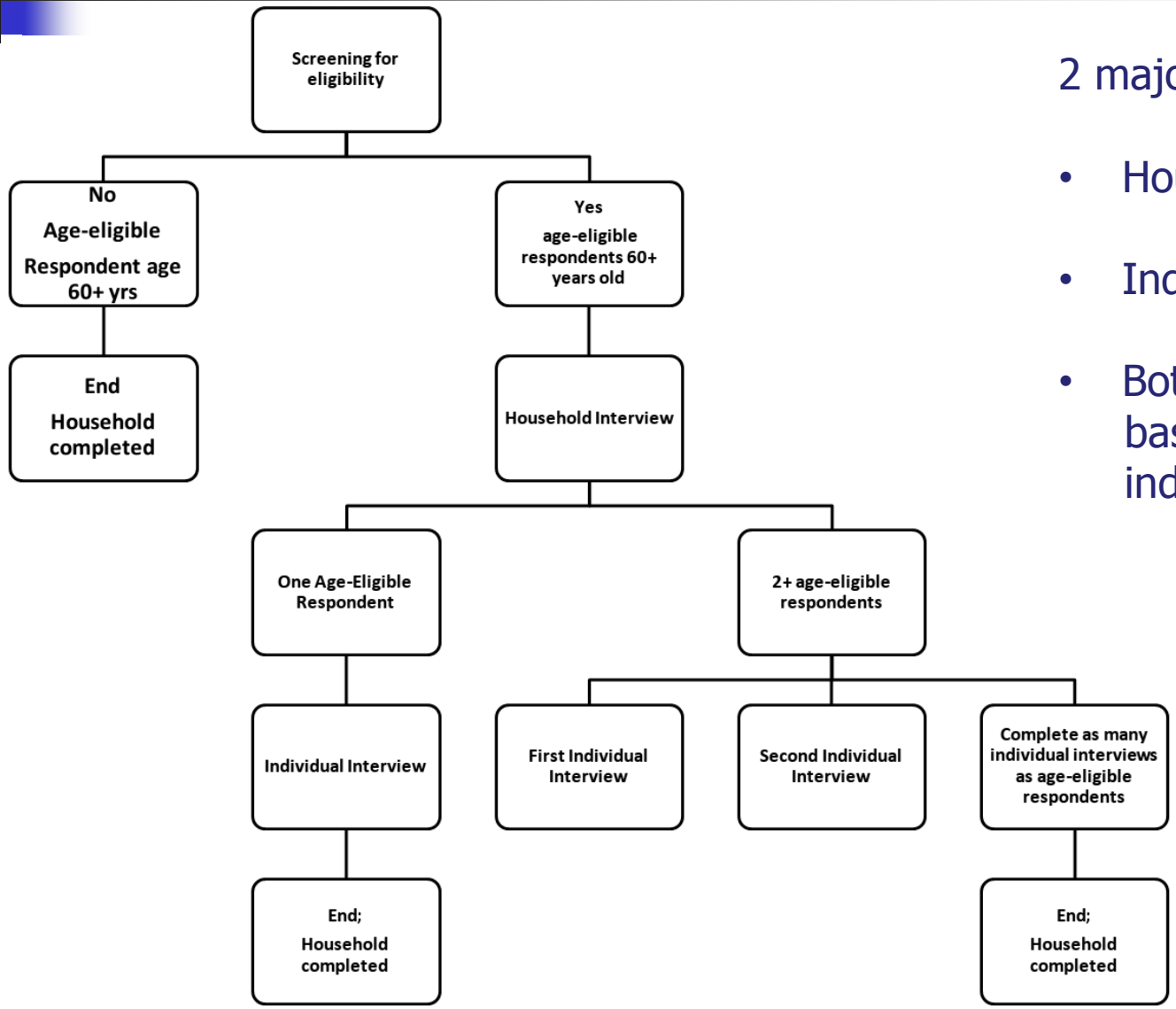
Sampling Design of The Survey on Aging in sub-Saharan Africa

Iliana V. Kohler

Population Studies Center
University of Pennsylvania

Overall Structure of the Survey on Aging in SSA



2 major components:

- Household Interview
- Individual Interview
- Both will be linked based on HH ID and individual ID

Overall Sampling Design

- Nationally representative stratified random sample of households that include at least 1 household member age 60 years and older

- Household sample surveys:
  - Key source for data on social phenomena
  - Are among the most flexible methods of data collections
  - In theory almost any population-based subject can be investigated through household surveys

- Only probability samples following well-established sampling procedures are suitable for making inferences from the sample population to the larger population that it is designed to represent
  - Snow-ball or convenience samples are not suitable for this survey

Overall Sampling Design cont'd

- Probability sampling in the context of household surveys:
  - Refers to the means by which elements of the target population are selected for inclusion in the survey
- In order to be cost-effective, most household surveys are not implemented as simple random samples
- Sampling procedure usually includes stratification to ensure that the selected sample actually is spread over geographic sub-areas and population subgroups
- This sampling design usually uses clusters of households in order to keep costs to manageable level

General Principals of the Survey on Aging in SSA

Target population: individuals age 60+ and older

Household sample: Nationally representative clustered random sample of households that include household members age 60+ yrs.

Selection of household members: All regular household members age 60+ in the sampled household and their spouses if these are age-eligible and co-resident

General Principals of the Survey on Aging in SSA

Use of an existing sampling frame: clustered random sample of households can only be obtained from existing sampling frame which is a complete list of statistical units covering the target population

- Census frame, complete list of villages/communities or sampling list from other nationally representative surveys

- Sampling frame: is a complete list of sampling units that entirely covers the target population

- Conventional sampling frame: list of enumeration areas (EA) from a recently completed census

- EA: geographic area which usually groups a number of households together for convenient counting purposes

General Principals of the Survey on Aging in SSA

Stratification: process in which the sample is designed into sub-groups or strata that are as homogeneous as possible;

- Within each stratum the sample is designed and selected independently;

Two-stage cluster sampling procedure:

- Cluster: a group of adjacent households which serves as the primary sampling unit (PSU)

General Principals of the Survey on Aging in SSA

Full coverage of the target population: should be nationally representative and cover 100% of the target population; that is no subpopulations age 60+ are systematically excluded;

Probability sampling: sample should be obtained as probabilistic sample based on existing sampling frame using established sampling procedures;

- Only way to obtain unbiased estimation and to be able to evaluate the sampling errors
- Excluded are purposive sampling, quota sampling, and other uncontrolled non-probability methods because they cannot provide evaluation of precision and confidence of survey findings

General Principals of the Survey on Aging in SSA

Full coverage of the target population: should be nationally representative and cover 100% of the target population; that is no subpopulations age 60+ are systematically excluded;

Probability sampling: sample should be obtained as probabilistic sample based on existing sampling frame using established sampling procedures;

- Only way to obtain unbiased estimation and to be able to evaluate the sampling errors
- Excluded are purposive sampling, quota sampling, and other uncontrolled non-probability methods because they cannot provide evaluation of precision and confidence of survey findings

Sample Size

- Sample size must take into account competing needs so that costs and precisions are optimally balanced
- Sample size must also address the needs of users who desire for sub-populations of sub-areas domains
- Sample size is determined by the trade-offs between survey precision, data quality, organizational capacities and survey budget;
- In the case of Malawi this is about 2,000 respondents (men and women)

Conducting a household listing and pre-selection of households

- Data quality is enhances if eligible households are preselected for participating in the study
- In many SSA countries recent and reliable household listings in EAs that carefully enumerates older individuals is not available
- Hence, we suggest to conduct a specific household listing in selected EAs that provides a well-grounded basis for selecting respondents
- Interviewers than interview only pre-selected eligible households

- STEPS:
  - Household listing operation conducted before the survey
  - Pre-selection of households from this list
  - Selected Households are interviewed

Overall Sampling Design cont'd

- Two stage sample design is well-established approach for implementing household surveys

- 1st stage: select a sample of EAs with probability proportional to size (PPS);
  - Within each stratum a sample of predetermined number of EAs is selected independently with probability proportional to size, where size is measured in terms of older individuals age 60+;
  - If size of pop age 60+ is not available, and variations in age structures are relatively modest, then total pop size can be used
  - All households in the EAs are listed
- 2nd stage: after complete listing in EAs, a fixed number of households with individuals age 60+ is selected by equal probability sampling in the EAs

Interviewing all individuals age 60+ in the HH

Advantages:

- Maximize the number of respondents for a given sample of HH
- Cost effective to achieve the sample size
- Analytical advantages so that interactions among spouses, within and between household variation of outcomes can be investigated

Disadvantages:

- Lower statistical power given the within household correlation of observations
- Logistical challenges in the fieldwork

Sample Take per Cluster

- How many eligible individuals to interview per EA
- DHS recommends 25-30 individuals
- Because there will be more than 1 age-eligible individual per household, less than 24-30 households per PA need to be selected
- If a sampled HH has 1.5 age-eligible individuals on average, than a sample take per cluster of 25-30 individuals results in the selection of 17-20 households per cluster
- With 2,000 individuals sample size: 67-80 clusters have to be selected
- If sample is stratified, these considerations should be conducted stratum-specific

Sample Take per Cluster

- This fixed sample take per cluster is:
  - Easy for survey management and implementation
  - But requires sampling weights that vary within clusters