

UTF/077/ZAM: Technical Assistance to the Zambia Aquaculture Enterprise Development Project (ZAEDP)

Session 3

Checklist 7

Study design and sampling

Dr Nihad Fejzic

Training Course on Development of an Active Surveillance for Epizootic ulcerative syndrome (EUS) and Tilapia lake virus (TiLV) using the FAO 12-point surveillance checklist (for non –specialist) and its implementation

University of Zambia, Lusaka, 14-17 October 2019

7	Study design and sampling	Used/described sampling frame a. Defined sampling units b. Described sampling method c. Explained consideration regarding sample size d. Describe sample selection process
----------	----------------------------------	---

Activities in each surveillance

- Determine what questions is being asked and how best to answer that
- Identify target population
- Choose the **right survey design**
- Prepare questionnaires and data collection form
- Decide if the survey is to use stratification
- Calculate **the best sample size**
- Plan field activities
- Train survey team
- Conduct a pilot survey

Activities in each survey:cont.

- **Select the sample**
- Carry out field work
- Collect information (from producers and/or animals)
- Process specimens ready for analysis
- Send the specimens to the laboratory
- Check the data for completeness and accuracy
- Enter the survey data and laboratory results into computer
- Check the data for mistakes during data entry
- Analyse the data to calculate estimates
- Report the data

Study design approaches

- **Passive surveillance**
- **Active surveillance**
 - Prevalence surveys
 - Production surveys
 - Incidence rate surveys
 - Surveys to demonstrate freedom from disease
- **Risk based surveillance**

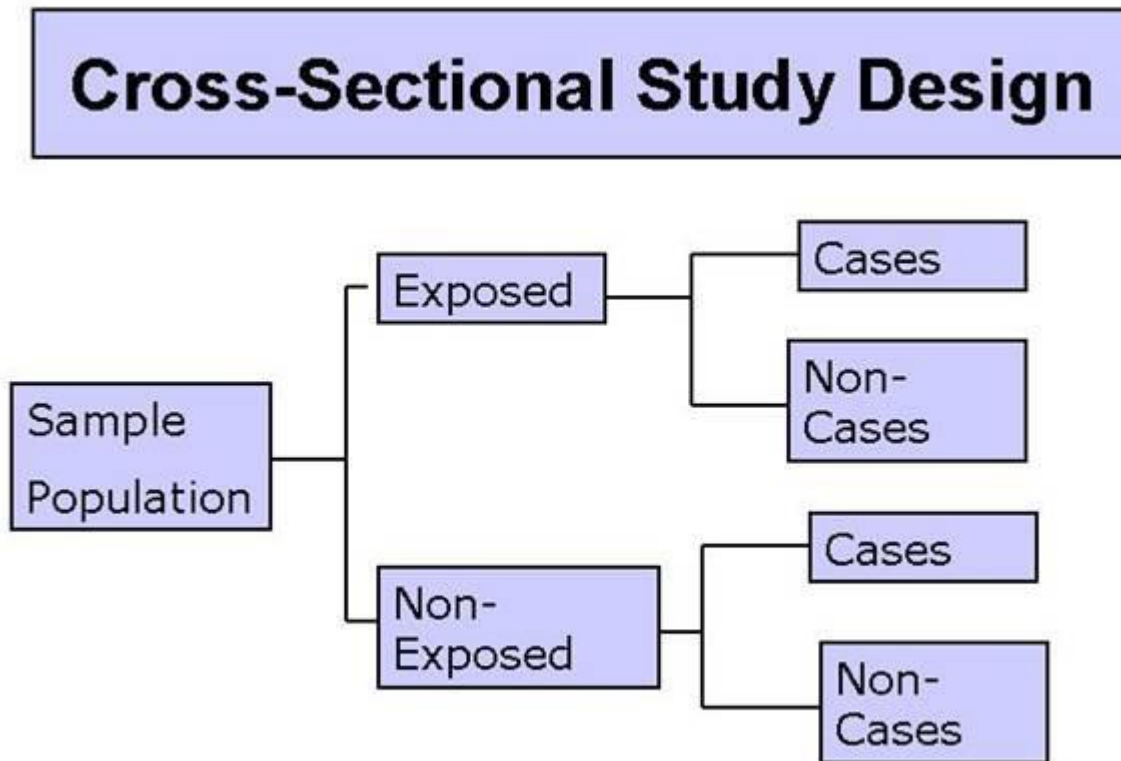
Surveys

- Surveys – surveillance / studies – epi research
 - Descriptive and explanatory
 - Experimental and observational
 - Retrospective and prospective
 - Cross section and longitudinal
 - Case – control and cohort

Main characteristics of different types of surveillance

- **Cross sectional survey** – prevalence study/outbreak investigation
 - Useful for common conditions
 - Short time frame
 - Low cost
 - Good for multiple risk factors or outcomes
 - Prospective or retrospective
 - No measure of disease incidence

Main characteristic of different types of surveillance



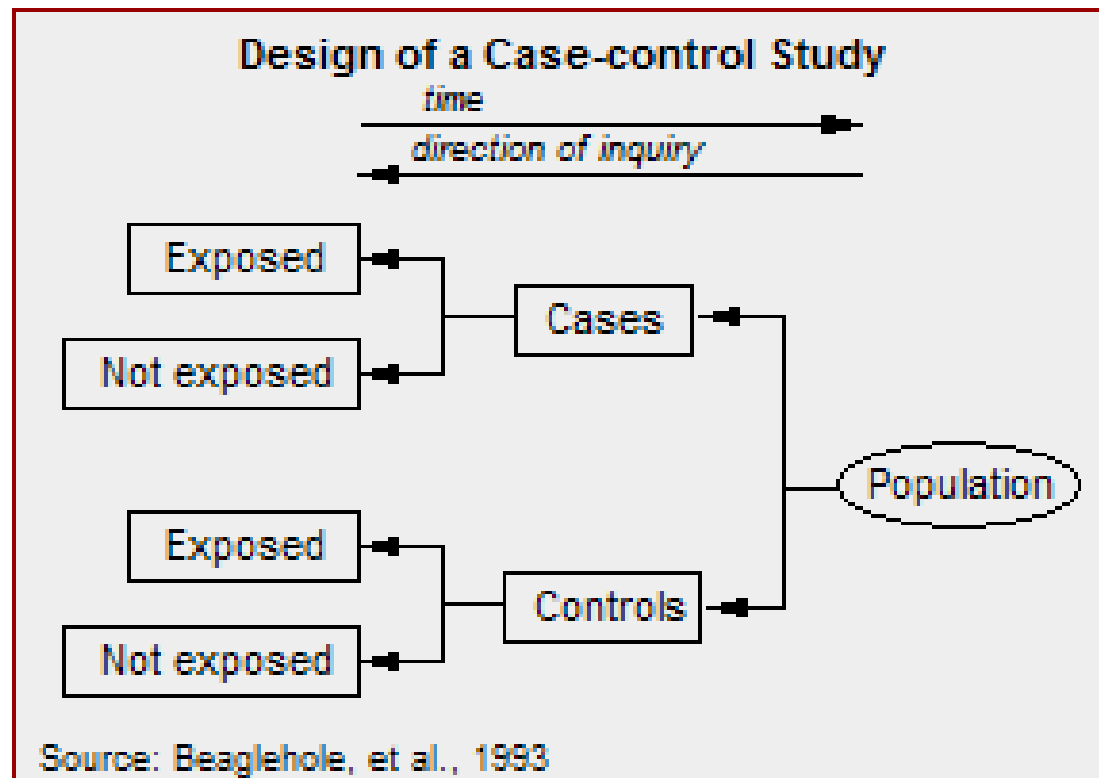
Main characteristics of different types of surveillance

- **Case – control studies**

- Groups according to disease status
- Mostly retrospective
- Useful for rare diseases
- Short time – low cost
- Good for multiple risk factors, not for multiple outcomes
- Non representative, prone to bias
- Can not calculate RR

Main characteristic of different types of surveillance

.

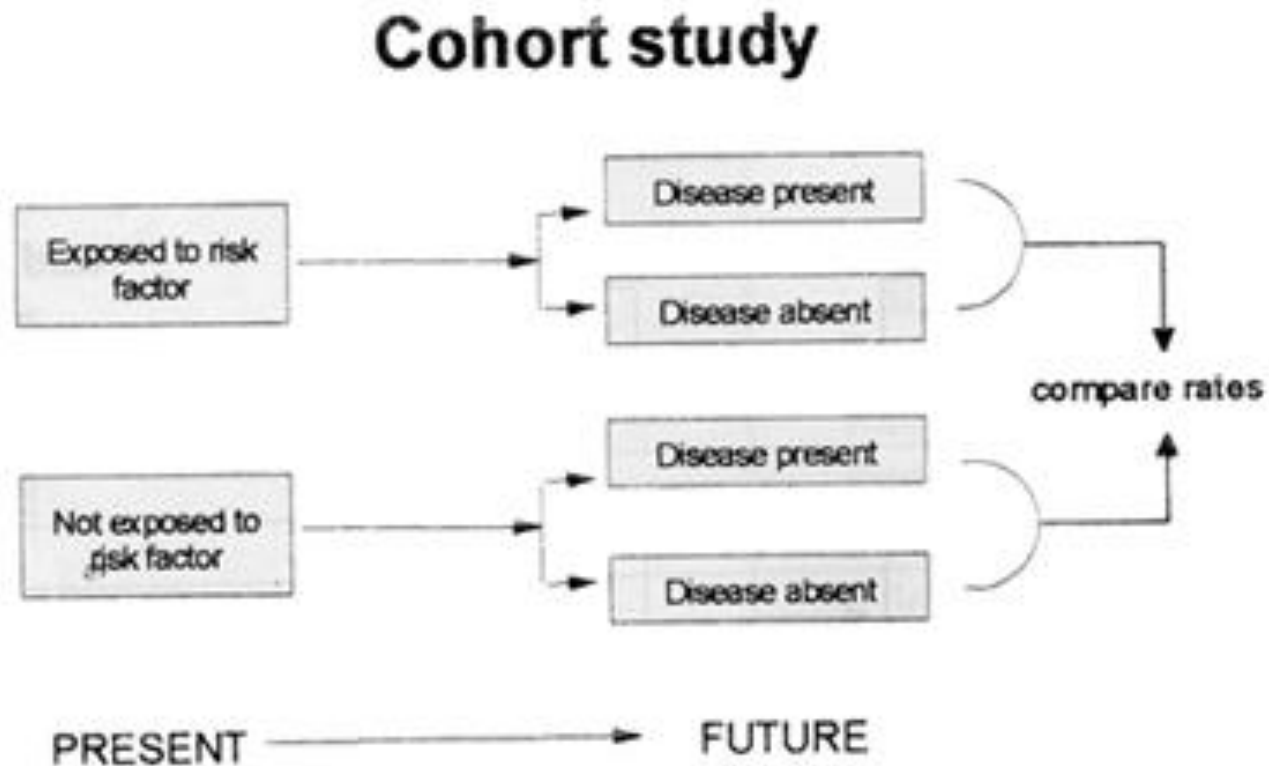


Main characteristics of different types of surveillance

- **Cohort study**

- Groups according to exposure
- Mostly prospective
- Useful for rare exposures
- Long time – high cost
- Good for multiple outcomes, not for multiple diseases
- Direct estimate of effect (incidence, RR)
- Control of bias

Main characteristics of different types of surveillance



Main characteristics of different types of surveillance

- Targeted surveillance – Risk based surveillance
 - More disease burden – less resources
 - Combines epidemiology, public health, economy, trade consequences
 - Looks for disease where is expected – risk assessment
 - Higher benefits-cost ratio
 - Estimation of effectiveness!?

Sampling

- Why different sampling methods?
- Farm management
 - Size of groups/pools?
 - With/without broodstock?
- Feasibility
 - Ensure randomness
 - Identify all animals/groups/farms
 - Access to all animals/groups/farms
- Disease biology
 - Infectious vs. Noninfectious diseases

- Sampling principles
- Sampling methods
- Sample size

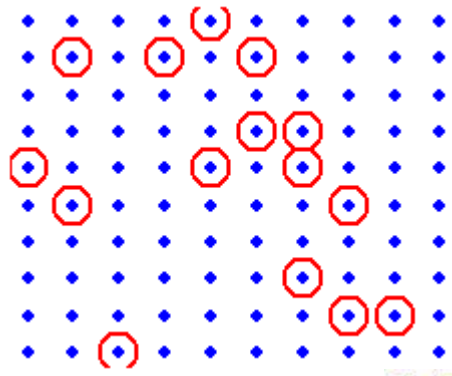
- Sampling method
 - Test entire population – census
 - Test sample (provide the greatest likelihood that the sample will be representative of the population)
 - Representative from population
 - Non representative
- Large populations - sampling frame not available -multi-stage sampling

Representative sampling – each individual in population has same and equal probability being selected into sample

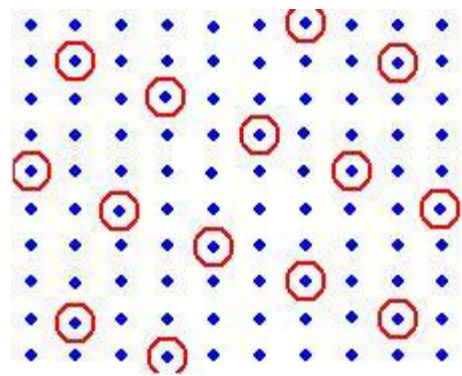
- Simple random sampling
- Systemic random sampling
- Stratified sampling
 - Proportionally stratified sampling
- Cluster sampling

Random sampling

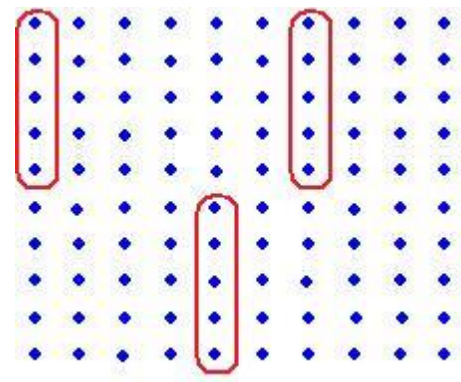
- Many farmed terrestrial animals are identified by an individual number
- Sampling frame in aquatic animals is different
- Random sampling can be applied using management practices (during grading or transfer of fish, during vaccination, during harvest)
- Most frequently used method in farm is capture sampling
- Likely to introduce some bias into the sample and it is important to be aware of the direction of bias.



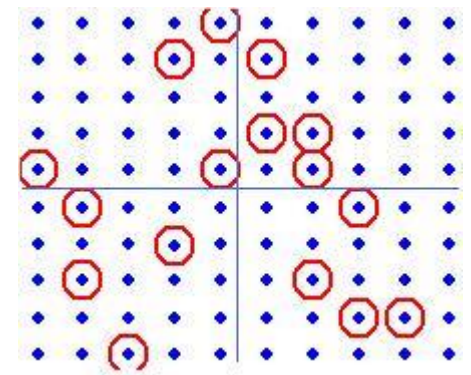
simple



systemic



stratified



cluster

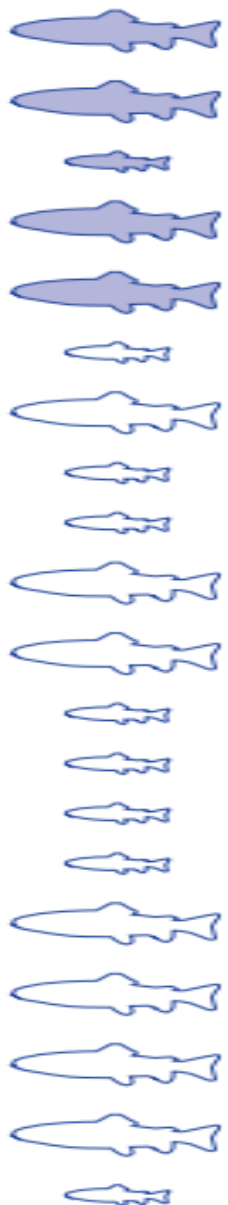
- Random sampling requires **sampling frame** (all individuals/units accessible and identified)
- NOTHING IS RANDOM IN RANDOM SAMPLING!!!
- Alternative use systemic/spatial sampling
- It should be possible to use sampling frame in aquaculture population for epidemiological or higher unit (vilages/farms..)
- For individual aquatic animals – no sampling frame
- Use any method to achieve random selection – documented and described
- Convenience sampling never acceptable

Non representative sampling

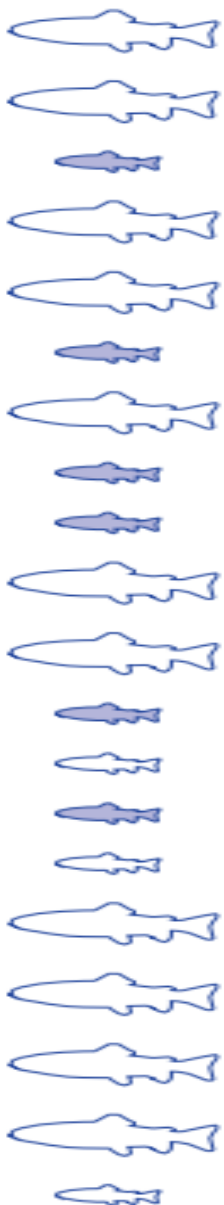
- Sampling of available animals – convenience sampling
- Targeted sampling
- Moribund /with lesions animals

Examples of non-probability sampling

Convenience

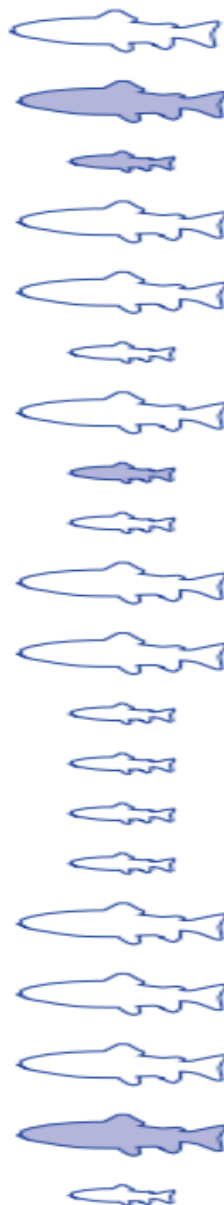


Purposive

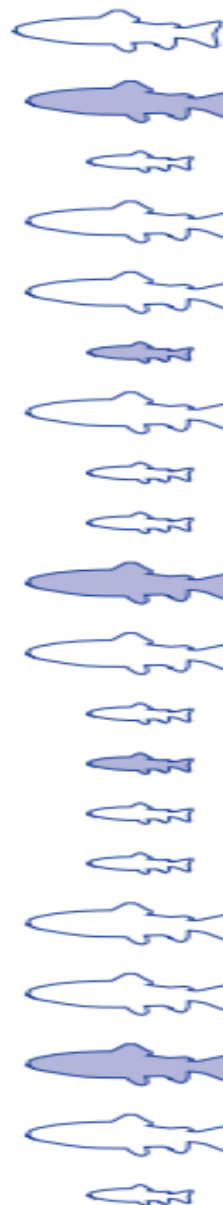


Examples of probability sampling

Random



Systematic



First five fish selected

Small fish selected for ease

Random numbers used to select fish

Second fish selected with random number, then every fourth fish selected

Sample size calculation

- The number of units to be sampled from a population should be calculated using a statistically valid technique, considering factors into account:
 - Imperfection of diagnostic test/s (Se, Sp)
 - the design *prevalence/s*
 - the level of confidence
- Other factors:
 - Population size (acceptable to assume infinitely large population)
 - The desired power of the survey
- Before – standardized tables
- Now – tailored calculation based on above factors

Design prevalence

- Design prevalence **is not** disease prevalence
- It forms part of the definition of the null hypothesis
- It is an abstract statement of what may be present in nature
- Design prevalence: minimum expected prevalence, maximum acceptable prevalence, minimum detectable prevalence

Design prevalence (DP)

- OIE Manual specified DP for certain terrestrial disease, not for aquatic
- At the individual animal level, the DP should be based on the biology of the infection
- A suitable DP value at the animal level may be
 - 1% - 5 % for infections that are transmitted slowly
 - Over 5 % for more contagious infections
- At higher levels (cage, pond, farm, village, etc) the DP usually reflect the prevalence of infection that is practically and reasonably able to be detected by a surveillance system.
- A suitable DP prevalence value for the first level of clustering (e.g. Proportion of infected farms in a zone) may be up to 2%

- For example, if a DP of 50 % is used when assessing prevalence of EUS on fish farms in country X, negative result will mean that we are 95 % confident that the prevalence of infection is less than 50 % !!!!!
- If a value is 1 %, negative result means that it is still possible that infection exist, but at prevalence of no more than 1 %
- Freedom of disease is zero prevalence (no one infected animals in country or zone)

Case study: closed system aquaculture

Type of sampling	Frequency	Class	Test type	Pathogens	Sample
Routine moribund	Quarterly, ongoing	Moribund or fresh mortality	Cell culture	Multiple exotic or emerging	70/year
Pathogen specific	Biannual, temporary	Random selection	RT-PCR	SAV, ISA, ..	165/year
Observational	Routine (daily, weekly, ongoing)	Moribund or fresh mortality	Veterinary investigation if above threshold	Pathogen clinical in this species	5 samples per event

EpiTools epidemiological calculators

This site is developed and maintained by Ausvet. The site is intended for use by epidemiologists and researchers involved in estimating disease prevalence or demonstrating freedom from disease through structured surveys, or in other epidemiological applications.

Surveillance utilities

- [1-Stage representative freedom surveys](#)
- [2-Stage representative freedom surveys](#)
- [Risk-based freedom surveys](#)
- [Random sampling from a population](#)
- [Estimating true prevalence](#)
- [Pooled prevalence calculator](#)
- [Survey Toolbox for livestock diseases and freedom in finite populations](#)
- [HerdPlus module for herd-sensitivity and freedom in finite populations](#)

Epidemiological studies

- [Sample size calculations](#)
- [Summarise categorical or continuous data](#)
- [Statistical significance testing](#)
- [Probability distributions](#)
- [Bioequivalence analysis](#)

Diagnostic tests

- [Application of diagnostic tests](#)

Case study data

- [GIS case study data from *Epidemiology for Field Veterinarians* text](#)
- [Epidemiological Problem Solving case studies and model answers](#)

Suggested citation: Sergeant, ESG, 2019. Epitools epidemiological calculators. Ausvet Pty Ltd. Available at: <http://epitools.ausvet.com.au>.

If you cite Epitools in your publications, please email the details or a copy of your paper to [Evan Sergeant](#) for inclusion in the reference list.



Sample size calculations

These utilities can be used to calculate required sample sizes to estimate a population mean or proportion, to detect significant differences between two means or two proportions or to estimate a true herd-level prevalence.

Epidemiological studies

[To estimate a single proportion](#)

[To estimate a single mean](#)

[Two proportions](#)

[Two means with equal sample size and equal variances](#)

[Two means with unequal sample size and unequal variances](#)

[To estimate true prevalence \(at animal or herd-level\)](#)

[Sample size for a cohort study](#)

[Sample size for a case-control study](#)

Sample size to demonstrate disease freedom

[Sample size assuming perfect test specificity](#)

[Sample size for pooled sampling in a large population](#)

[Sample size to achieve target confidence of freedom](#)

[Design prevalence required to achieve target population sensitivity for given sample size](#)

[FreeCalc sample size calculation for imperfect tests](#)

2-stage sampling, assuming **perfect** test specificity:

[Least-cost sample sizes where cluster sizes are known \(and select clusters for testing\).](#)

[Least-cost sample sizes where cluster sizes are *NOT* known.](#)



[Sample size calculations](#)

Sample size to estimate a proportion or apparent prevalence with specified precision

Estimated true proportion

Desired precision (+/-)

Confidence level

Population size (for finite populations)

This utility calculates the sample size required to estimate a proportion (or prevalence) with a specified level of confidence and precision.

Inputs are the assumed or estimated value for the proportion, the desired level of confidence, the desired precision of the estimate and the size of the population for limited population sizes. The desired precision of the estimate (also sometimes called the allowable or acceptable error in the estimate) is half the width of the desired confidence interval. For example if you would like the confidence interval width to be about 0.1 (10%) you would enter a precision of +/- 0.05 (5%).

The program outputs the sample sizes required to estimate the true value with the desired precision and confidence, for both an infinite population and for a population of the specified size. If population size is left blank or zero, only the sample size for an infinite population is calculated.

Note: Adjustment for finite population size may underestimate required sample size unless this is also taken into account when estimating variance and resulting confidence interval

Sample size to estimate a simple proportion (apparent prevalence)

Analysed: Tue Oct 08, 2019 @ 18:26 UTC

Inputs

Estimated Proportion	0.02
Desired precision of estimate	0.05
Confidence level	0.95
Population size	N/A

Results

Sample size required for specified inputs

Large population 31

Design: one stage sampling

- Sample frame available
- Random sample is collected (simple random sampling)

Design: two stage sampling

- Design 1: Probability proportional to size
- Design 2: Simple random sampling
- Design 3: Random geographic coordinate sampling

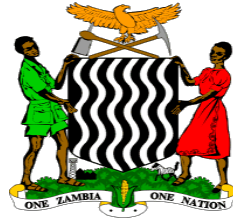
Example of EUS (if probability sampling is achievable)

Sampling frame	Exist			Don't exist		
Surveillance scenario	Infected	Unknown	Considered free	Infected	Unknown	Considered free
Sampling	Probabilistic			Convenience/purposive/hap hazard		
Design Prevalence	>5 %	50 %	< 2%			
Dg test Sn/Pp			100 %			
Confidence	95 %	95%	95%			
Sample size*						
First stage (epi unit)	73	385	165			
Second stage (within farm)	165	165	165			
Methods	Simple random	Simple random	Simple random			

Sample size, based on defined parameters, calculated using <http://epitools.ausvet.com.au/content.php?page=home>

Example of TiLV (if probability sampling is not achievable)

- Epidemiological Unit: A tilapia farm
- Unit of sampling: A mix of 30 moribund or sick tilapia from ponds at the farm.
- Total number of enrolled and participant farms: 40 to 60 tilapia farms, which should be visited twice (total field visits = 80 to 120, per country - at least 1+ve farm, at 2% Prev).
- Dynamic of sampling: 1st and 2nd semesters



UTF/077/ZAM: Technical Assistance to the Zambia Aquaculture Enterprise Development Project (ZAEDP)

Thank you for your attention