SAFOSO

Animal Health Matters.
For Safe Food Solutions.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs,
Education and Research EAER
State Secretariat for Economic Affairs SECO

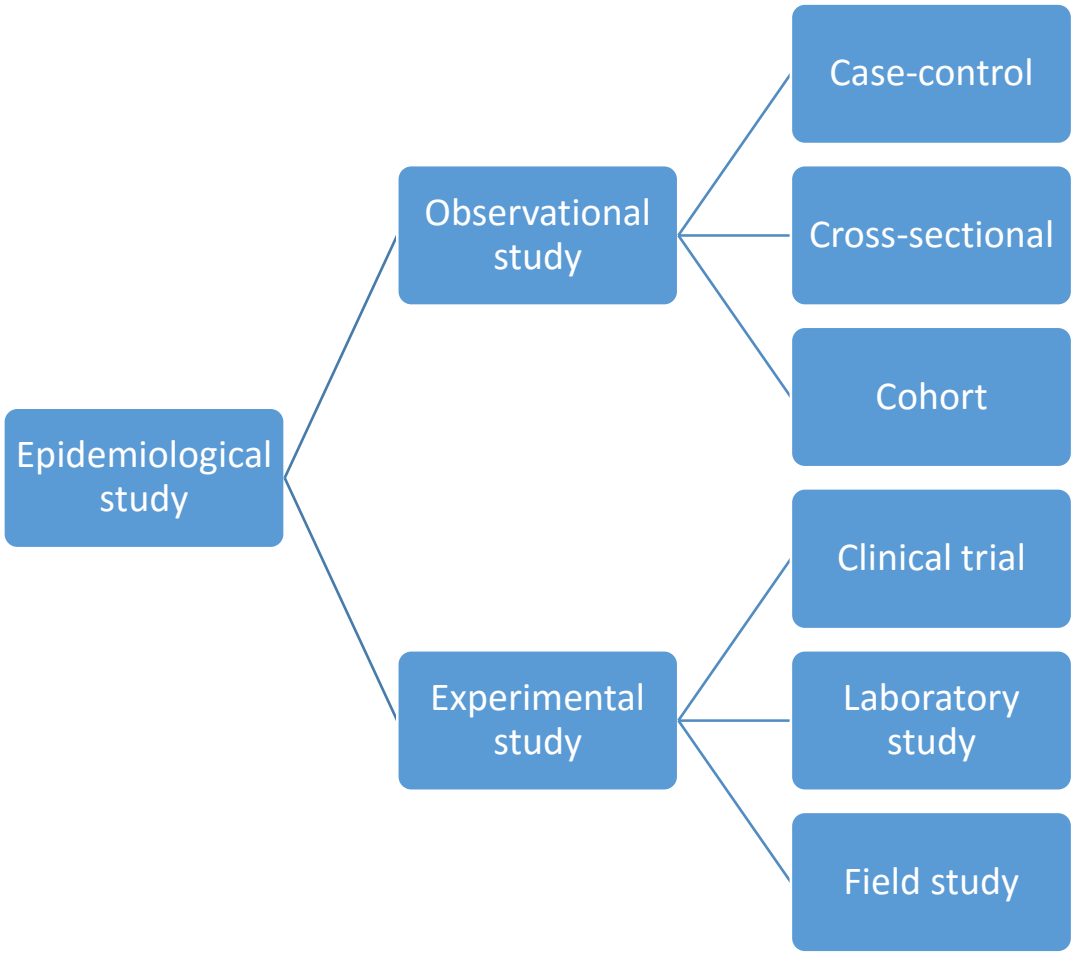
Introduction to basic concept of sampling and sample size

Marco De Nardi

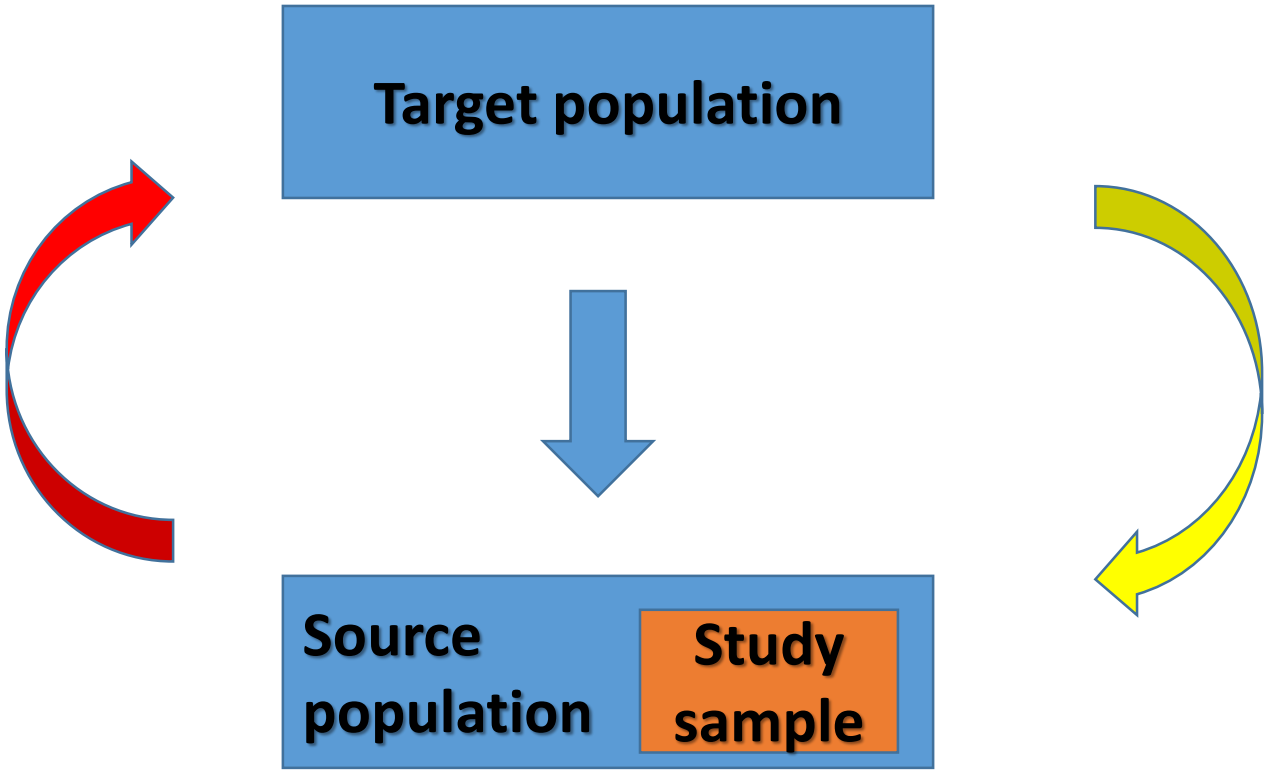
Active surveillance

- **<u>Variety of methods:</u>**

1. Exhaustive surveillance/Census
1. Surveillance on representative samples
1. Risk Based/Targeted surveillance
1. Sentinel
1. Combination of methods:


Epidemiological study
Observational study
Case-control
Cross-sectional
Cohort
Experimental study
Clinical trial
Laboratory study
Field study

Target population, Source population, Study sample



Target population

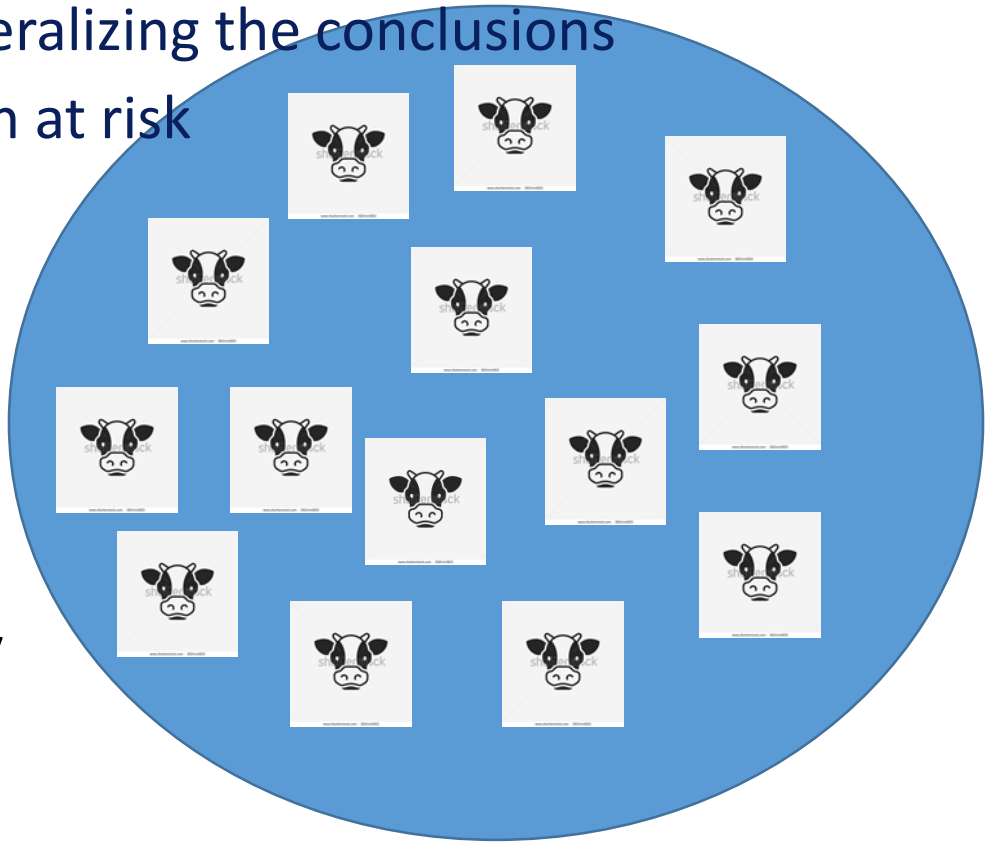
- It is the population targeted by the surveillance about which information is required
- Not always clearly defined (might be very difficult)
- It refers to the ENTIRE group of individuals or objects to which researchers are interested in generalizing the conclusions
- Ideally it should be the population at risk

- **Mastitis in Holstein Friesians in middled size farms in Ukraine:**

  Target population=

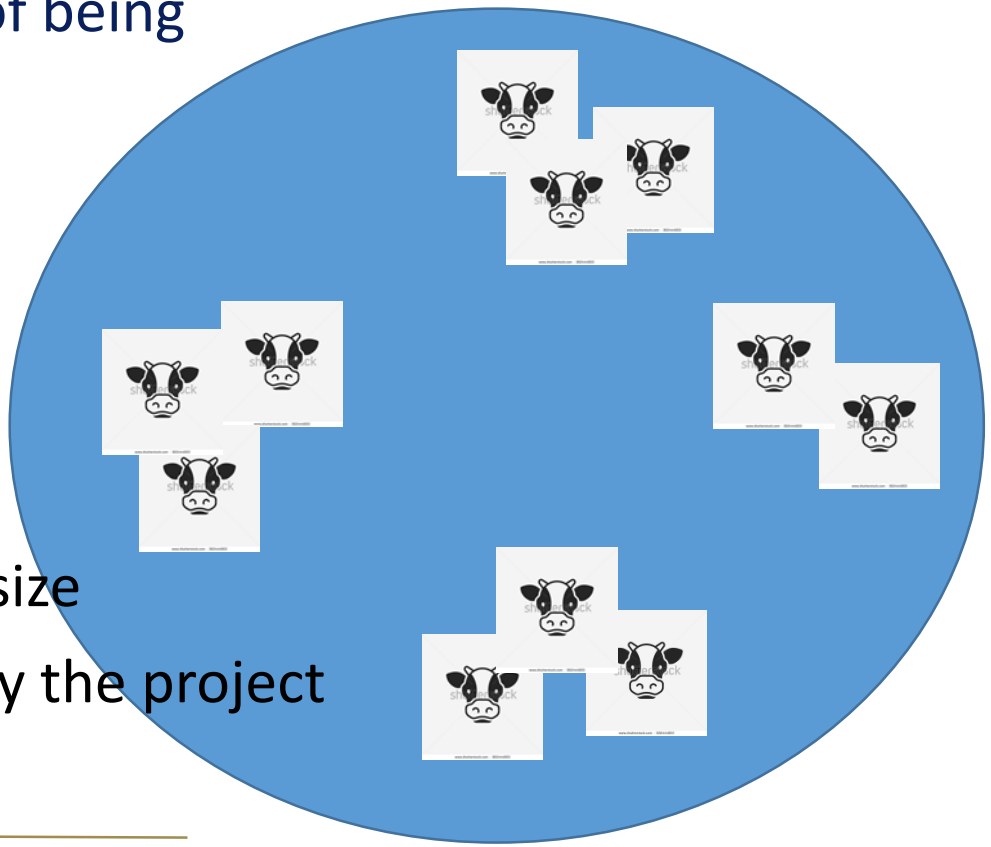
  all Holstein Friesians in

  middled size farms in the country



Source population

- It is a subset of the target population
- It is the population from which the individuals are sampled
- All individuals (animals, farms) should be “listable” and have a non-zero probability of being included in the study



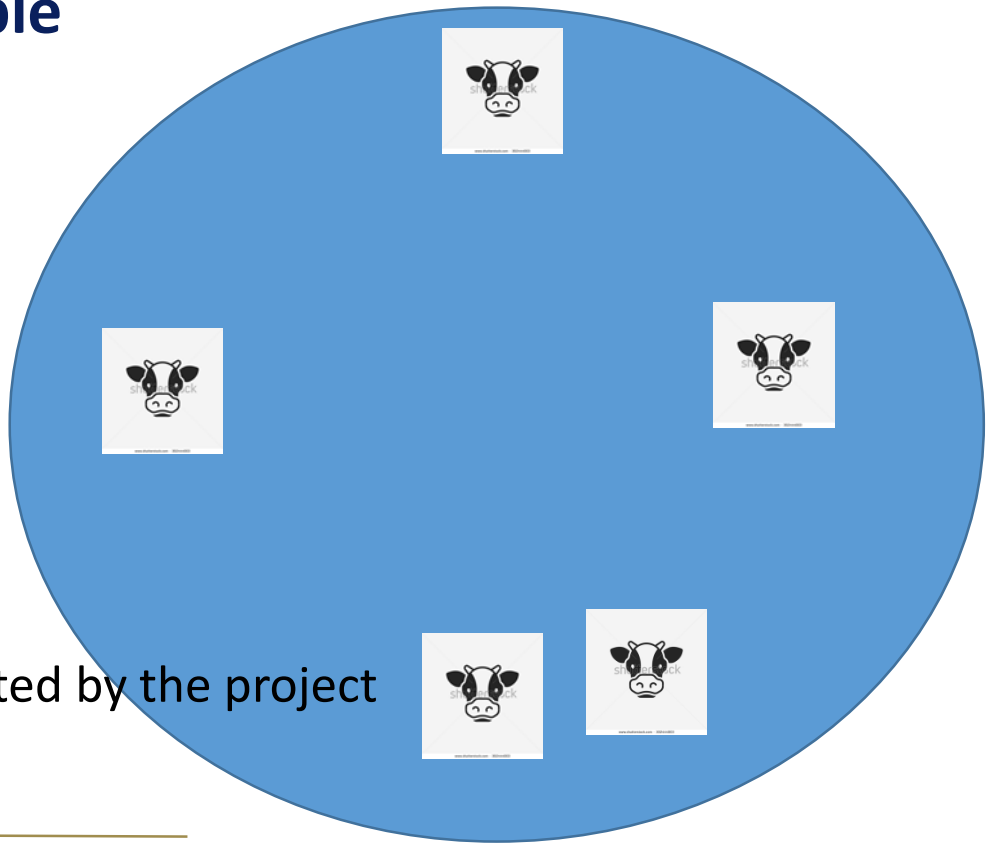
- **Mastitis in Holstein Friesians in middled size farms in Ukraine**

  Source population=

  all Holstein Friesians in middled size farms in the 4 regions targeted by the project

Study sample

- The study sample consists of the **individuals** (e.g. animals, herds, villages) that end up in the study
- Prior to conducting the study, the researchers would determine the necessary **sample size** required (i.e. how many individuals have to be included in the study)



Mastitis in Holstein Friesians in Ukraine:

Study sample=

a sample (statistically appropriate) of Holstein Friesians in the 4 regions selected by the project

Internal validity and external validity

- **Internal validity of the study:**

- Whether or not the study results (obtained from the study sample) are valid also for members of the source population

- This indicates whether or not the study has obtained the "correct" answer for the source population

- Statistics and epidemiological study design aims to maximize the probability of getting this answer correct

- **External validity of the study**

- Relates to how well those results can be generalized to the target population
- Subjective assessment of whether or not the source population is broadly representative of the target population
- The assessment of the external validity is much more difficult than assessing the internal validity

NBB: if the source population is not "representative" of the target population, then the results should not be generalized beyond the study population

Sampling frame, sampling units

- Before a sample is taken, members of the **source population** must be identified: **sampling frame** (list of villages, herds, animals etc…)

- Each member of the sampling frame is a **sampling unit**:
  - individual animals,
  - herds, abattoirs, administrative regions,

- Prevalence (or other indicators) can be calculated in relation to the different units:
  - individual animal prevalence = proportion of affected animals
  - herd prevalence = proportion of affected herds

Sampling frame: limits

a) Not always straighforward,

- List of members of the frame are incomplete,
- Information is obsolescent,
- Segments of the frame are untraceable,
- Lack of cooperation by some members-categories of the frame,
- Adopted sampling procedure are not random,

Results are BIASED estimates

Sampling frame

b) A frame doesn't exist (herds in nomadic areas)

Sampling Strategies

1) **Non-probability sampling**
2) **Probability sampling**

1) Non-probability sampling

- The choice of the sample is left to the investigator
  - convenience sampling: easily accessible s. units,
  - purposive sampling: sampling unit though to have characteristics (age, sex, breed…) similar with those of target population

- Very unlikely the sample will be truly representative of the study population

- Biased estimates

Sampling Strategies

2) Probability sampling

- The selection of the sample is made using a deliberate, unbiased process so that…

  each sampling unit in a group has the same probability of being selected (random sampling)

a) **simple random sampling**

a) **systematic sampling**

a) **stratified random sampling**

a) **cluster sampling**

a. Simple random sampling

- Draw the sampling-frame of all animals or sampling units:
- Use a random number table
- Computer programs (SurveyToolbox, Episcope, EpiInfo, Excel) or pocket calculators will select random samples,
- Published tables are available:

Random Number Table

7766	5509	2494	**0122**	9027
7481	**2870**	0407	4411	0892
0588	5444	6724	**3822**	8928
0579	9369	6921	1533	6138

Simple random sampling

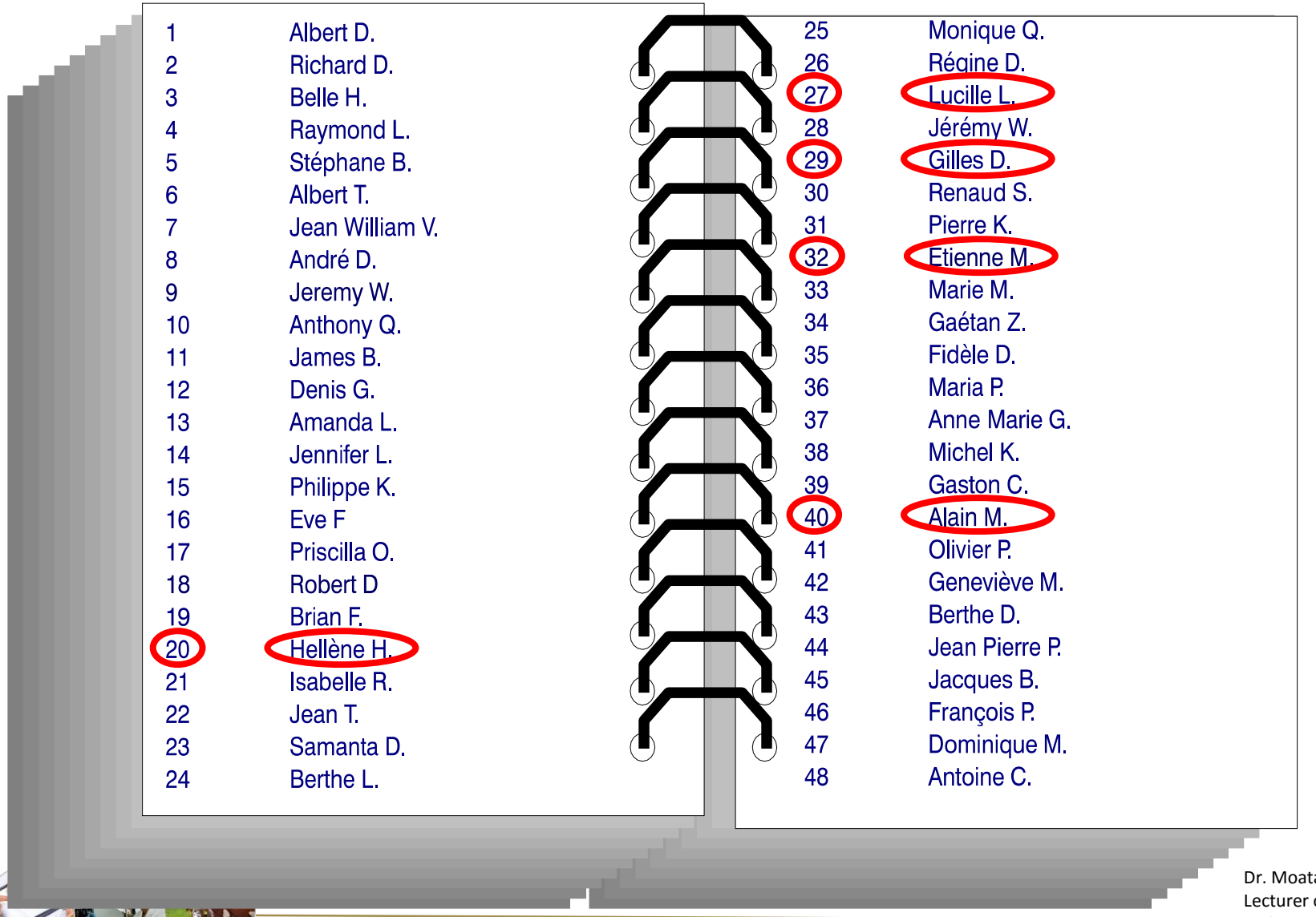
Advantages

- Simple to set up (if you have a complete list frame),
- Useful for specific situations e.g. selecting a sample of 10% of lab. confirmed rabid cases over the last 5 years
- Can be a component of more complex designs

Disadvantages

- Need a complete sampling frame
- May be difficult to implement in the field
- Tendency of over-representing or under-representing some sections of the sampling frame (i.e. s. r. s. of herds in a country: animals in very small herds have less chance to be selected!

Simple random sampling



Dr. Moataza Mahmoud Abdel Wahab
Lecturer of Biostatistics
High Institute of Public Health
University of Alexandria

b. Systematic sampling

- Give results similar to simple random sample

- Involves:

- Selection of sampling units at equal intervals,
- Only the first animal have to be selected randomly.

- Field application:

- animals entering a crash individually,

i.e. 1 animal in every 100 is required:

1) Randomly selection of 1 animal from the first 100 (s.r.s) : n.63
2) Systematic sampling: 63, 163, 263 etc…

- industrial quality control: selecting sample on conveyor belt

Systematic sampling

Advantages

- Simplifies the selection of samples : need to only randomize at the start
- Doesn't require knowledge of the total size of the study population

Disadvantages

- If population not randomly presented but fluctuate periodically then the estimation may biased.

  i.e. farmer only send animal for slaughter on Monday and samples are collected only on Thursday at the abattoir then farmer's animals will not be represented in the sample

c. Stratified Sampling

- Divide the study population into **strata** according to factors that are expected to influence the outcome (herds size, management system, geographical areas),
- Overcome the limit of s.r.s. of under-representing some group,
- Number of sampling unit in each strata:

a. Simple random sample

b. Proportional allocation

Stratified sampling

- Proportional allocation: the number of sampling units selected from each strata is proportional to the number of each strata:

i.e. 500.000 dairy cattle so divided:

Region x: 300.000 cattle

Region y: 200.000 cattle

Need 5% final sampling size: 25.000

Region x: 300.000 * 0.05 = 15.000

Region y: 200.000 * 0.05 = 10.000

Stratified sampling

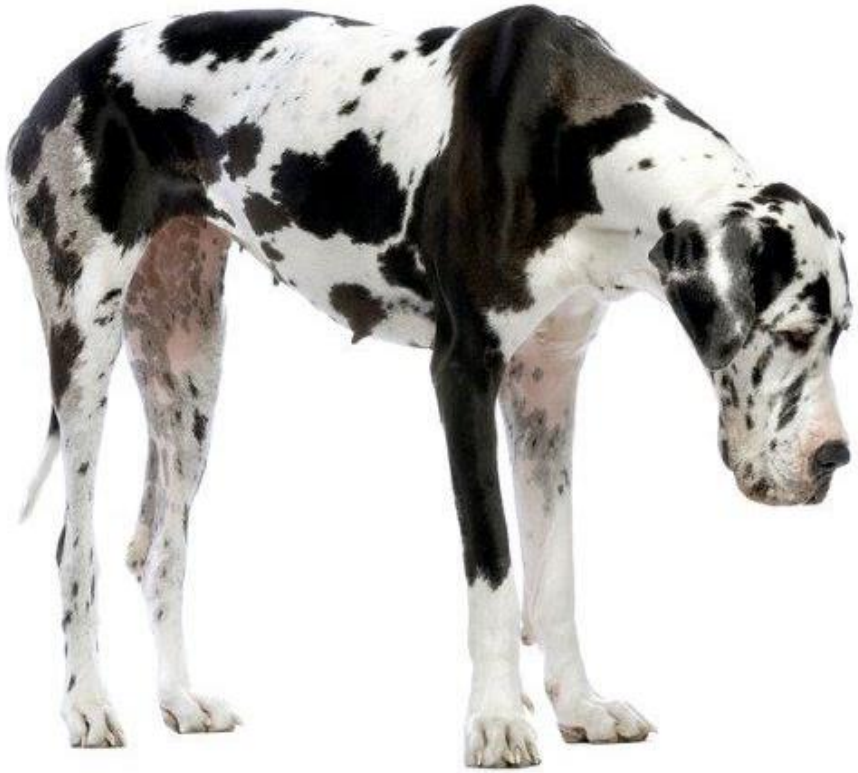
Advantages

- Increased precision of sample estimates may be obtained
- Within-stratum estimates are available

Disadvantages

- Knowledge of the elements in the population must be known in advance to place in appropriate strata
- Multiple factors may affect the outcome of interest but it is not practical to stratify by more than one or two factors
- Poor choice of factors for stratification could lead to a decrease in precision

Sample size



Sample size

- Depends on:

1) Non statistical considerations

- Man power, time, logistics,
- Resources,
- Availability of sampling frame

2) Statistical consideration

- Desired precision of the estimate of prevalence
- Expected prevalence of the disease

Sample size

- **Desired Precison**

- The ability of an estimator to determine, through representative sampling, the **true population value** of a variable,

- It can be expressed in terms of the bound on the **error of estimation** that can be tolerated:

- i.e. absolute error of ± 2% of an estimate of 40% of prevalence: 38-42%

  relative error: 2% of 40%. Range: 40 ± 0.8%

Sample size

- **Expected prevalence**

- Paradoxical?

- Prevalence with same degree of precision:

  expected prevalence level closer to 0% or 100% need smaller sample size than a prevalence around 50%,

- Information not available: estimates=informed guesses (guestimates)

Sample size to estimate prevalence

- **What do you need?**

- Expected prevalence (P),
- Desired absolute precison (d: 1%, 5%, 10%)
- Level of confidence in your results (C.I.: 90%, 95%, 99%)

- **Large population**: $n = 1.96^2 P(1-P)/d^2$ where 1.96 express 95% C.I.

- **Small population**: adjusted formula

Sample size to estimate **prevalence**

- **Imperfect tests** (Se, Sp < 100%)

$$n = \frac{((1.96/d)^2 * ((Se * P) + (1-Sp)(1-P))*((1-Se*P)-(1-Sp)(1-P)))}{(Se+Sp-1)^2}$$

- Published **tables** (*Cannon and Roe, 1982*)

- **Software**: Epiinfo, Stata etc…

Sample size to **detect disease**

- $n = (1-(1-P)^{1/d})((N-d/2)+1)$

N = population size (i.e. herds with 200 animals)

d = number of affected animals expected in the population

Highly contagious disease: use minimum seroprevalence (i.e. at least 5%)

Not highly contagious: use maximum sero-prevalence

P = probability of finding at least 1 case in the sample (i.e. 0.95)

n = 51 "If the seroprevalence is 5%, 51 animals need to be sampled to detect at least one seropositive animal with a 0.95 probability"

AI Active Surveillance in **Poultry**

- Decision of the **European Commission** of 13 April 2007 on the implementation of **surveillance programmes** for Avian Influenza in poultry and wild birds to be carried out in the Member States.

Sampling strategy

- Sampling shall be stratified throughout the territory of the whole country, taking into account:

(a) the number of holdings to be sampled;

that number shall be defined so as to ensure the identification of at least one infected holding if the prevalence of infected holdings is ≥ 5 % (95 % confidence interval- 99% for ducks, geese and turkeys);

Number of holdings to be sampled of each poultry category (except turkey, duck and goose holdings)

Number of holdings per poultry category per Member State	Number of holdings to be sampled
Up to 34	All
35-50	35
51-80	42
81-250	53
>250	60

Number of turkey, duck and goose holdings to be sampled

Number of holdings per poultry category per Member State	Number of holdings to be sampled
Up to 46	All
47-60	47
61-100	59
101-350	80
>350	90

Sampling strategy

(b) the number of birds sampled from each holding shall be defined so as to ensure 95 % probability of identifying at least one positive bird if the prevalence of sero-positive birds is ≥ 30 %.

Blood samples for serological examination shall be collected from at least 5 to 10 birds per holding for each poultry category (except ducks geese and quail),

If more than one shed is present on a holding it is recommended to take at least 5 birds per shed.

SAFOSO

Animal Health Matters.
For Safe Food Solutions.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs,
Education and Research EAER
State Secretariat for Economic Affairs SECO

Thanks