Food and Agriculture Organization of the United Nations

SUSTAINABLE DEVELOPMENT GOALS

VIRTUAL COURSE

26 March to 15 April 2021

Design of an Active Surveillance for Tilapia Lake Virus (TILV) Disease and Its Implementation

TCP/INT/3707: Strengthening biosecurity (policy and farm level) governance to deal with Tilapia lake virus

Food and Agriculture Organization of the United Nations

SUSTAINABLE DEVELOPMENT GOALS

CHECKLIST 9

9 April 2021

Data analysis

Nihad Fejzic
nihad.fejzic@vfs.unsa.ba

Fernando Mardones
fomardones@gmail.com

TCP/INT/3707: Strengthening biosecurity (policy and farm level) governance to deal with Tilapia lake virus

Learning objectives

- To understand the requirements and criteria for Checklist 9
- To understand data analysis methodology
- To apply basic epidemiological analysis on collected data
- To apply 2x2 tables (contingency table) in order to evaluate association between a possible factors (exposures) and disease (outcome)

Presentation topics

- Introduction to data analysis and test hypothesis
- Epidemic curves and spatial distribution
- Prevalence
- Incidence
- Contigency table
- Odds ratio

Analysis/interpretation/reporting

- According to surveillance objective and purpose, TiLV surveillance data analysis will give answer on different questions: If TiLV is present or absent, prevalence of disease if present and confirmation of disease free status of zone/country.

- Analysis and interpretation of data collected through surveillance will give results in accordance with surveillance purposes/objectives, survey design applied, sampling methodology/type applied, with prescribed sample/epi unit level (individual animals, ponds, farms, zones, etc) and restrictions of applied methodology.

- If **measurements of risk exposures are reported** – interpretation will be dependent on type of risk measure (AR, OR, RR) and relevant for restrictions of survey design applied.

Testing of hypothesis

- **Null hypothesis: disease is present** at a level equal to or grater than that specified by the design prevalence

- If we reject null hypothesis and accept **alternative hypothesis disease is not present** at the level equal to or greater then that specified by the design prevalence.

- **The required level of confidence** in the surveillance system must be greater than or equal to **95%**

- **Reject Ho= disease free**
- Probability of rejecting true null hypothesis = alfa (α) (disease present country declare free)
- Consequence: **the spread of infection between countries**
- 1-α = strength of evidence confirming null hypothesis – measure of confidence ≥95% (account for test characteristics)

- **Accept false Ho: country determines infected, in fact free (type II error)**
- **Power** of the analysis is probability of avoiding a type II error
- No international standards
- Consequence: **loss of trade opportunity, however no increased risk for spreading of disease, more samples**
- In practice, many test system involve one confirmatory test that is considered for all intents and purposes to have a specificity of 100 %

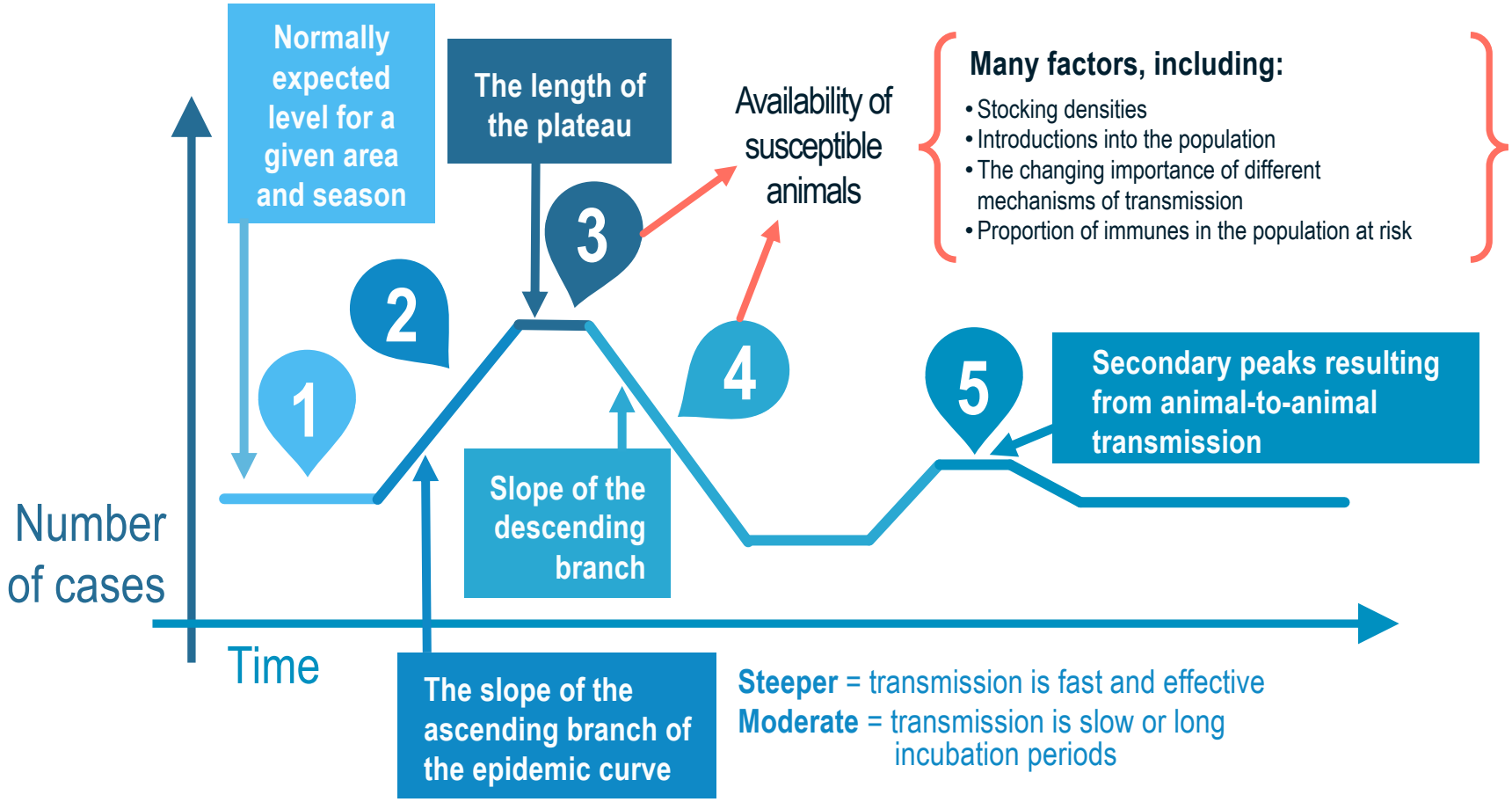
Analysis of surveillance data:

- Account for **survey design**
- Account for **diagnostic test imperfection**
- Account for **design prevalence(s)**

Epidemic curves

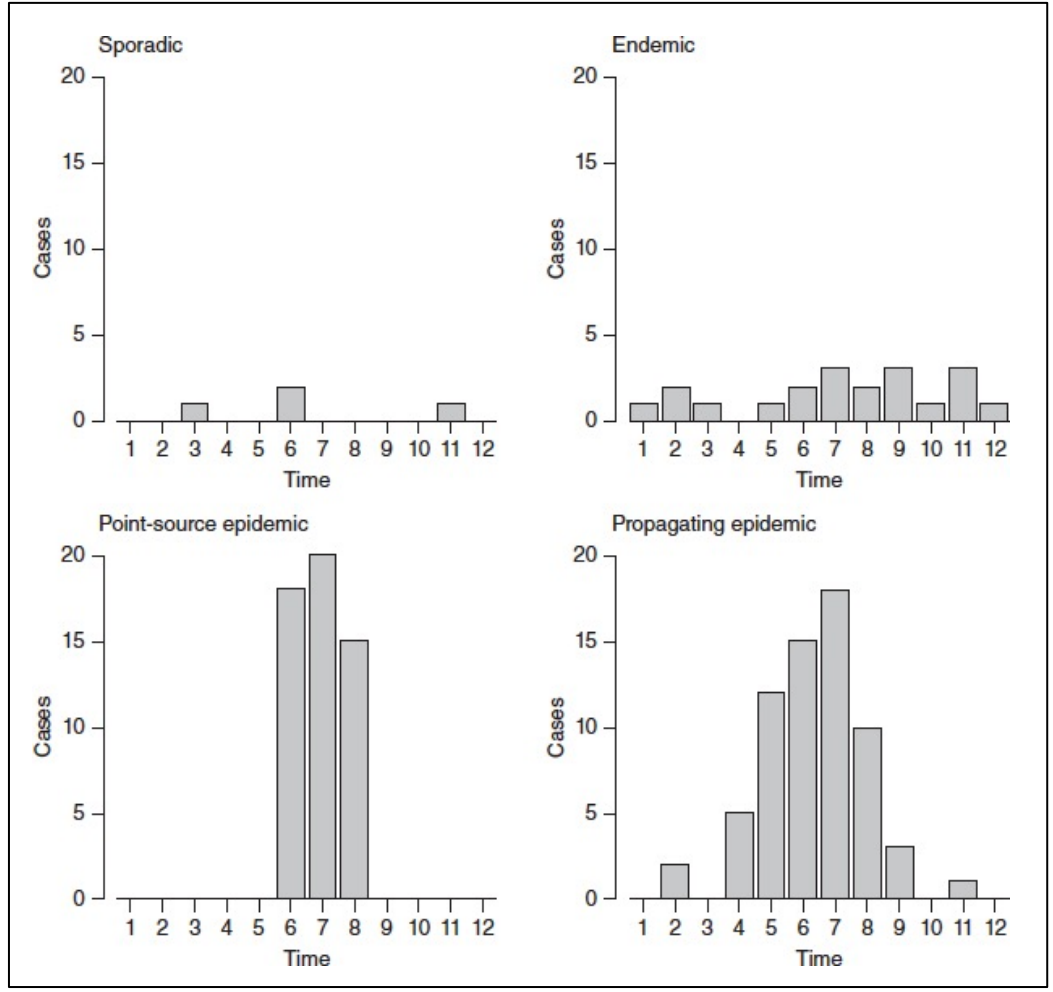
- Summary of the temporal pattern of disease events
- Provides a visual display of the scale or magnitude of the event and the rate at which new cases are occurring.
- Represents in a graphic form the onset of cases of the disease, either as a histogram, a bar graph or a frequency polygon.
- The frequency of new cases (or outbreaks) is plotted on the y-axis over a time scale on the x-axis.
- A typical epidemic curve may be conceived of as having four and occasionally five segments

Five stages of an epidemic curve


Normally expected level for a given area and season
The length of the plateau
Availability of susceptible animals
Many factors, including:
• Stocking densities
• Introductions into the population
• The changing importance of different mechanisms of transmission
• Proportion of immunes in the population at risk
1
2
3
4
5
Secondary peaks resulting from animal-to-animal transmission
Slope of the descending branch
Number of cases
Time
The slope of the ascending branch of the epidemic curve
Steeper = transmission is fast and effective
Moderate = transmission is slow or long incubation periods

Different types of epidemic curve

- For **sporadic** disease, most time periods have no cases, with occasional periods experiencing small numbers of cases.
- For the **endemic** disease, the number of cases fluctuates between time periods but remains at a fairly stable level.
- For the **epidemic** disease, the number of cases increases sharply from its initial endemic level and then declines slowly back to that level.


Sporadic
Endemic
Point-source epidemic
Propagating epidemic
Cases
Time
0
5
10
15
20
1 2 3 4 5 6 7 8 9 10 11 12

The shape of the curve

An epidemic is said to occur when the frequency of cases (or outbreaks) in a population clearly exceeds the normally expected level for a given area and season.

Spatial distribution of diseases

- Disease occurrence can be described by spatial pattern
- Such pattern are typically influenced by differences in environment and farming practices
- Computerized mapping and statistical methods for spatial analysis permit formal analysis of spatial pattern

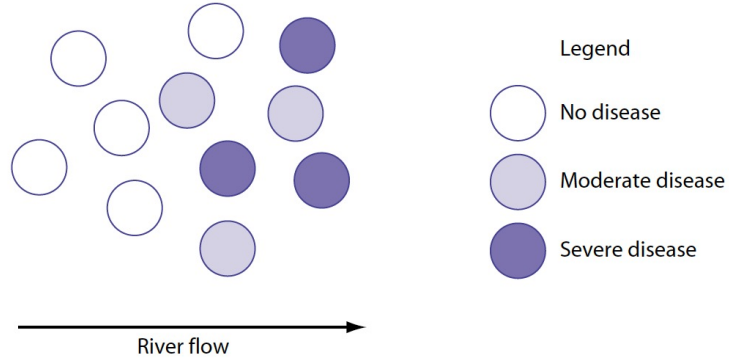


Figure 2.8 Theoretical spatial pattern of disease in fish culture cages in a river

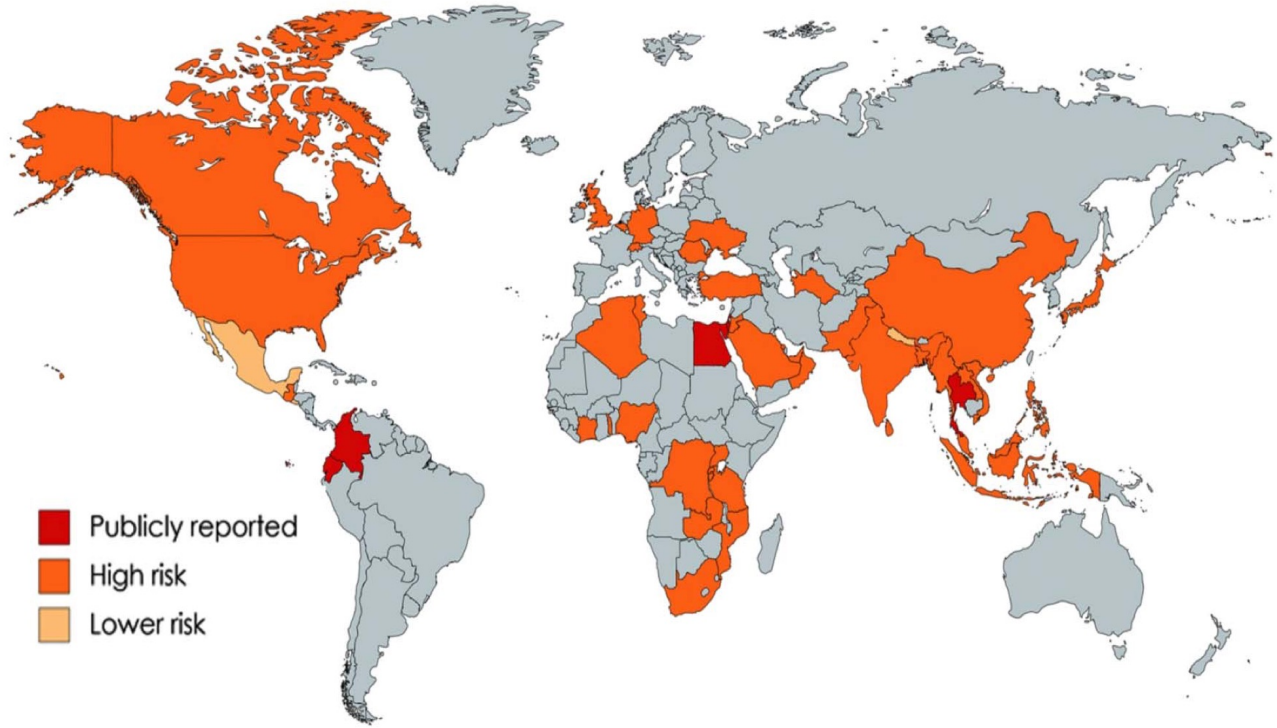
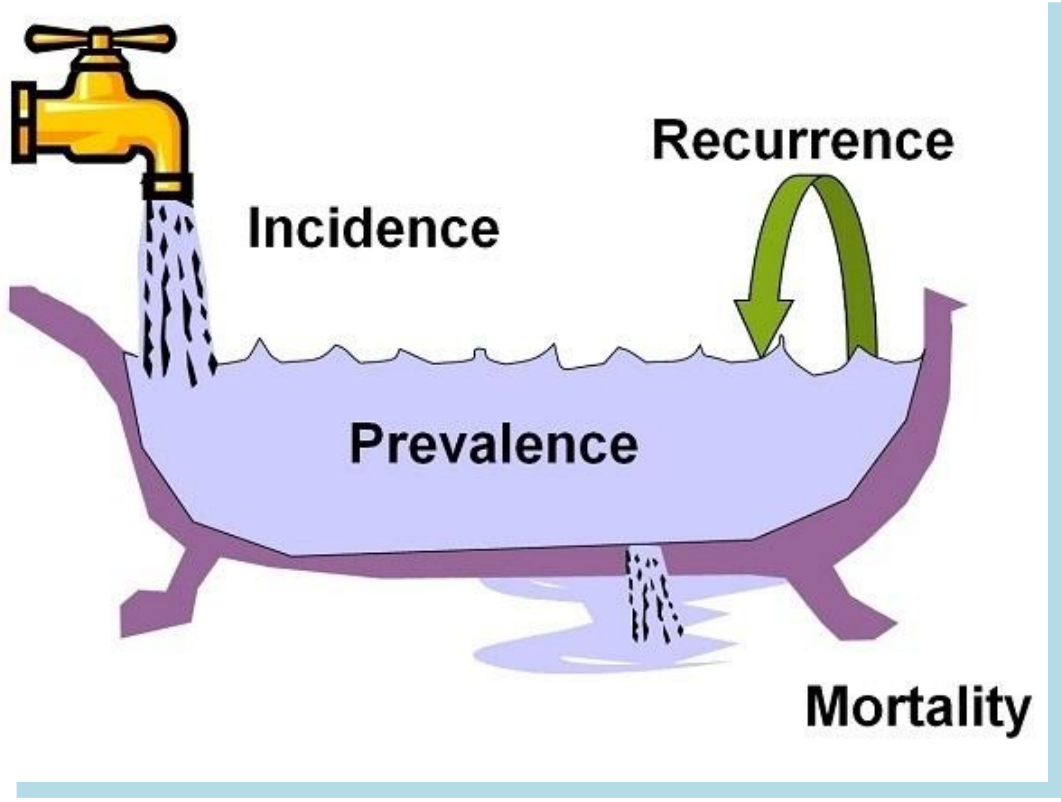


Fig. 3. The geographical distribution map of tilapia lake virus (TiLV). Red colour indicated 5 countries with firmed evidence of the presence of TiLV. Orange and light orange colors represent 40 and 3 countries with respective high risk and lower risk of TiLV spread through translocation of tilapia fry/fingerlings that may have been infected with TiLV. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Measuring disease frequency

- By absolute and relative numbers
- Absolute numbers – number of cases
- Relative numbers – proportions, rates, odds

Measuring disease frequency



Prevalence

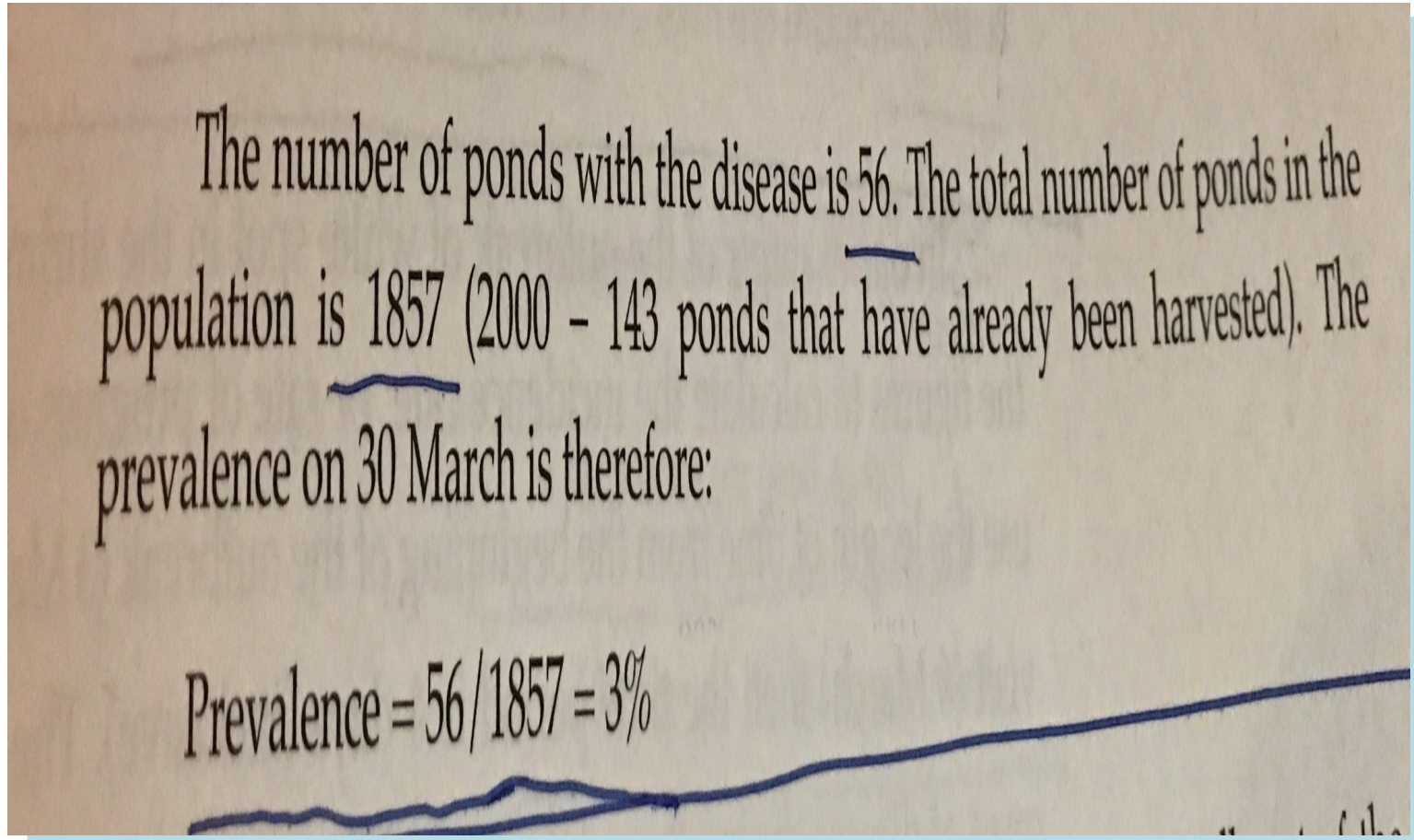
- is a measure of **the number of fishes/ponds/farms/villages** with the disease at one point in the time, as a proportion of the total number of animals/ponds/farms/villages in the population at risk

- Is the number of sick fishes, ponds or other units of interest at at single point of time as proportion of the total population at risk at that time

- **P = Number of cases at one point of time/population at risk at the same point of time**

Example

An intensive shrimp farming area with 2000 ponds suffers an outbreak of white spot disease. The first ponds start to show signs of sick shrimp on 3 March. By 29 March many ponds have sick and dead shrimp. The local fisheries officer visits on 30 March. On that day, the officer counts 56 ponds showing signs of disease, and the producers report that a further 143 have already been emergency harvested, and 28 ponds had been diseased, but recovered. There are 1801 apparently unaffected ponds remaining. What is the pond prevalence of white spot disease in the area on 30 March?



The number of ponds with the disease is 56. The total number of ponds in the population is 1857 (2000 – 143 ponds that have already been harvested). The prevalence on 30 March is therefore:

Prevalence = 56/1857 = 3%

Prevalence

- National level prevalence
- Province/district/village level prevalence
- Farm level prevalence
- Animal (fish) level prevalence

Incidence

- Is a measure/rate of the average speed at which the disease is spreading;
- Incidence rate is the total number of new cases of disease divided by the total time that each animal in the population was at risk of getting the disease.
- Incidence rate measure the number of new cases over period of time

Incidence rate

- Incidence rate=total new cases during a period of time/average number of animals at risk x time period

- Example of incidence rate:
  - Total number of new TiLV cases 30
  - Population at risk 100
  - Time period 10 days
- Result:
  - 0,03 cases per pond per day
  - 3 cases per pond per 10 days

Prevalence versus incidence rate

- A disease with a high incidence rate but of very short duration will have a relatively low prevalence

- A disease of relatively low incidence rate with long duration will have a high prevalence.

	Incidence rate	Cumulative incidence	Prevalence
Numerator	New cases occurring during a period of time among a group initially free of the disease in question	New cases occurring during a period of time among a group initially free of the disease in question	Existing cases at a point in time
Denominator	Sum of time periods during which individuals could have developed disease	All at-risk individuals present at the beginning of the period	All at-risk individuals examined, including cases and non-cases
Time	From beginning of follow-up until disease occurs for each individual	Duration of period of observation	Single point in time
How measured	Prospective cohort study	Prospective cohort study	Cross-sectional study
Interpretation	Rapidity with which new cases develop over a defined time period	Risk of developing disease in defined time period	Risk of having disease at a particular point in time

Measuring disease frequency

- **Disease prevalence** – proportion of diseased animals in a population
- Static measure
- Good for common, low contagious, chronic diseases

- **Disease incidence** – rate of new cases of disease in a population
- Dynamic measure
- Good for acute, highly contagious diseases
- Measure of disease risk

Do not forget!

1. Incidence is a dynamic measure of disease whereas prevalence is only a static measure of disease.

2. Incidence and prevalence are related. The prevalence of disease in a PAR reflects both the incidence of new cases of disease and the duration of disease in individual cases:

   Prevalence = incidence x duration under certain conditions.

3. Changes in the incidence or the duration of a disease will change the prevalence. The incidence rate is usually greater than prevalence if the disease is short in duration and/or fatal. Prevalence is usually greater than the incidence if the disease is chronic in nature.

Do not forget!

4. True rate describes the average speed at which the event of interest occurs per unit of animal time at risk. It is often called incidence rate. True rate has no meaning on the individual level. However, it can be interpreted on a population basis.

5. Risk rate (cumulative incidence rate) provides a direct estimate of the likelihood of an animal experiencing the event of interest during the internal time period. Risk rate has a meaning on an individual basis as well as on a population basis.

Do not forget!

6. Counting the PAR (i.e. the denominator):
   i. With prevalence, the total number of animals examined during the time you counted the frequency of disease is the denominator.
   ii. With incidence rates, however, we are looking at a population over a period of time; therefore, the number of animals at risk can change. There are a number of ways to deal with this problem, but the two most common are:
      - Use an estimate of the population, either by counting the population at a time midway in the time interval, or by taking the average of the population at the beginning and end of the time interval.
      - Calculate the population on each day of the time interval and arrive at the number of animal-days-at-risk (incidence density rate).

Attack rate or attack risk

- A specific type of incidence rate that applies to outbreaks or situations where the period of observation is relatively short.

- An attack rate is the number of cases of the disease divided by the number of animals at risk at the beginning of the outbreak (the outbreak covers a defined time interval).

- Attack rate = number of animals affected/number of animals exposed e.g., the attack rate can be used to measure mortality due to ISAV virus infection in farmed Atlantic salmon. If, over a 10-day period, 35,000 of the 50,000 salmon in a cage die, the attack rate is 0.7 or 70%.

Measures of Mortality (Death) in a Population

1. Crude death rate:

$$\frac{\text{NUMERATOR}}{\text{DENOMINATOR}} = \frac{\text{deaths in a given time}}{\text{total population at risk}}$$

2. Cause-specific death rate (a measure of the risk of death from a specific cause):

$$\frac{\text{NUMERATOR}}{\text{DENOMINATOR}} = \frac{\text{deaths in a given time due to the disease of interest}}{\text{total PAR}}$$

3. Age/cause-specific death rate (limits numerator and denominator to specific age/cause of interest):

$$\frac{\text{NUMERATOR}}{\text{DENOMINATOR}} = \frac{\text{deaths in a given time in the group of interest}}{\text{total PAR for the group of interest}}$$

Measures of disease attributes among the ill or dead animals

1. Case recovery rate (actually a proportion rather than a true rate):

$$\frac{\text{NUMERATOR}}{\text{DENOMINATOR}} = \frac{\text{number of cases recovering}}{\text{total cases for which outcome known}}$$

2. Case fatality rate (a proportion rather than a true rate):

$$\frac{\text{NUMERATOR}}{\text{DENOMINATOR}} = \frac{\text{number of cases dying}}{\text{total cases for which outcome is known}}$$

3. Proportional mortality rate: The proportion of total deaths attributable to a specific cause:

$$\frac{\text{NUMERATOR}}{\text{DENOMINATOR}} = \frac{\text{deaths due to specific cause of interest}}{\text{total deaths in population}}$$

Comparing Disease Frequencies

- Since incidence rates reflect risk, then the incidence rates (or attack rates) of two different groups may be compared in a ratio called the risk ratio or relative risk (RR).
- The RR compares disease among individuals of the one group to another group.
- Relative risk and a number of other commonly used measures can be used to compare disease frequency between risk groups in the population.

Relative risk

- The **relative risk** (or risk ratio, relative incidence rate ratio, etc.) is the ratio of the incidence rate (IR) in the exposed group to the IR in the unexposed group.

- You can use cumulative incidence, incidence density or attack rate for the calculations, as long as you use the same type of measure in both parts of the ratio.

- Since RR is the ratio of incidences, RR cannot be calculated for case - control studies (because incidence cannot be calculated in case-control studies).

Calculation of relative risk from an attack rate table

	Diseased	Not diseased	Total
Exposed	a	b	$a+b$
Not exposed	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d$

$$\text{Incidence}(\text{exposed}) = \frac{a}{a+c}$$

$$\text{Incidence}(\text{unexposed}) = \frac{c}{c+d}$$

$$\text{Relative risk} = \frac{a/(a+c)}{c/(c+d)}$$

Relative Risk can vary from zero to infinity

- An estimate of how much more likely disease is to occur in the exposed group compared to the unexposed group and has a null value (no association or no increase in risk) of 1, which is equivalent to equal incidence rates.

- If RR is >1 the factor increases the risk of disease. If RR is <1 the factor decreases the risk of disease.

- However, a confidence interval for the estimate should always be calculated and the value can only be considered to vary significantly from 1 if the confidence interval does not include 1.

- As a rule of thumb, RR values greater than about 3 (or less than about 0.33) are considered potentially biologically important.

	Diseased	Not diseased
Exposed	a	b
Unexposed	c	d

The relative risk is a/a+b / c/c+d

	TiLV positive farm	TiLV negative farm
Exposed (using TiLV non certified eggs)	10	2
Unexposed (using TilV certified eggs)	3	5

Relative risk or risk ratio

- 10/(10+2) = 10/12= 0.83
- 3/(3+5) = 3/8= 0,37
- **RR = 0,83/0,37 = 2,24**
- **Interpretation:**
  - RR less then 1; exposure is protective
  - RR = 0; exposure no effect
  - RR more than 1; exposure is **positively associated with disease**

In class exercise

Higher Tilapia density	Diseased	Not diseased	Total
Yes	43	11	54
No	3	18	21
Total	46	27	75

Please, report the incidence in exposed, non exposed and the RR…

Interpretation

- Tilapia farms that were at higher densities were almost six times more likely to become sick than were those that were at lower densities.
- In fact, 95% confidence limits for the RR estimate are 1.96 to 16, suggesting that this difference is likely to be statistically significant.
- We could also use a Chi-square statistical test to see if this relationship is significant.

Odds ratio (OR)

- Measure of the strength of association that is very useful in epidemiological studies of all types (cohort, case-control, cross- sectional).

- As the name implies, this is a ratio of the odds of exposure: non-exposure in disease-specific groups or the ratio of the odds of disease/no disease in exposure-specific groups.

Odds Ratio

$$Odds\ Ratio = \frac{a/b}{c/d} = \frac{ad}{bc} \quad or \quad \frac{a/c}{b/d} = \frac{ad}{bc}$$

	Diseased	Not diseased	Total
Exposed	*a*	*b*	*a+b*
Not exposed	*c*	*d*	*c+d*
Total	*a+c*	*b+d*	*a+b+c+d*

Interpretation OR

- Similar to the relative risk: values >1 indicate increased risk, while values <1 indicate a protective factor. Just like a RR, the null value of the OR is 1, and the OR has no units.

- The OR's significance can also be tested using a Chi-square statistical test or confidence intervals. In the above example the 95% confidence interval is from 5.8 to 94.

- The odds of developing disease were 23.5 times greater for those farms that were under high densities compared with those that did not were exposed to higher densities.

Comparing the RR and the OR

- Attributable risk also called risk difference is the absolute difference between the two incidence rates.
- $AR = IR_{exposed} - IR_{unexposed} = \left(\frac{a}{a/b}\right) - \left(\frac{c}{c/d}\right)$
- The AR tells us how much of the disease in the exposed group is attributable to being exposed.
- It implies the rate of disease that could be prevented if the exposure were removed completely from the population.

Attributable Risk

- Attributable risk also called risk difference is the absolute difference between the two incidence rates.

- $AR = IR_{exposed} - IR_{unexposed} = \left(\frac{a}{a/b}\right) - \left(\frac{c}{c/d}\right)$

- The AR tells us how much of the disease in the exposed group is attributable to being exposed.

- It implies the rate of disease that could be prevented if the exposure were removed completely from the population.

Attributable Risk

- If you get a negative AR, the AR is telling you the rate of disease that was prevented by the exposure.
- The AR has the same units as the IR and can theoretically vary from −1 to +1; the null value is zero.
- Remember that the RR has no units and has a null value of 1.0.

Results

Expected values

	Disease +	Disease -	Total
Exposure +	12.5	87.5	100
Exposure -	12.5	87.5	100
Total	25	175	200

Measures of association

	Estimate	Lower 95% limit	Upper 95% limit
Overall Incidence/Prevalence	0.12	0.08	0.18
Incidence/Prevalence in Exposed	0.2	0.13	0.29
Incidence/Prevalence in Unexposed	0.05	0.02	0.11
Odds Ratio	4.75	1.71	13.23
Relative Risk (Cross-sectional study)	4	1.56	10.24
Attributable Risk	0.15	0.06	0.24
Attributable fraction in Exposed	0.75		
Population Attributable Risk	0.08	-0.01	0.16
Population Attributable Fraction	0.6		

Key points to be noted in relation with data analysis of TILV surveillance:

- After the completion of survey treat data analysis and preparation of report as a priority. The user of the information need to know what is happening, report cases of EUS after confirmation and produce reports as soon as possible.

- When analyzing data and producing reports, keep in mind that objective of the survey is to answer one or two questions and that question is usually answered by one number (example prevalence of EUS is 5 % of farms, or there is no disease case confirmed).

Key points to be noted in relation with data analysis of TILV surveillance:

- After the completion of survey treat data analysis and preparation of report as a priority. The user of the information need to know what is happening, report cases of EUS after confirmation and produce reports as soon as possible.

- When analyzing data and producing reports, keep in mind that objective of the survey is to answer one or two questions and that question is usually answered by one number (example prevalence of EUS is 5 % of farms, or there is no disease case confirmed).

Key points to be noted in relation with data analysis of TILV surveillance:

- Information should be presented in a way that makes it easy and quick to understand and should be distributed to everybody who may need it, and everybody who participated in generating it. Distribution of the result internationally and neighboring countries will help in coordinating regional approaches and efforts to TiLV control.

- If possible, the results should be publish in international journals. A demonstrated ability to carry out high quality surveillance using internationally recognized standards greatly improves the international reputation of a country's aquatic authorities.

Food and Agriculture Organization of the United Nations

SUSTAINABLE DEVELOPMENT GOALS

Thank you for your attention!

Nihad Fejzic
nihad.fejzic@vfs.unsa.ba

Fernando Mardones
fomardones@gmail.com

TCP/INT/3707:
Strengthening biosecurity (policy and farm level) governance to deal with Tilapia lake virus

This was also made possible with the support of the Norwegian Agency for Development Cooperation under the project GCP/GLO/979/NOR Improving Biosecurity Governance and Legal Framework for Efficient and Sustainable Aquaculture Production.

Norad