The Online R-FETPV 1st Module : Basic Epidemiology and Surveillance Data Analysis

5 April -28 May 2021

USAID FROM THE AMERICAN PEOPLE

Department of Livestock Development

Food and Agriculture Organization of the United Nations

Lesson 4:Apply Basic Measures and Tools of Descriptive Epidemiology

David Castellan DVM, MPVM, ACVPM, ACPV

USAID FROM THE AMERICAN PEOPLE

Department of Livestock Development

Food and Agriculture Organization of the United Nations

Regional R-FETPV Field Epidemiology Training Program for Veterinarian

Content/Outline

At the end of this lesson, you will be able to:

4.1. Describe and explain the importance of describing a health or disease event

4.2. Describe and apply the types of data including continuous vs categorical data

4.3. Describe and apply the concepts of ratios and proportions

4.4. Describe and apply the concepts of incidence rates and prevalence

4.5. Describe and apply other measures including attack rate (AR), case fatality rate (CFR) and other useful measures of disease impact

Lesson 4: Part 1 of 4

Review: Epidemiology

Epidemiology…
is a scientific discipline…
that involves the study of…
the frequency…
and distribution…
of health and disease…
in populations…
in order to find risk factors…
for prevention and control.

Usefulness of Data to Explain Disease Events

Descriptive Epidemiology

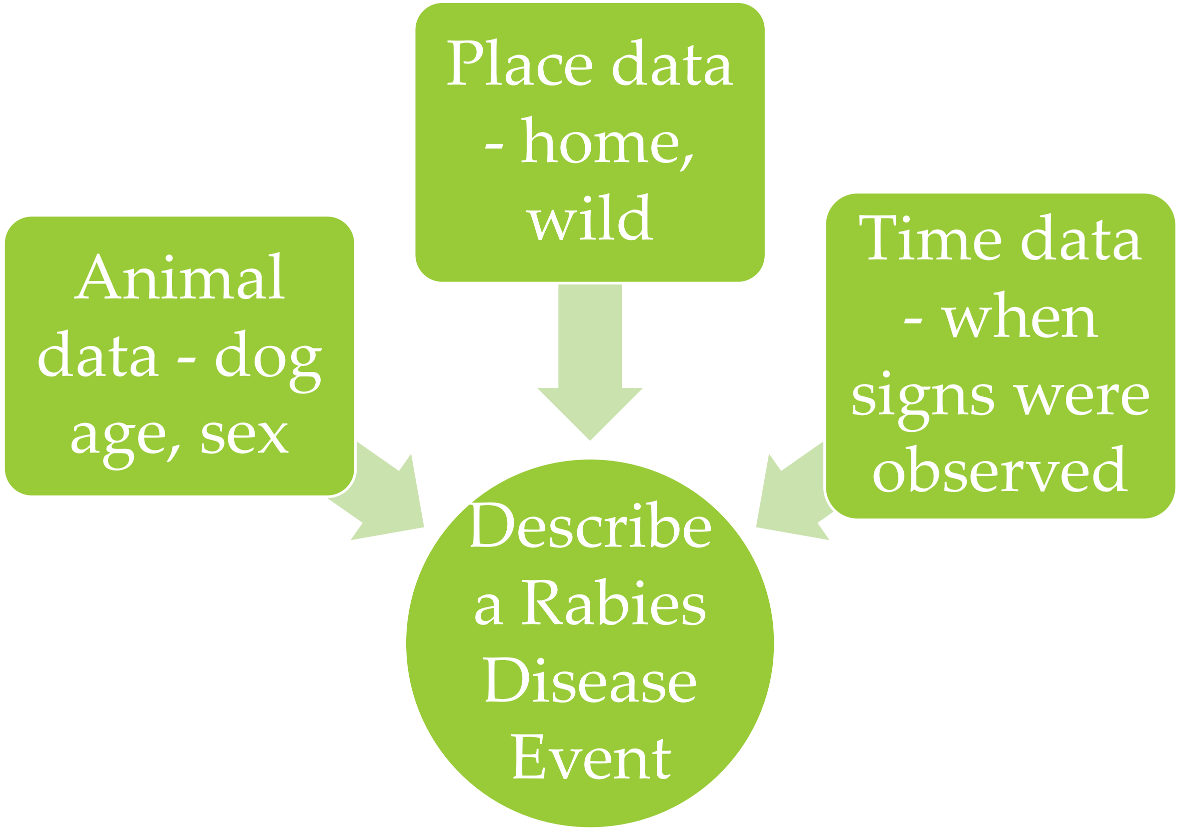
- **What** happened? (DISEASE)
- **When** did it occur? (TIME)
- **Where** did happen? (PLACE)
- **Who** is affected (pigs, poultry, etc.) ? (ANIMAL)

Analytical Epidemiology

- **Why** did it happen (research hypothesis)? (RISK FACTORS)
- **How** will the data be processed and used? (INTERVENTIONS)

Importance of Describing Animal and Human Health and Disease Events

- Disease is not random: There are patterns we must describe so that we can understand how to prevent and control such disease events
  - Example: dog deaths in a rabies outbreak or epidemic
- We must collect the following detailed data to 'fully describe' an animal or human health or disease event
- After describing the event we can then analyse data to identify risk factors leading to the disease event



Data, Information and Knowledge

DATA
(FACTS)

→

INFORMATION
(SHARED/PROCESSED DATA)

→

KNOWLEDGE
(INTERPRETED INFORMATION)

We count 20 cows on a farm

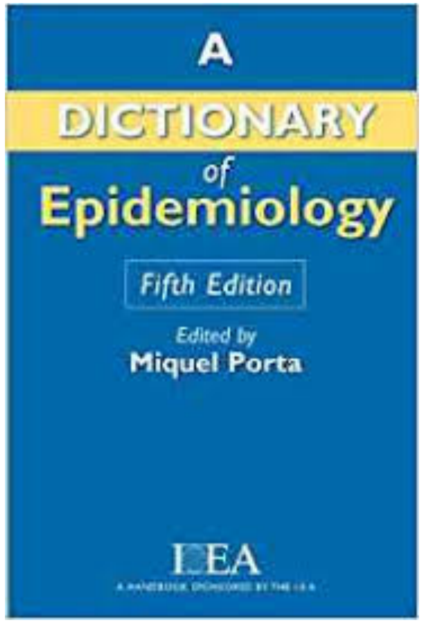
I count 10 cows are Holstein and you count 10 cows are native breed

This farm has 20 cattle including two breeds

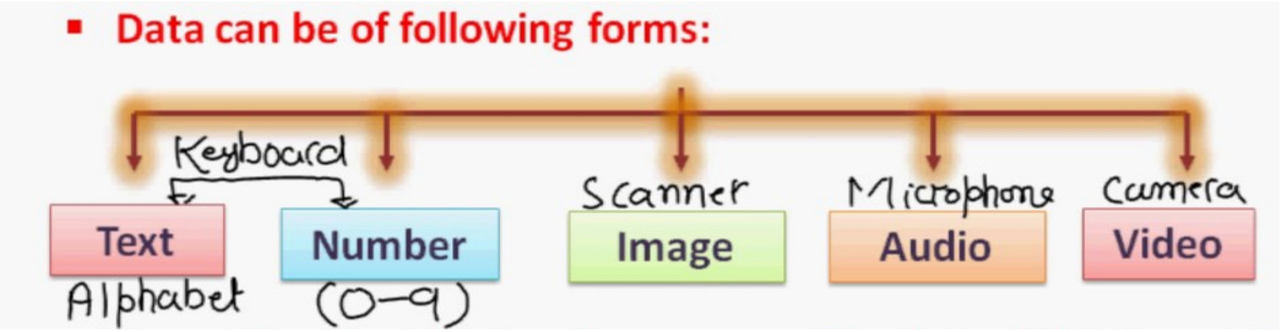
What is data?

DATA: "A collection of items of information." *Dictionary of Epidemiology, 2008*

- The individual elements of measurements recorded during data collection



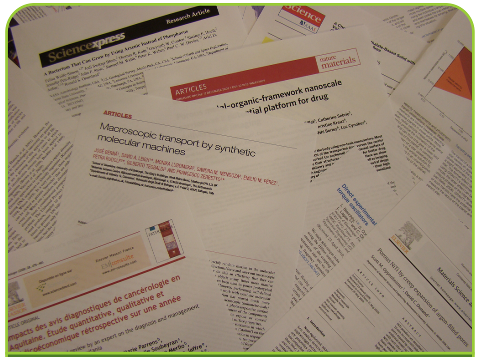
Google Images



Forms of Data

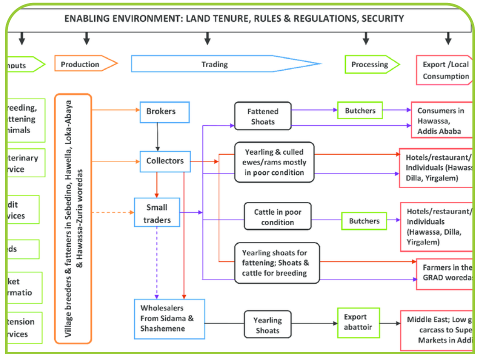
	2010	
Species	Diagnosed cases (%)	D ((
Buffaloes	67 (3.1)	0
Cattle	1879 (86.4)*	4
Goats	220 (10.1)	1(
Sheep	8 (0.4)	0

Numbers (Forms, Excel spreadsheets)

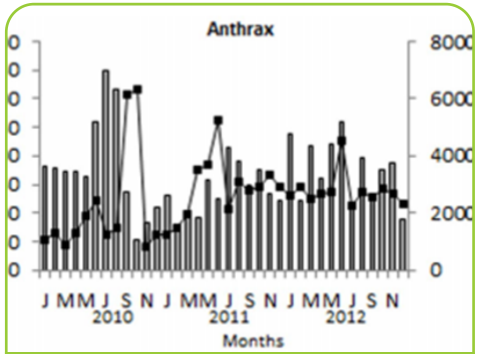


Words (Reports)

Maps (Paper / electronic)



Images (Diagrams)



Graphic symbols

Reference: Google Images

1. Data Collection: Data variables

(Ch. 1.4 Terrestrial Animal Health Code. OIE, 2018)

Populations at risk (ANIMAL)

- Species, breed, age, sex
- Clinical signs and PM findings
- Production drops

Location (PLACE)

- x,y coordinates
- Infected premises
- Suspect premises
- Dangerous contacts

Timeframe (TIME)

- Onset of clinical signs
- Onset of death
- Movements in/out

Goal for Descriptive Data

- Since **data is intended to be used** and shared then it must be:
  - High quality
  - Dependable
  - Have a clear meaning
  - Organized
  - Understandable to the person or organization receiving the data

Usefulness of Data

- **BEFORE** collecting data consider the following questions:

  - **Why** collect data and what do you need it for?
  - **What** data is needed and at what level of detail?
  - **How** will the data be processed and used?
  - **When** will the data be collected?
  - **Where** will the data be collected?

Tools to Collect Data

- Forms
- Questionnaires
- Excel spreadsheets
- Official reports and records
- Field notes and records
- GPS transponders and maps
- Telephone, fax, computer
- Photo camera
- Bar code and electronic identification instruments

Sources of Data

- Surveys, surveillance and monitoring
- Markets
- Slaughter plants
- Laboratory reports
- Official government reports
- Communities
- Quarantine stations
- Traders and marketers
- Others…

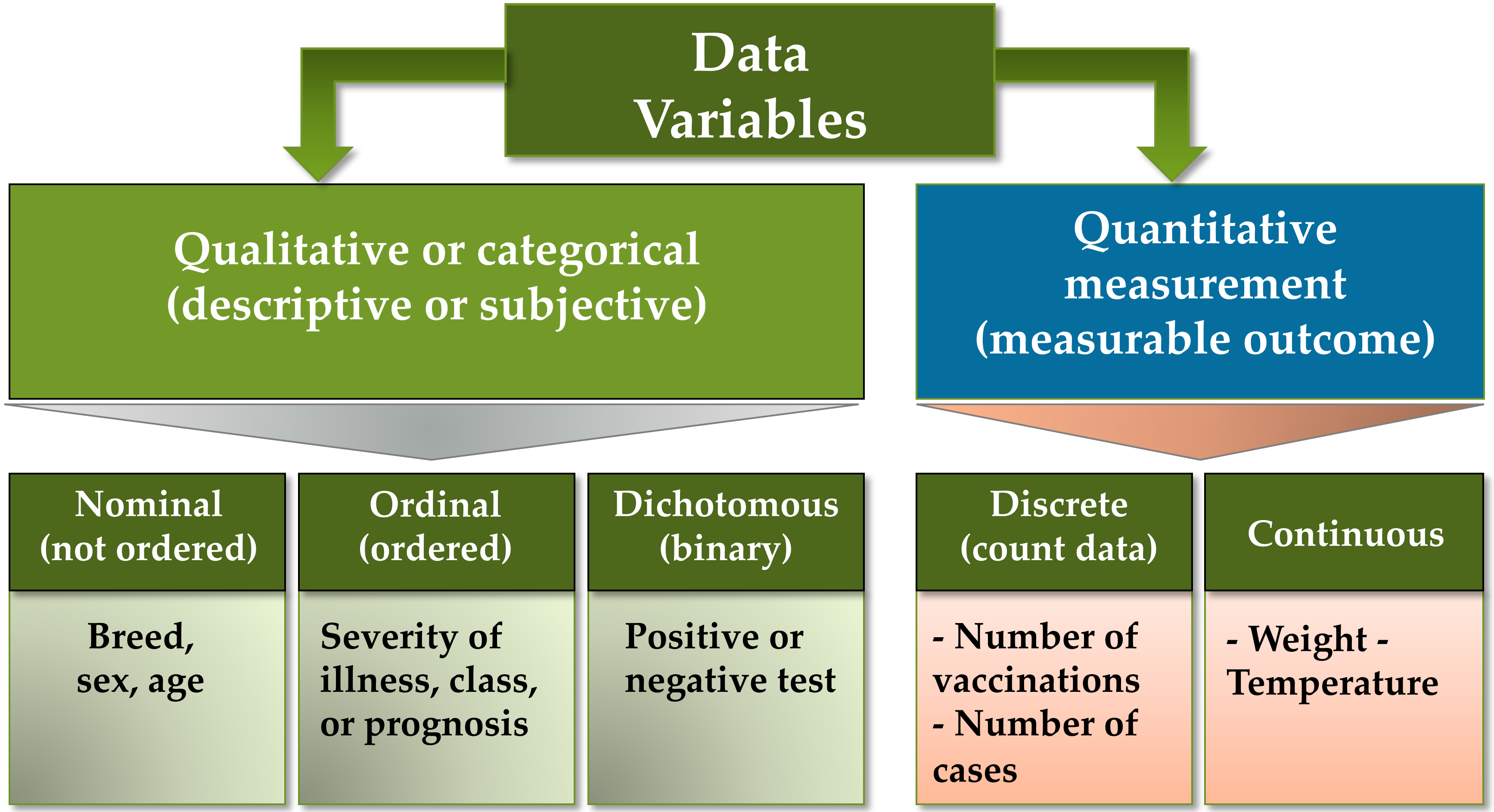
Data Collection in Time

- Prospective (Forward in Time)
  - Using data collected in a forward direction in real time
  - Opportunity to influence data quality
  - Example: Passive laboratory reporting in 2021

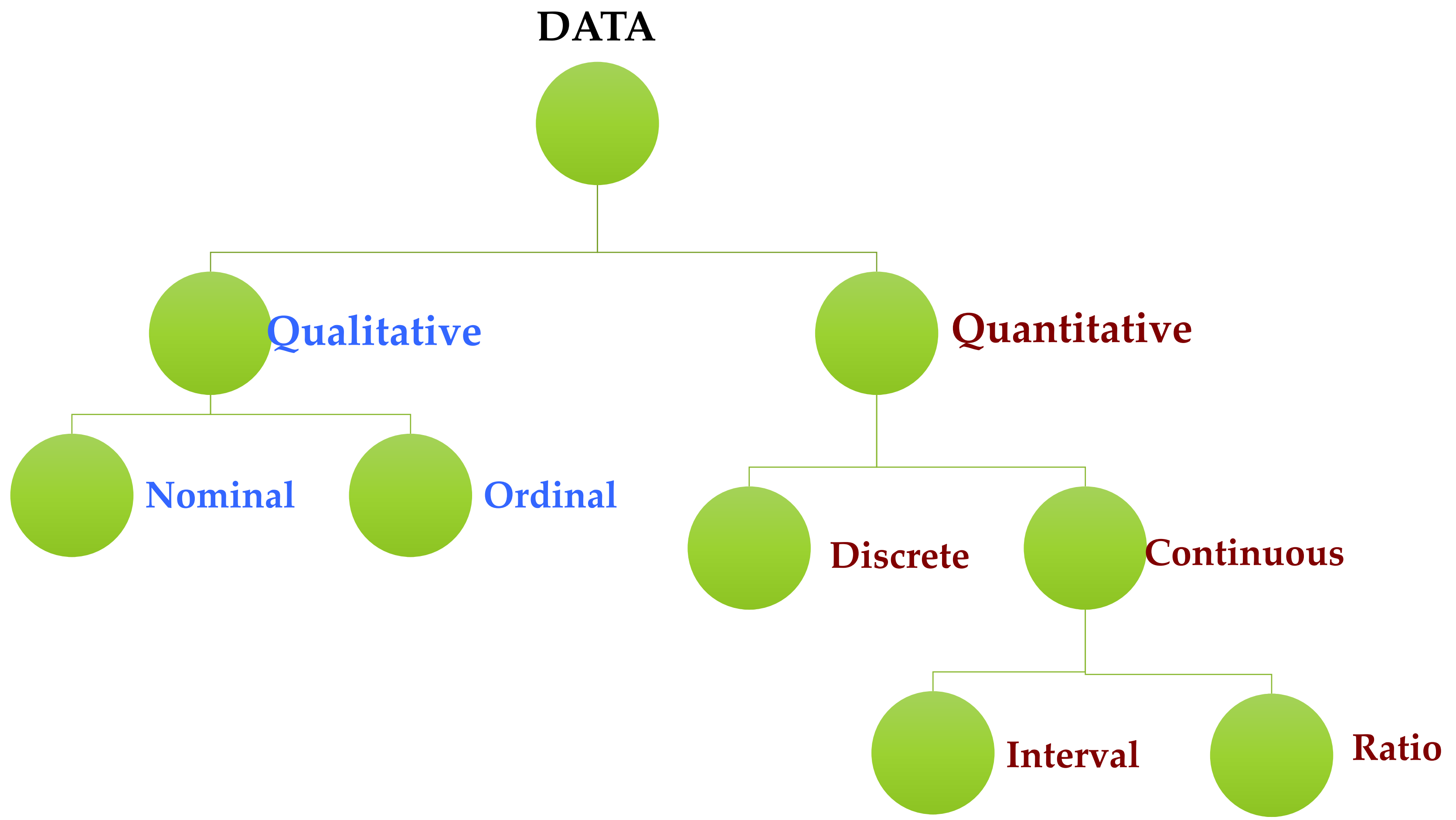
- Retrospective (Back in Time)
  - Using data that already exists in a database
  - Rely on past methods without control of data quality
  - Example: Review all cases of PRRS in Thailand over the past 2 years

Lesson 4: Part 2 of 4

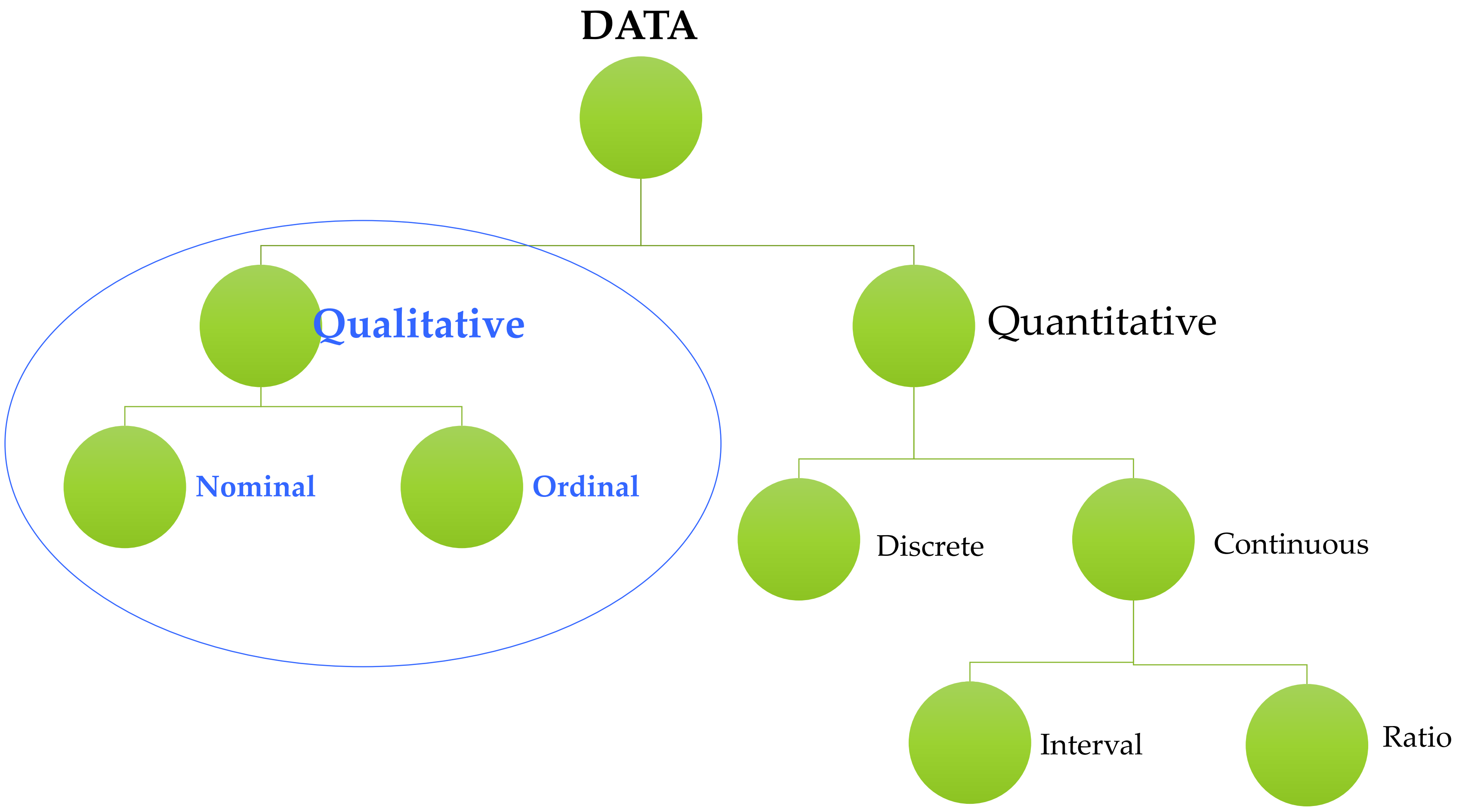
Types of Data Variables



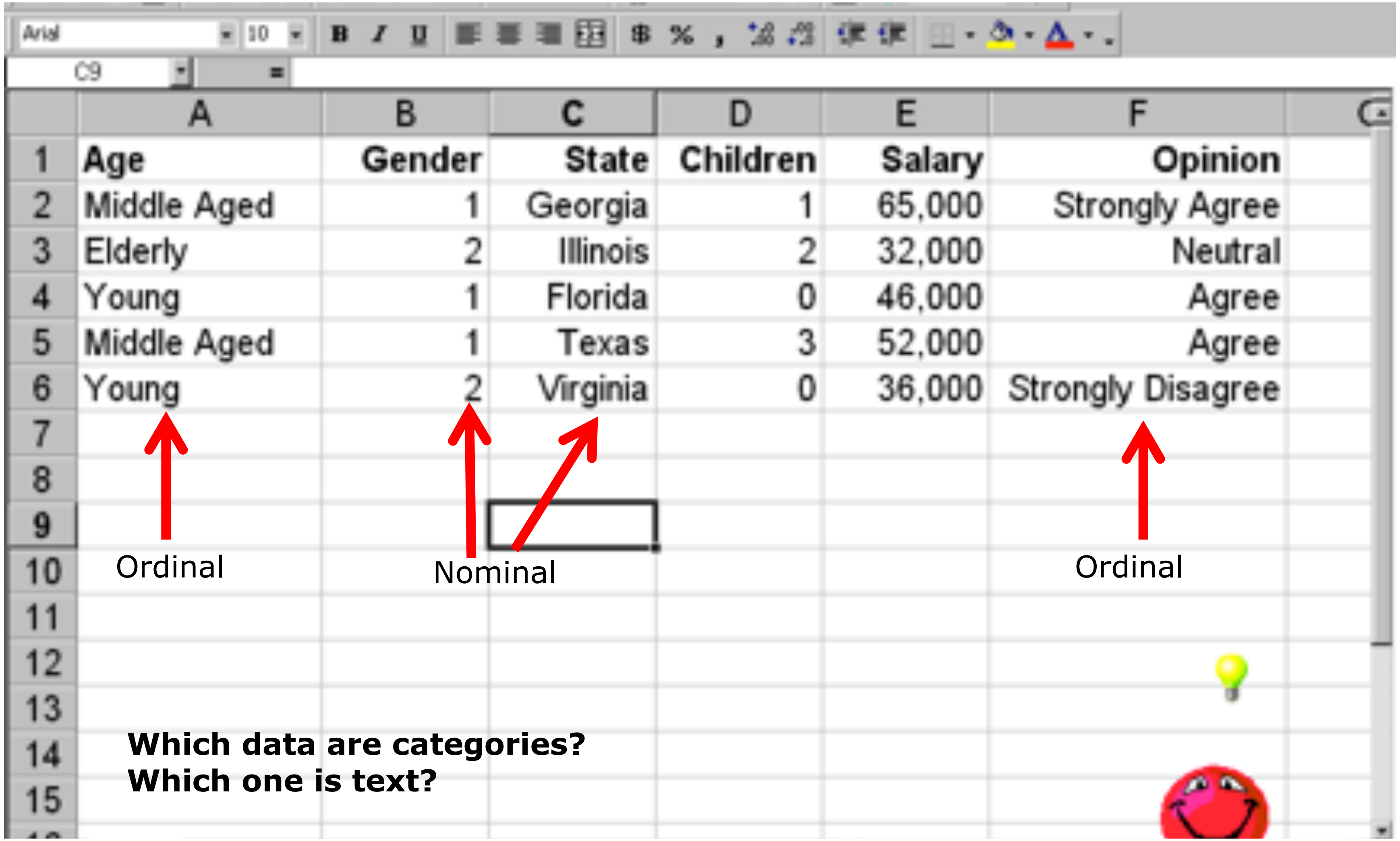
Types of Data Variables: Qualitative and Quantitative



Qualitative Data: Nominal and Ordinal



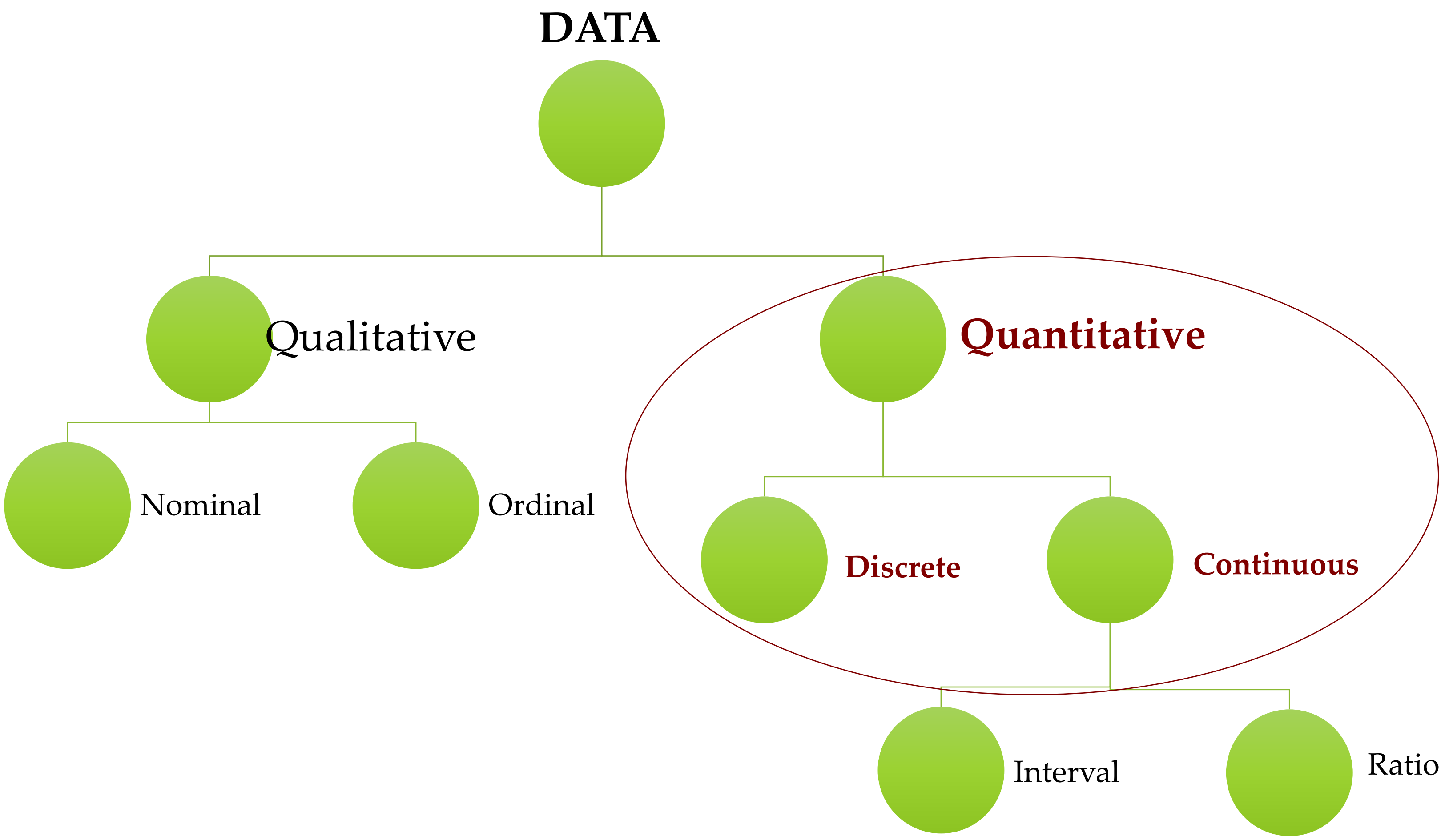
Example: Qualitative Data



	A	B	C	D	E	F
1	Age	Gender	State	Children	Salary	Opinion
2	Middle Aged	1	Georgia	1	65,000	Strongly Agree
3	Elderly	2	Illinois	2	32,000	Neutral
4	Young	1	Florida	0	46,000	Agree
5	Middle Aged	1	Texas	3	52,000	Agree
6	Young	2	Virginia	0	36,000	Strongly Disagree

Which data are categories?
Which one is text?

Quantitative Data: Discrete and Continuous

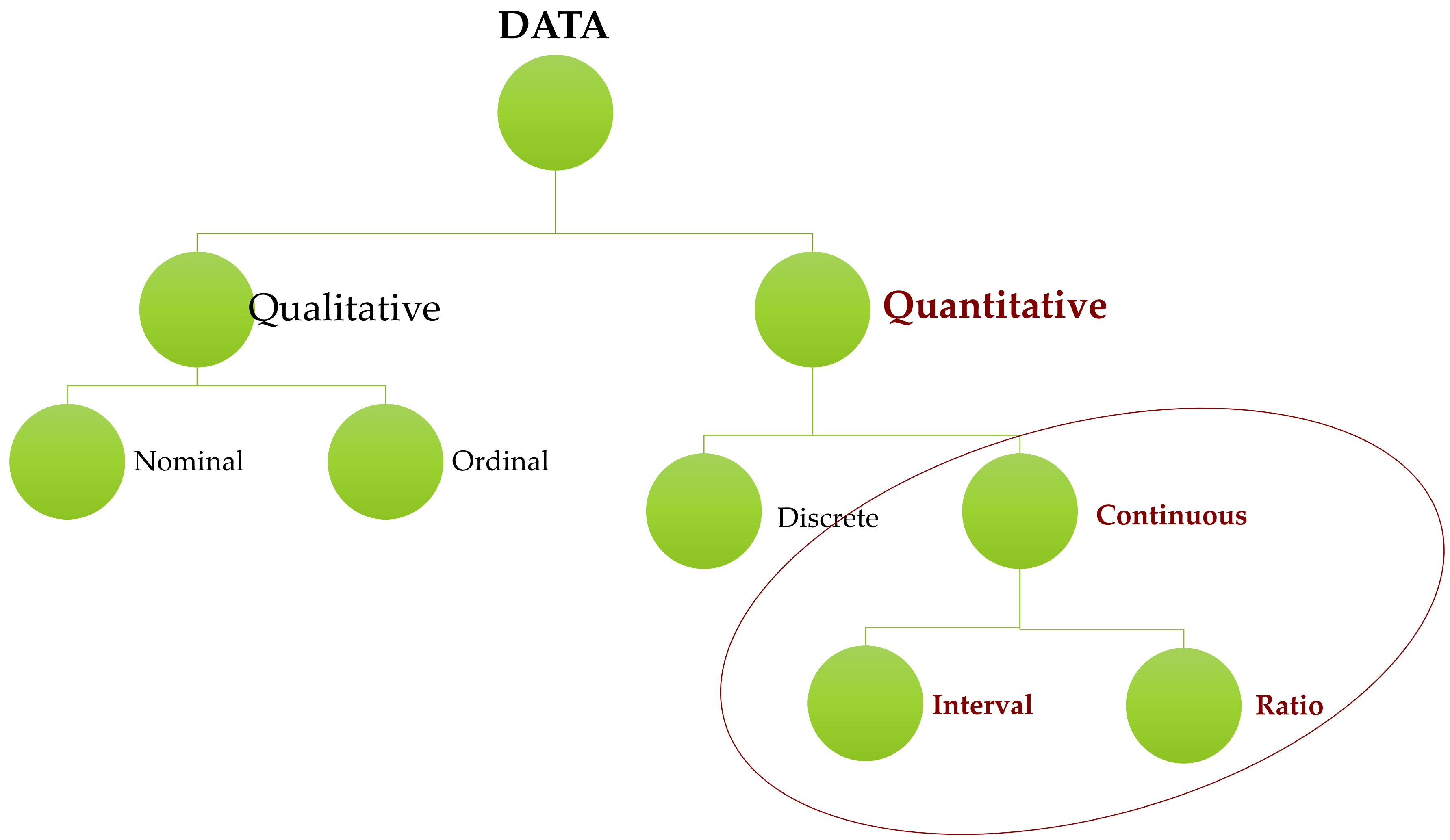


Example: Discrete and Continuous

	A	B	C	D	E	F
1	Age	Gender	State	Children	Salary	Opinion
2	Middle Aged	1	Georgia	1	65,000	Strongly Agree
3	Elderly	2	Illinois	2	32,000	Neutral
4	Young	1	Florida	0	46,000	Agree
5	Middle Aged	1	Texas	3	52,000	Agree
6	Young	2	Virginia	0	36,000	Strongly Disagree

Discrete (Children) **Continuous** (Salary)

Quantitative Data: Interval and Ratio

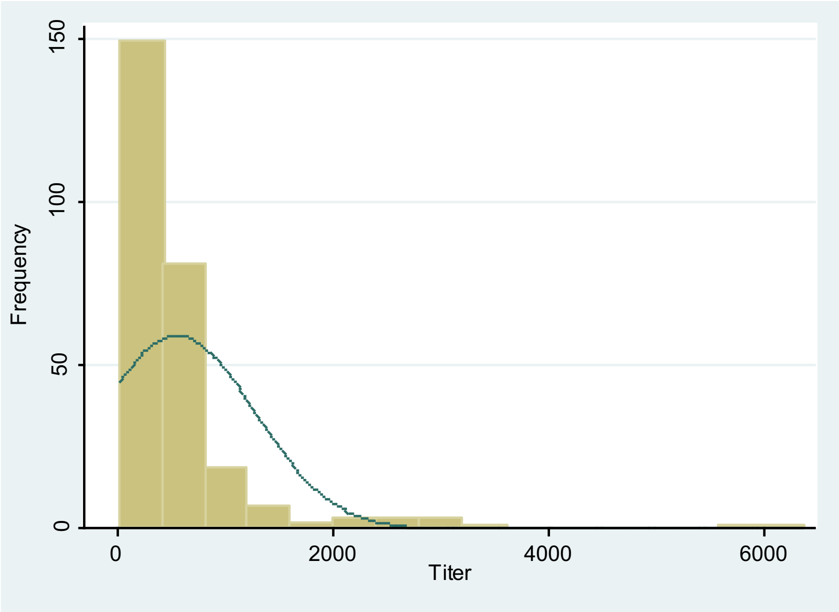


Continuous Data: Population Census

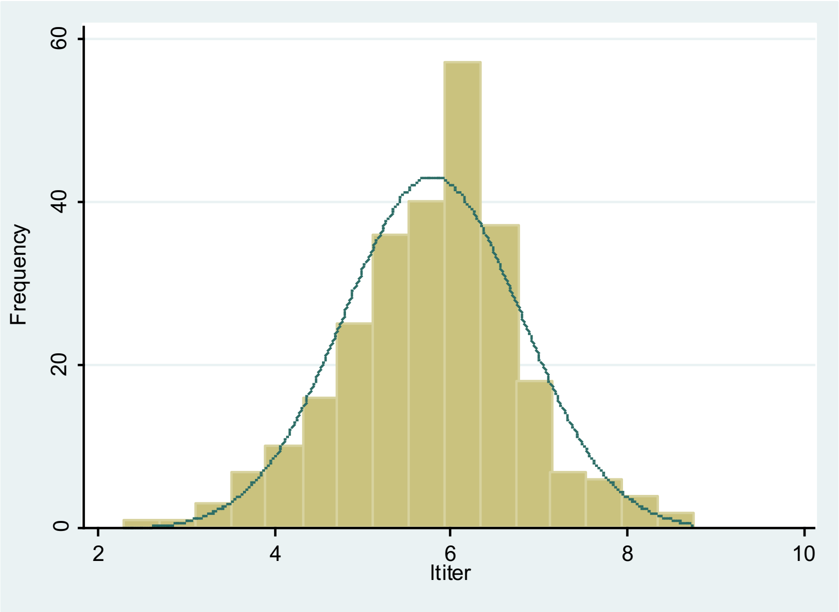
District	Total Cattle	Beef Cattle	Milking Dairy Cows	Sheep	Swine	Broilers	Egg Layers	TOTAL
A	18,000	8,000	500	4,224	4,581		1,556	28361
B	15,000	10,000		6,336	120		133	21,589
C	12,000	1,000	3,300	71	27	150	229	12,477
D	60,000	16,000	17,900	6,722	2,362		764	69,848
E	55,000	20,000	16,200	3,601	1,561		1,552	61,714
F	7,000	4,000		1,607	1,128		6,133	15,868
G	44,000	25,000		4,138	913		459	49,510
H	32,000	9,000	10,200	11,146			358	43,504
I	18,000	10,000		9,418	2,408	510	4,451	34,787
J	67,000	46,000		7,055	143		359	74,557
TOTAL	328,000	149,000	48,100	54,318	13,243	660	15,994	412,215

Distribution of Continuous Data
Examples: Age, weight, population census

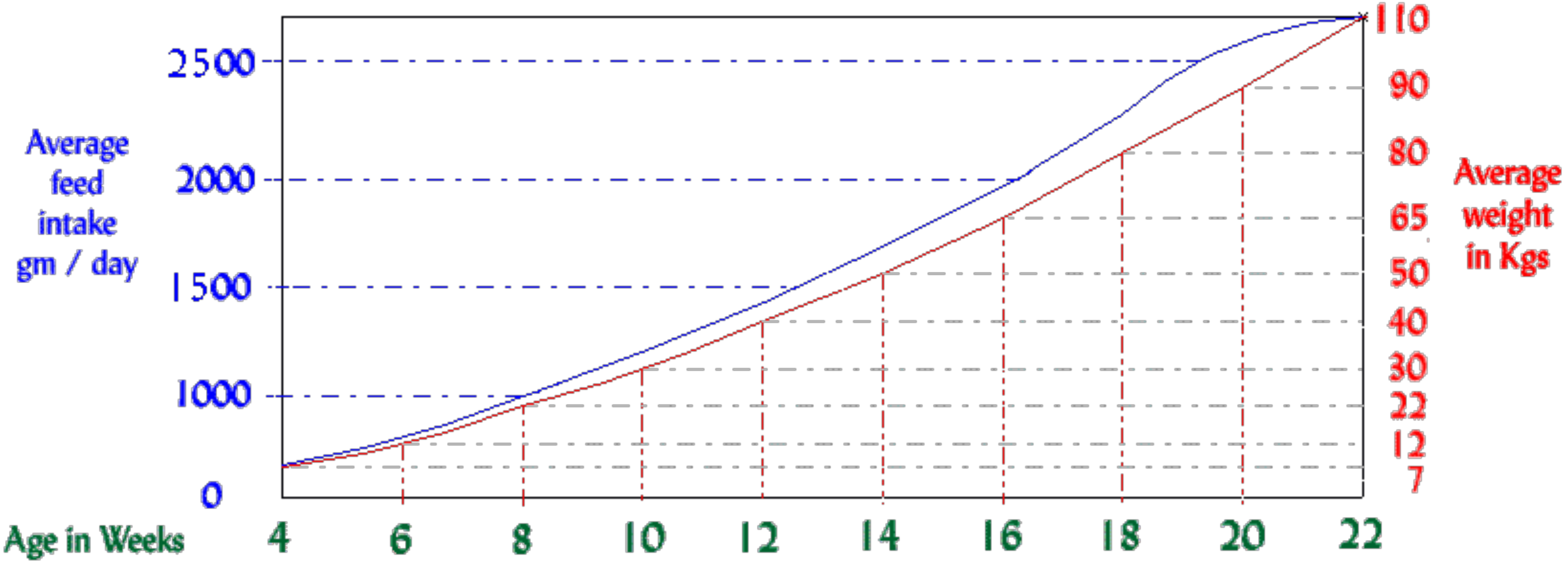
Not normal (skewed)



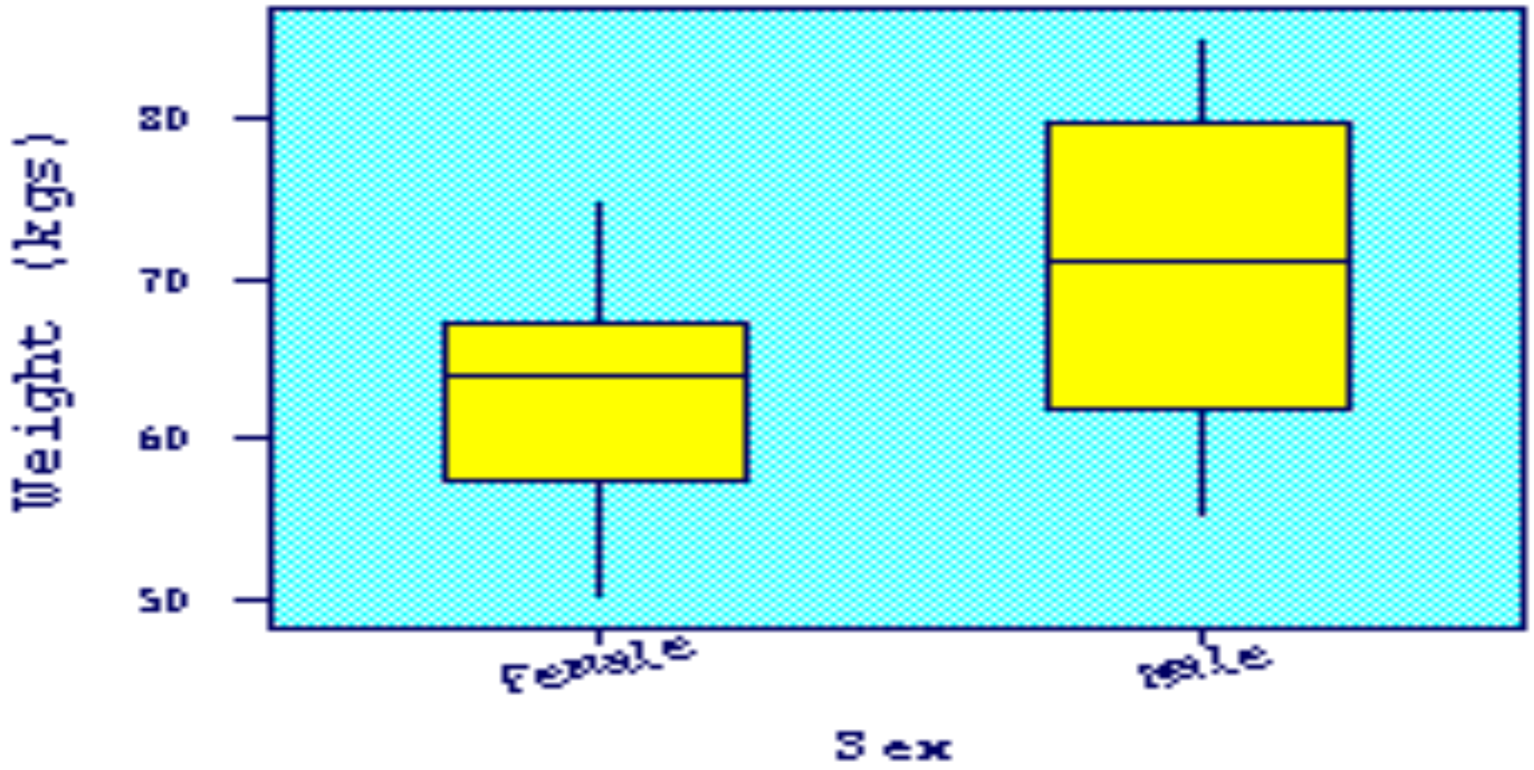
Normal



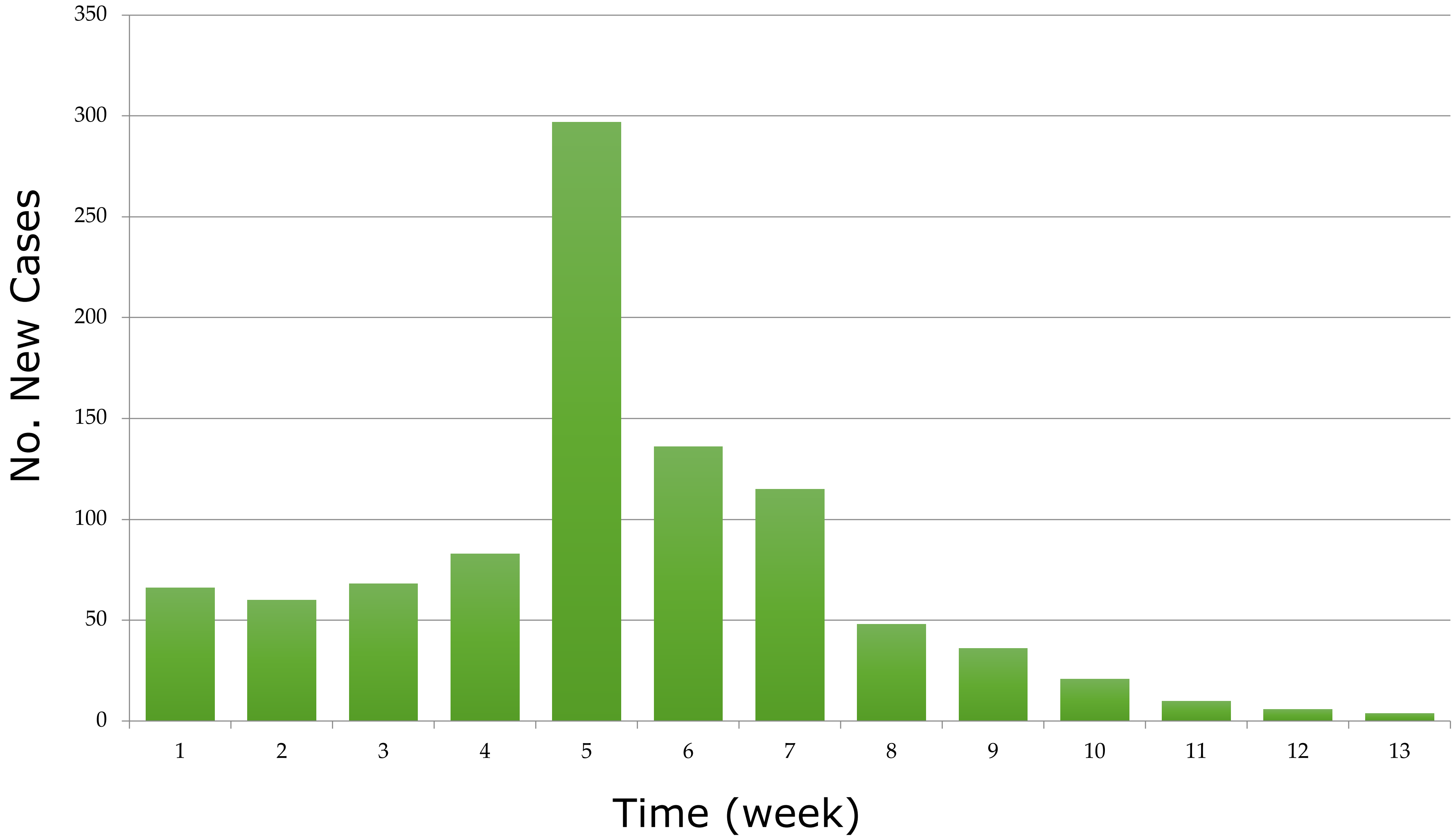
Continuous Variables
Example: Pig Growth Weights



Convert Continuous Data into Interval Data Example: Boxplot Display of Pig Weights as



Interval Data
Example Histogram: No. new positive laboratory cases of HPAI by RT-PCR



Exercise: Describe Rabies Post-Vaccination Reactions in Dogs

QUALIITATIVE DATA:

1. Which variables are qualitative?
   - State, sex, reaction

2. Which variables are nominal
   - State, sex

3. Which variables are ordinal?
   - Reaction

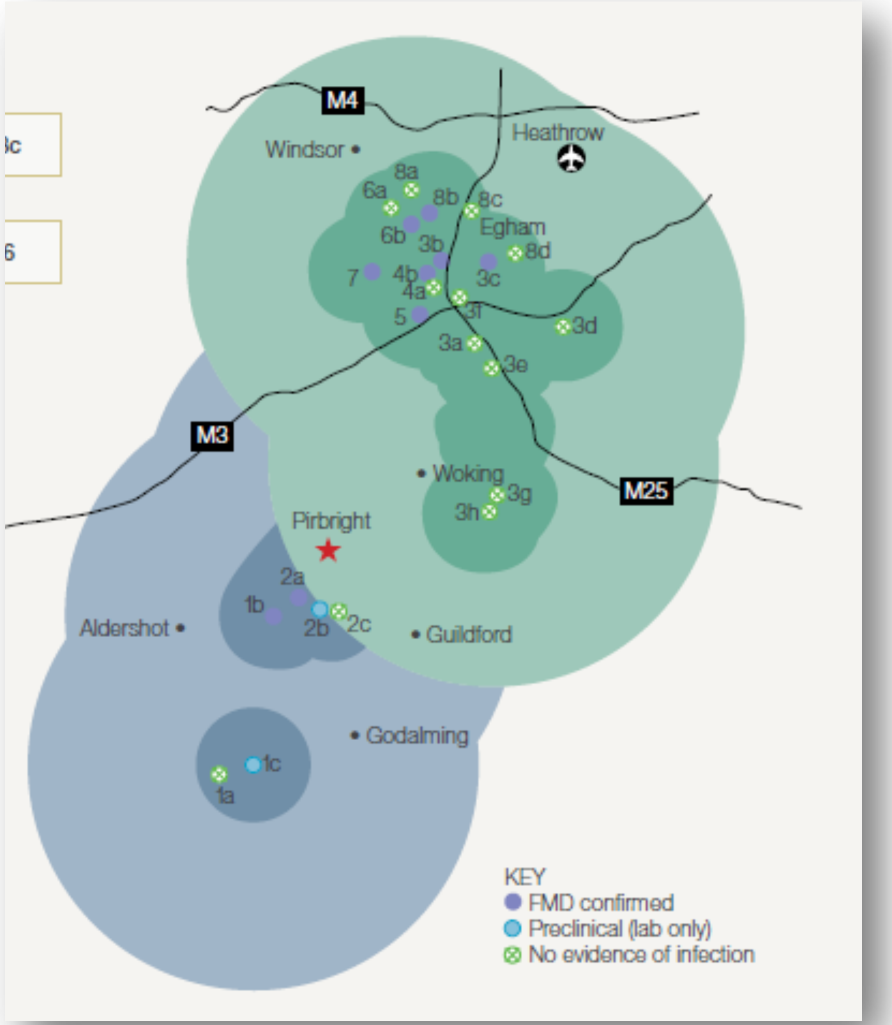
QUANTITATIVE DATA:

1. Which data is discrete?
   - Age, days, dose

2. Which variables are continuous?
   - None

State	Age	Sex	Days	Dose	Reaction
A	2	M	3	1	+
B	3	M	3	1	+
C	6	M	3	1	+
D	2	M	4	1	++
E	4	F	4	1	+
F	7	M	4	2	+++
G	2	F	5	1	+
H	3	M	5	1	++
I	3	F	5	1	+
J	4	F	5	1	++++
K	11	M	5	1	+
L	3	F	7	1	+
L	4	F	14	2	+++
M	2	M	29	1	+
N	5	M	59	1	+

Examples: Continuous Disease Data by Place and Time



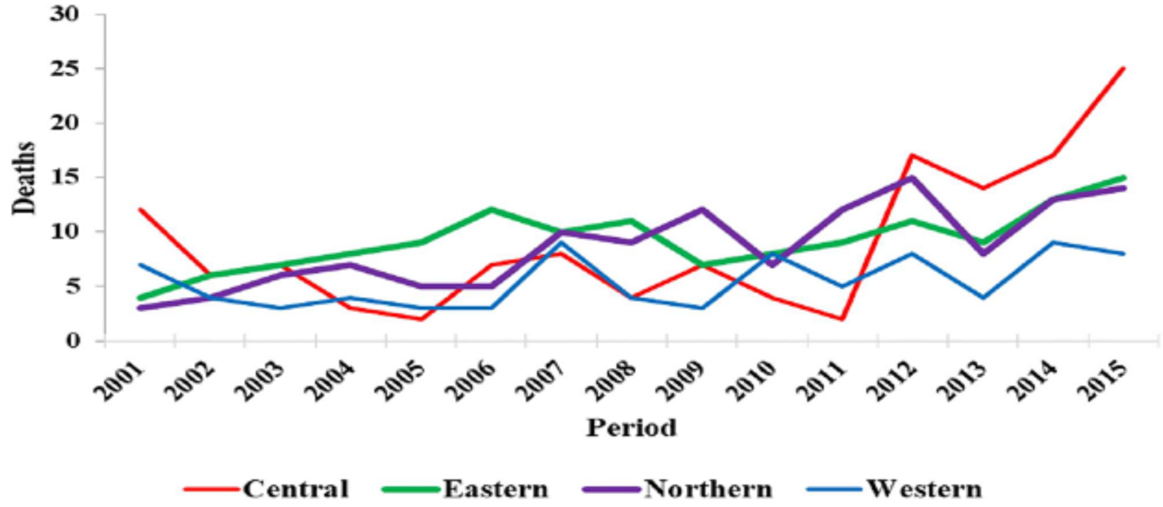


Fig 6. Trends of suspected human rabies deaths by region; 2001–2015.

https://doi.org/10.1371/journal.pone.0198568.g006

Reference:

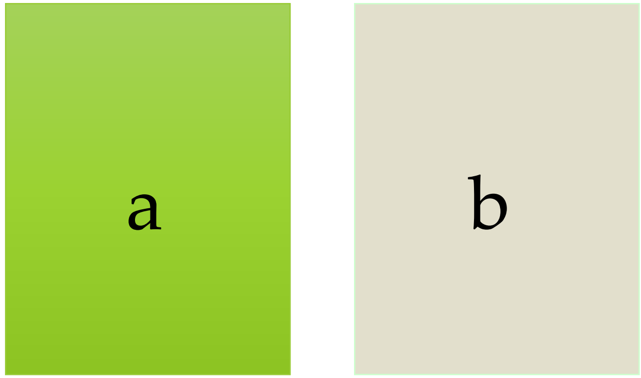
Masiira B, Makumbi I, Matovu JKB, Ario AR, Nabukenya I, Kihembo C, et al. (2018) Long term trends and spatial distribution of animal bite injuries and deaths due to human rabies infection in Uganda, 2001-2015. PLoS ONE 13(8): e0198568. https://doi.org/10.1371/journal.pone.0198568

Lesson 4: Part 3 of 4

Ratios and Proportions

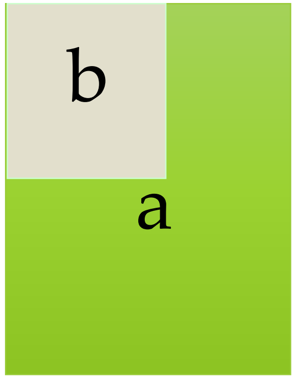
RATIO

- Comparing two separate things e.g. male (a) : female (b)
- Numerator is separate from denominator
- a:b = 1:1

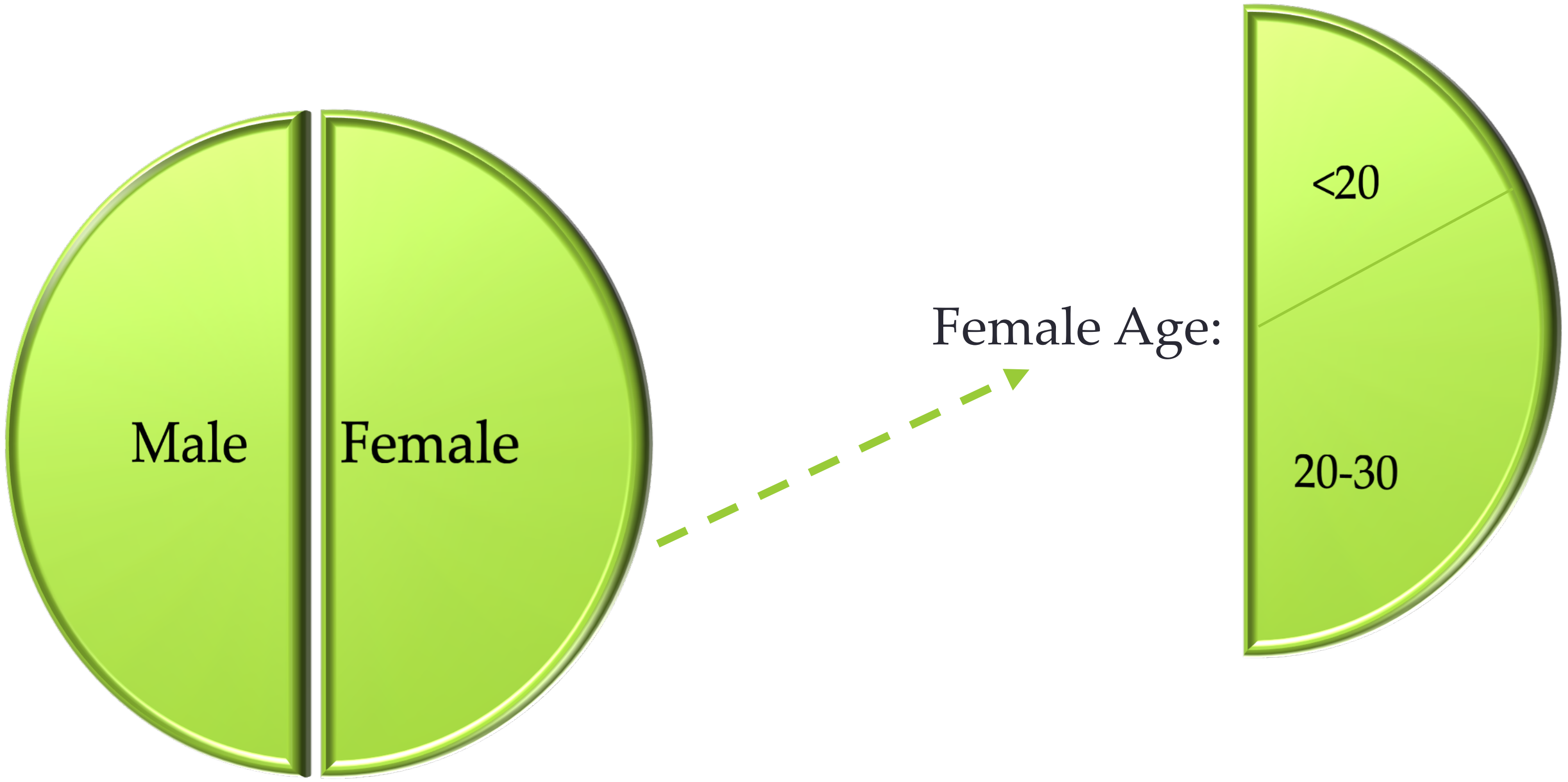


PROPORTION

- Compare how one part exists within something larger
- Numerator is part of denominator
- a / a+b = 3 / 3+1 = ¾



Example of Ratio and Proportion



Ratio M/F: 1:1
Proportion of female = 0.5

Ratio of Age Groups <20: 20-30 = 1:2
Proportion of females < 20 = 0.33
Proportion of females 20-30 = 0.66

Measures of Disease Frequency: Ratio

- <u>A Ratio Compares</u> two counts
- Expressed as a fraction where the numerator is separate from the denominator

Ratio = a / b

<u>Assumption:</u>

1. *The numerator is not included in the denominator*

<u>Application</u>: A field epidemiologist counts 1020 ducks and 310 geese in one village. There are many more ducks than geese present and it can be expressed clearly using numbers in the form of a ratio. The ratio of ducks to geese is as follows:

Ratio (ducks/geese) = 1020 / 310 = 3.3

There are 3.3 many times more ducks as there are geese in this village.

Measures of Disease Frequency: Risk Proportion

- <u>Compares</u> one part to a larger population from which it comes
- The numerator is also included in the denominator
- **<u>Is a measure of "risk" or "probability" of an event occurring</u>**

Using the same village count data a field epidemiologist may want to know what proportion of all waterfowl in a village are geese.

Proportion = a / a+b

Proportion = 310 / 310+1020 = .23 (23%)

Approximately 23% of the waterfowl in the village are geese at this time. Therefore the percentage (proportion) of remaining waterfowl that are ducks is 77% (0.77).

Key Concept: Risk versus Rate

Risk

- **Is a probability expressed as a proportion:** Example 2 of 10 animals are dead = 20%
- Equation: Risk = a / (a+b)
- Units: Percentage affected (%)

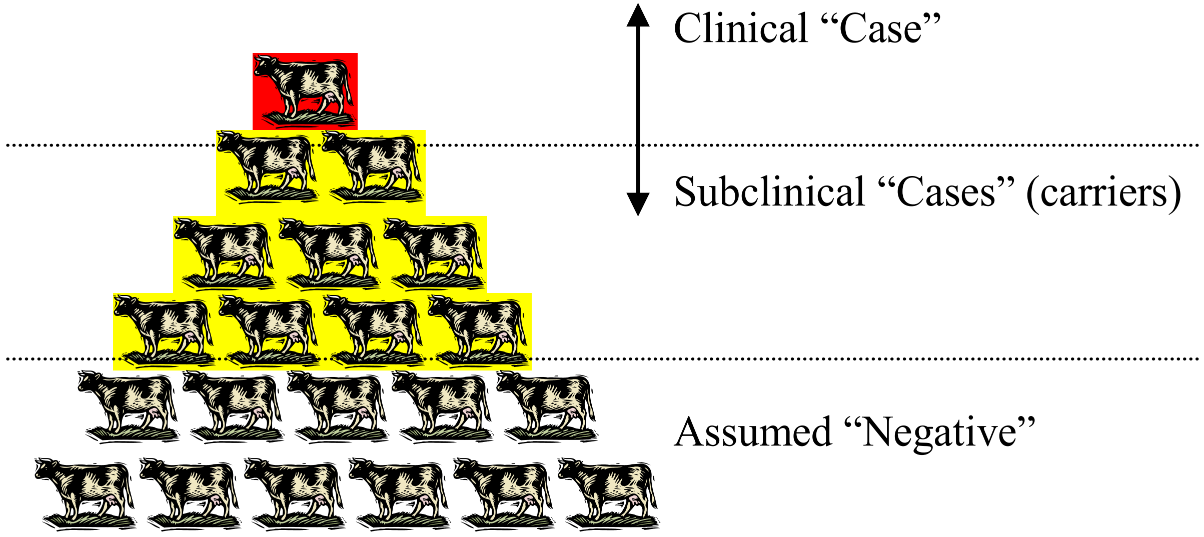
Rate

- **Measures disease risk per unit time**
- Equation: Count / Animal-time at risk
- Units: Animal time at risk e.g. 40 cattle-weeks at risk

Risk Proportion

Example: What is the Risk of TB in the Herd?

(Adapted: C. Gay)



Case Definition	Diagnostic Criteria	Prevalent Cases
Based on clinical signs	Age, body score, herd history	1/21 = 0. 04
Based on subclinical intradermal test	Reaction to tuberculin CC test	9/21 = 0.43
Based on clinical/subclinical	Both criteria above	10/21 = 0.48
Assumed negative	Other than above	11/21 = 0.53

Measuring Disease Risk in Time: Rate

- The number of disease events occurring <u>over a given period of time</u>
- EXAMPLE: A slaughter plant had 250 employees where a Salmonella

outbreak occurred. 100 people were initially positive and 5 of 100 people still carry Salmonella (carriers) 20 weeks following Exposure (Kotova, 1988) Salmonella Disease Rate:

100 cases / 5000 person-weeks at risk = 2 cases per 100 person-weeks at risk

Weeks Following Infection	# Salmonella Positive Fecal
2	92
4	41
9	17
10	12
20	5
TOTAL POSITIVE	167

Measures of Disease Risk: Comparing Proportions

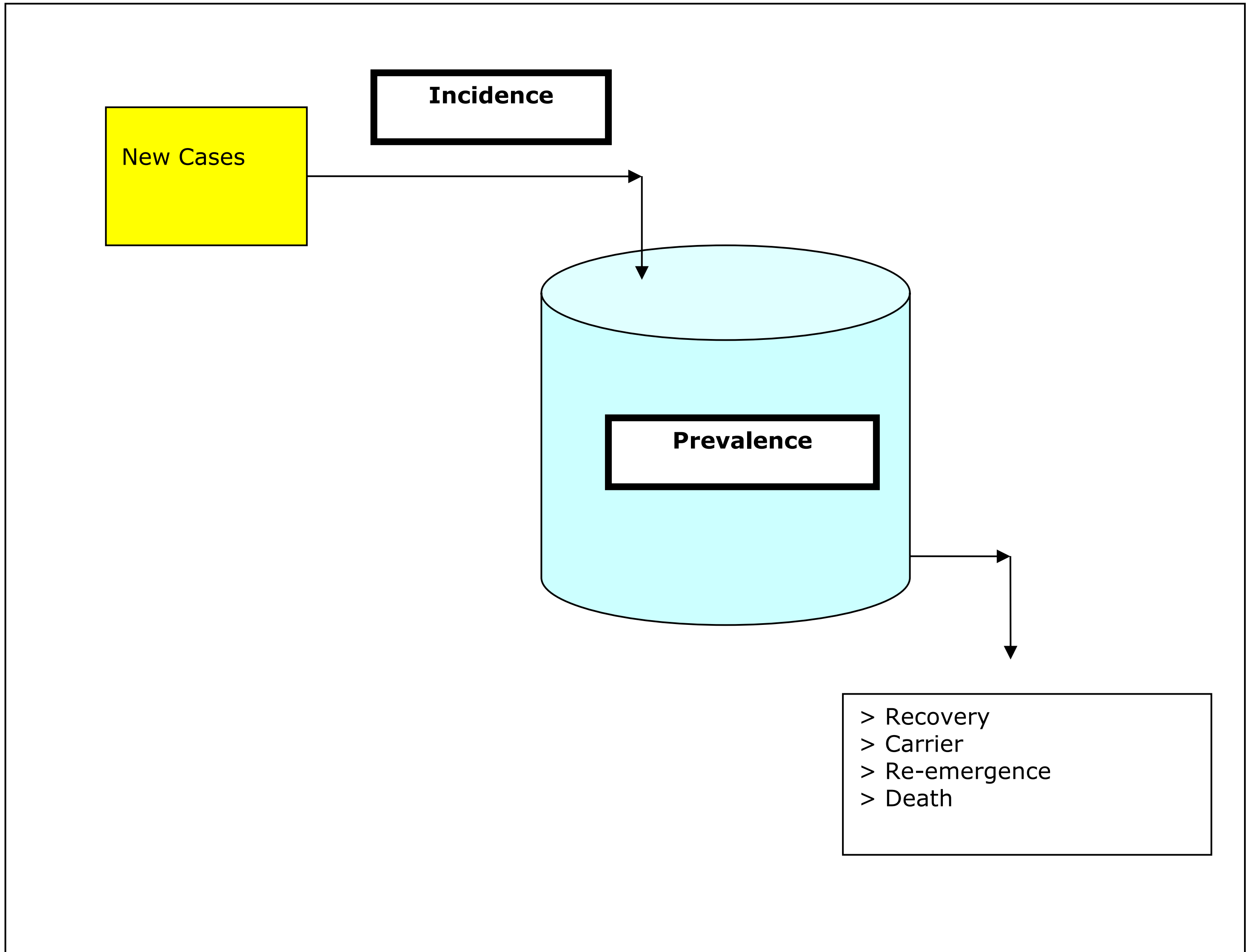
Exposure Factor	**Disease +**	**No Disease -**	Total
Yes - Loose Chickens	10	22	32
No- Confined Chickens	11	37	48
Total	21	59	80

Pr(Loose/D+) = 10/21 = .48 (48% of D+)
Pr(Confined/D+) = 11/21 = <u>.52</u> (52% of D+)
TOTAL = 1.00 (100% of D+ farms)

Pr(Loose/D-) = 22/59 = .37 (37% of D-)
Pr(Confined/D-) = 37/59 = <u>.63</u> (63% of D-)
TOTAL = 1.00 (100% of D- farms)

Disease Incidence and Prevalence

- **Incidence** = Count only NEW cases of disease
- **Prevalence** = Includes both NEW and EXISTING cases of disease



Measuring Disease Prevalence

- Number of existing cases and new cases at some point during a time period

- ***Point prevalence***
  - Counting the existing cases at one brief point in time divided by the PAR at that time

- ***Period prevalence***
  - Counting cases over a longer period of time

Measuring Disease Prevalence

- Prevalence = *P* = *new cases + existing cases / Population at Risk (PAR)*

Sentinel Chicken Example:

Point Prevalence of HPAI on the first day of week 5	=	*1/8*	=	*0.125*	=	*12.5%*
Period Prevalence of HPAI during the 10 week period	=	*4/10*	=	*0.40*	=	*40%*

Incident Disease Rate

- *Review: A rate is a risk (probability) that is calculated over a given time period*
  - *Probability expressed through time*

- Populations change over time (birth, death, migration)

- Cumulative Incident Rate = new cases over specific time period
  - There are two methods to calculate incidence:
    - Approximate
    - Exact

Cumulative Incidence Rate: Approximate Method

- Use when the population at risk (PAR) is changing (open population) over a period of time

$$\textit{(Cumulative) Incident Rate} = IR = \frac{\text{\# Events in a specific time period}}{\text{Average PAR at mid-point}}$$

Assumptions:

1. *All animals are negative for the disease in question at the beginning of the time period;*
2. *All animals that died are due to the disease (although mixed infections do occur);*
3. *The number of animals at the beginning and midpoint are known.*

Application: *There were 40 new cases of rabies diagnosed in cattle in a district over a one year period. The cattle population was estimated to be 1,000 in January at the beginning of the year but many cattle were marketed in May of that year leaving 660 cattle remaining by the end of June.*

$$IR = \frac{\text{40 rabies cases}}{\text{660 cattle}} = .06$$

The cumulative incident rate is .06 with no unit of measure.

Cumulative Incident Rate: Approximate Method

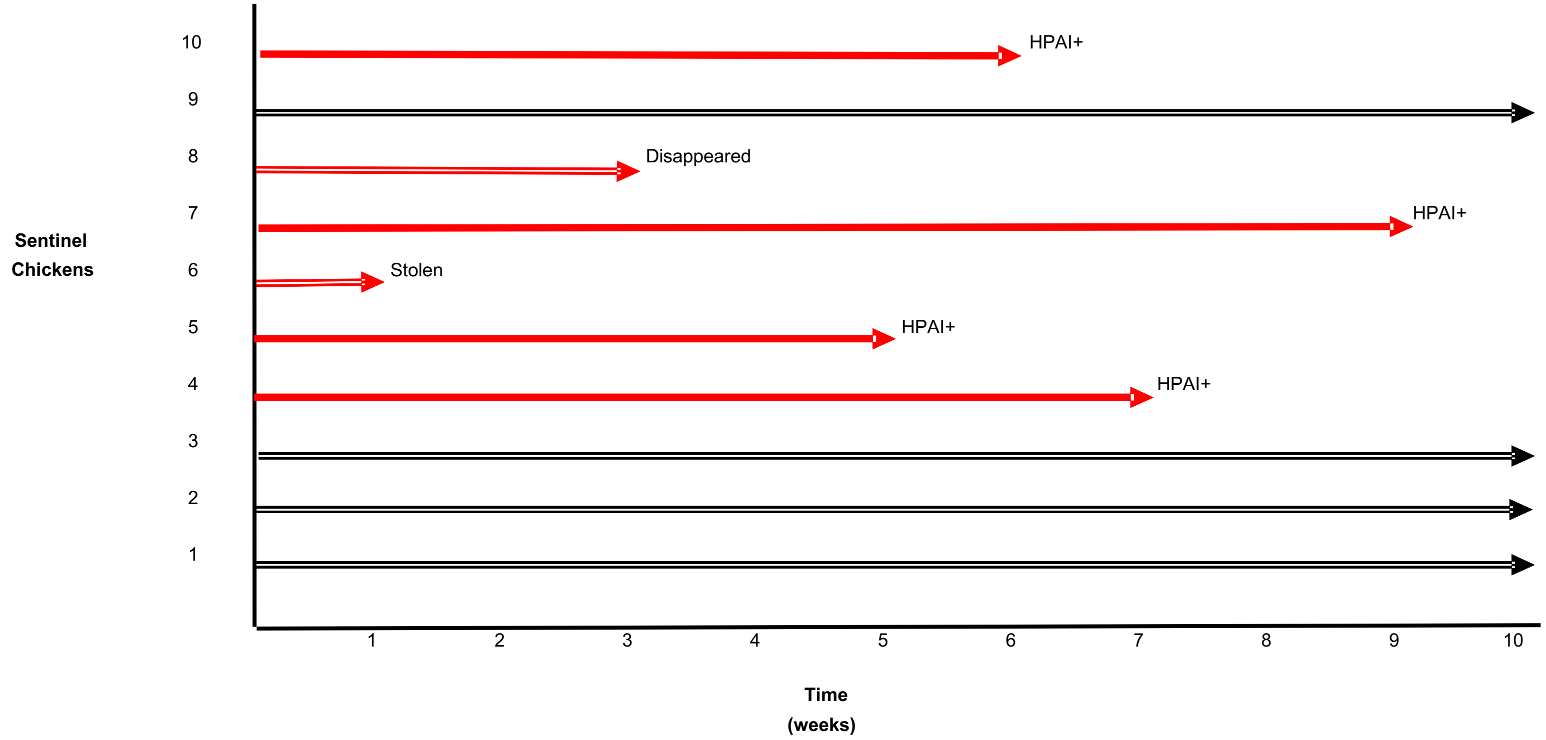
- Difficult to compare two unequal populations
- Therefore we need to "standardize" the rate
- Method:
  - Multiply the incidence rate (IR) by either 100, 1000 or 10,000 or some other number (human health incidence rates are often compared per 100,000 population).
  - Example:

There were 60 cases of rabies in cattle per 1,000 head of cattle in a one year period in this district (IR per 1,000 = .06 X 1000 = 60)

Lesson 4: Part 4 of 4

Exercise 1: Exact Cumulative Incident Rate:

- Counting animal/herd/village "time at risk"
  - Example: Sentinel chicken time at risk for HPAI



Exercise Questions

1. *Calculate total chicken-weeks at risk by constructing a summary table summarizing time at risk for villages that are Negative, Positive and Withdrawn*
2. *What is approximate IR?*
3. *What is the exact IR?*
4. *What is the point prevalence in week 7?*
5. *What is the period prevalence over the entire period?*

1. Calculating total chicken-weeks at risk

Chicken Status	Total	Chicken-Weeks at Risk
Positive (6)	*31*	*Sum the total number of weeks at risk for positive sentinel chickens before they died* *6+3+9+1+5+7 = 6 chickens contribute 31 chicken-weeks*
Withdrawn from Study	0	Calculate the time chicken were at risk until they were lost
Negative (4)	40	Step 1: Calculate the total number of negative sentinels = positive – withdrawals = 10 (Sentinels) - 6 (Positive) – 0 (Withdrawals) = 4 Step 2: Multiply the answer by the total time at risk in weeks = 4 X 10 weeks = 40
TOTAL Time at Risk	**71**	Chicken-Weeks

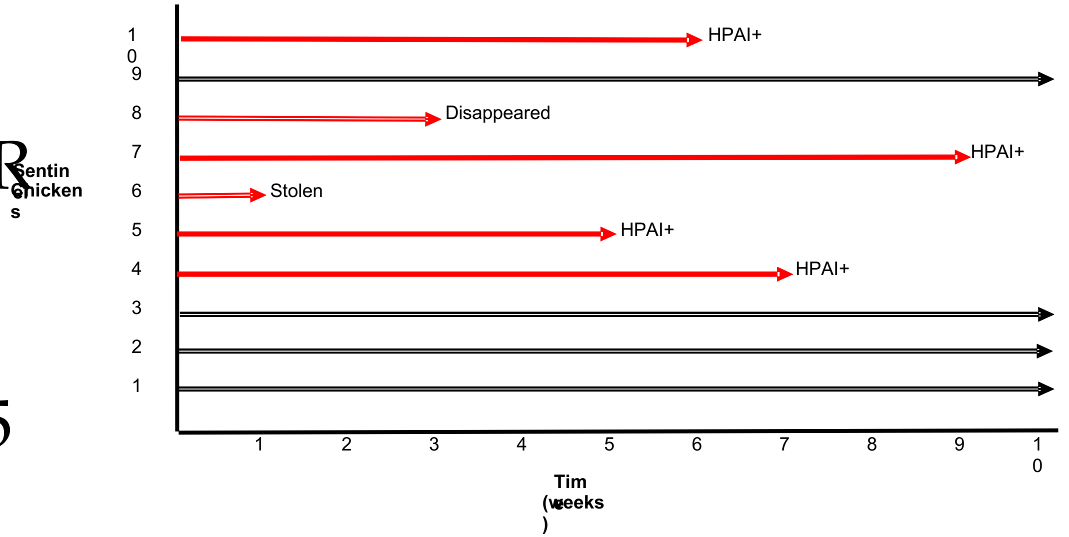
Question 2

Approximate Incidence Rate

= # New Cases Over 10 weeks / PAR at Midpoint of Time (By Week 5)

= 6 / (10 - 3 Pos. – 0 withdrawn) at 5 weeks = 6 / 7 = 0.86

= 86%



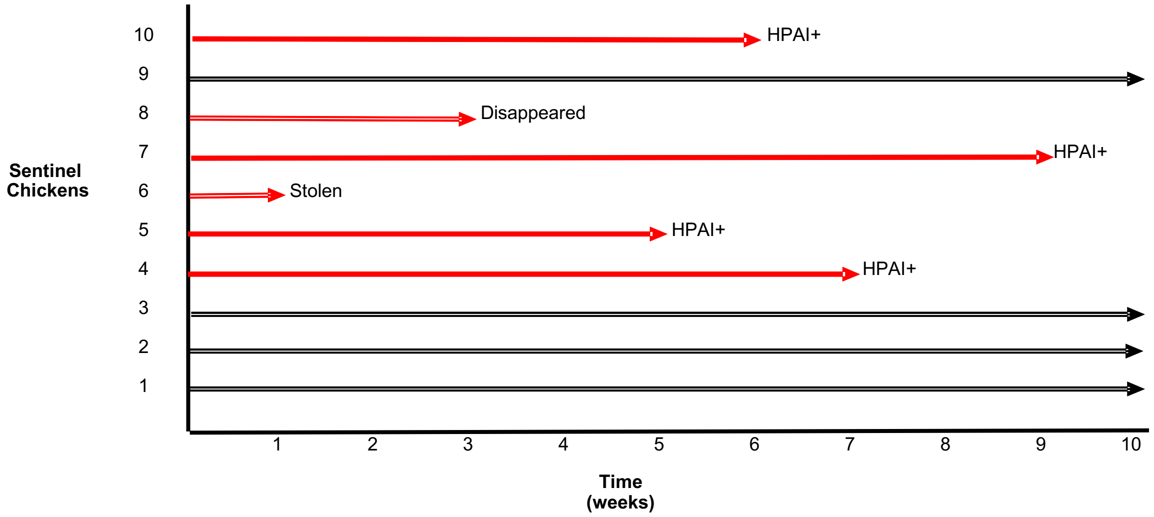
Question 3

Exact Incidence Rate

= # Positive Cases / Total Chicken-Weeks at Risk

= 6 / 71 = 0.08 Cases per Village-Week at Risk

= 8 cases/100 V-W at risk

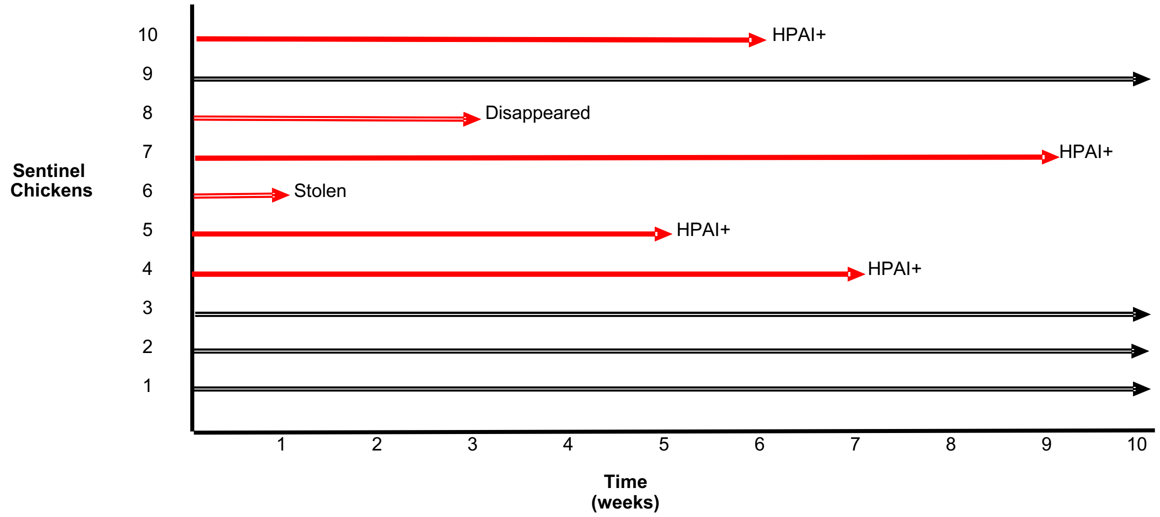


Question 4

Point Prevalence (week 7)

= 1 case (week 7)/ PAR (week 7)
= 1 / 5 = 0.2 = 20 %

(PAR Week 8 = 5 total - 0 withdrawn in previous 8 weeks = 5)

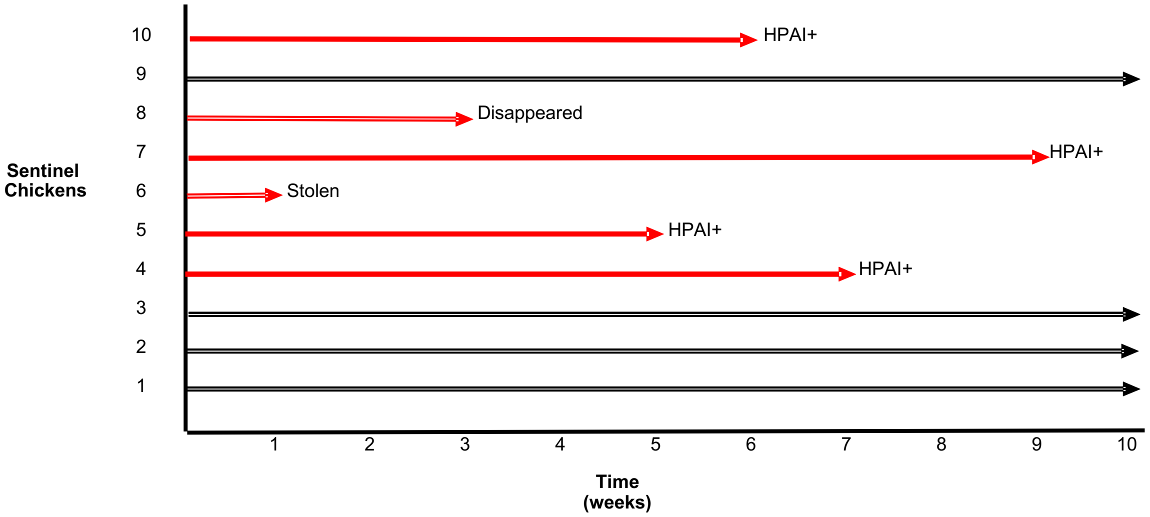


Question 5

Period Prevalence

= # Cases/Initial PAR = 6 / 10 = 0.6

= 60%



Quantitative Relationship Between Incidence and Prevalence

$$P = \frac{I \times D}{I \times D + 1}$$

Where: *P is prevalence*
I is incidence
D is duration of time

Assumptions:

1. *The population is stable*
2. *The incidence of disease remains constant*

Unless these two assumptions can be met, then it is difficult to estimate disease prevalence from incidence data.

Example: *The sub-clinical incidence rate of udder infection in a goat herd was 0.07/goat-year (7 new cases/100 goats). The mean duration of udder infection is 1.5 months (0.125 years) and the population is stable.*

$$P = \frac{0.07 \times 0.125}{0.07 \times 0.125 + 1} = \frac{0.0088}{0.0788} = 0.11 = 11\%$$

At any time, 11% of the goats in this herd can be expected to have sub-clinical udder infection

Crude Disease Rates

1. Crude Morbidity Rate: Describes the number of cases that are clinically affected of the population at risk over some identified time period.

$$\textit{Morbidity Rate} = \frac{\#\ \textit{clinically ill}}{PAR}$$

Example:

$$\textit{Morbidity Rate} = \frac{150\ \textit{ill}}{1100\ \textit{at risk}} = 13.6\%$$

Crude Rates

1. Crude Mortality Rate: Describes the number of deaths in the PAR over some identified time period.

$$\textbf{Mortality Rate} = \frac{\textbf{\# deaths}}{\textbf{PAR}}$$

Example: $$\text{Mortality Rate} = \frac{\text{50 dead}}{\text{1100 at risk}} = 4.5\%$$

Crude Rates

1. Infection Rate: Describes the number of infected individuals in the PAR over some identified time period.

$$\textit{Infection Rate} = \frac{\# \text{ infected}}{\text{PAR}}$$

Example: *Infection Rate* $= \frac{20}{30} = 67\%$

Crude Rates

1. Secondary Attack Rate: Describes how much the disease agent spreads to other animals (secondary cases) over a certain period of time.

$$\textit{Secondary Attack Rate} = \frac{\#\ \text{cases now} - \#\ \text{initial cases}}{PAR}$$

Example: $\textit{Secondary Attack Rate} = \frac{100 - 50}{10{,}000} = 0.005\%$

Crude Rates

1. <u>Case Fatality Rate</u>: Describes the number of deaths among all infected cases over a certain period of time.

$$\textit{Case Fatality Rate} = \frac{\#\ deaths}{\#\ clinically\ ill}$$

Example:

$$\textit{Case Fatality Rate} = \frac{50\ deaths}{150\ clinically\ ill} = 33\%$$

Specific Rates

1. Specific Rates: Describes the number of clinical cases or deaths within a certain part of the population being considered based on **sex, age, breed, production level, etc**.

Example: The crude mortality rate in a flock of pekin ducks was 50 / 1100 = 4.5%. The farmer is sure that more ducklings died than adult ducks. Before the disease occurred, 20% of the population was ducklings and 30 of 50 deaths occurred in ducklings (a duckling is defined as a duck less than 20 weeks of age). The age-specific rate in this case is as follows:

Duckling Age-Specific Mortality Rate $= \dfrac{30}{1100 \times 0.2} =$ *13.6% of ducklings died*

Duck Age-Specific Mortality Rate $= \dfrac{20}{1100 \times 0.8} =$ *2.3% of adult ducks died*

Other Measures of Association

- Relative Risk Ratio

<u>Relative Risk</u>

What is the risk of being a positive case if exposed to a risk factor or not exposed to a risk factor? This is measured by calculating the relative risk:

$$\textbf{Relative Risk} = \textbf{Risk}_{\text{exposed}} \; / \; \textbf{Risk}_{\text{unexposed}}$$

Other Measures of Association

Attributable Risk

How much of the risk of being a positive case is due to exposure to the risk factor? This risk is measured by calculating the attributable risk:

$$\textbf{Attributable Risk} = \textbf{Risk}_{\text{exposed}} - \textbf{Risk}_{\text{unexposed}} \; / \; \textbf{Risk}_{\text{exposed}}$$

Example of RR and AR

Recall Lesson 1, Example 9: vND and Loose Chickens

Exposure	Disease +	No Disease -	Total
Yes- Loose Chickens	10	22	32
No- Confined Chickens	11	37	48
Total	21	59	80

$R_{Exposed}$ = 10/32 = 0.31
$R_{Unexposed}$ = 11/48 = 0.23

Relative Risk = .31 / .23 = 1.3

The relative risk for vND positive cases is 1.3 times higher when loose chickens are observed than when loose chickens are NOT observed.

Attributable Risk = .31 - .23 /.31 = 0.26

The proportion of vND positive cases attributed or associated with observing loose chickens is 0.26.

Assignment: Exercise 2 Incidence and Prevalence

- An assignment will be given to practice incidence and prevalence calculation.

Lesson Summary

1. Descriptive epidemiology is based on Animal-Place-Time data

2. Data variables can be qualitative (categorical) versus quantitative (continuous or discrete)

3. We describe disease impact using the following health and disease indices:
   1) Ratios and proportions
   2) Incidence and prevalence
   3) Incidence rates and prevalence proportions
   4) Crude rate; specific rate; attack rate; morbidity rate; mortality rate; infection rate; secondary attack rate; case fatality rate (CFR); relative risk (RR); and attributable risk (AR).

References

Altman, N. and M. Krzywinski. 2015. Association, correlation and causation. NATURE METHODS. VOL.12 NO.10

Dohoo, I. R., Martin, W., and H. E. Stryhn. 2003). Veterinary epidemiologic research. Charlottetown, P.E.I.: University of Prince Edward Island.

Martin S.W., Meek A.H. and P. Willeberg. 1987. Veterinary Epidemiology. Ames Iowa. Iowa State Press.

Thrusfield M. 2007. Veterinary Epidemiology, (3rd Ed.). Ames, Iowa. Blackwell Science Ltd.