

# QEPI NOTES

SOLYMOSI NORBERT  
2020.11.03.



# *Contents*

## *Lecture 1*      5

<i>Goal of the course</i>	5
<i>Epidemiology</i>	6
<i>Causation</i>	8
<i>Reasoning</i>	11
<i>Models</i>	12

## *Lecture 2*      13

<i>Measures of health</i>	13
<i>Incidence &amp; prevalence</i>	13
<i>Survival analysis</i>	16
<i>Adjusting measures</i>	17

## *Lecture 3*      19

<i>Diagnostic Tests</i>	19
<i>Test evaluation</i>	20
<i>Predictive values</i>	22

## *Lecture 4*      25

<i>Multiple diagnostic tests interpretation</i>	25
<i>Prevalence adjustment</i>	26
<i>Herd level interpretation of diagnostic test results</i>	28
<i>Sample size for disease detection</i>	29

*Lecture 5*      31*Analytic epidemiology*      31*Association of disease and exposure*      32*Lecture 6*      39*Association with more than 2 levels of categories*      39*Relationship between continuous & continuous variables*      40*Logistic regression*      40*Associations in survival studies*      41*Evidence Based Medicine*      42*Lecture 7*      49*Spatial epidemiology*      49*Disease mapping*      49*Spatial pattern analysis*      51*Ecological analysis*      54*Bibliography*      57

# Lecture 1

## *Goal of the course*

Lifelong learning is a fundamental prerequisite for maintaining professional standards. An important element of that to follow the veterinary literature continuously. Since in this area – just like in other disciplines – a remarkable part of publications based on studies of populations, and accordingly the results are presented by quantitative features, the readers must be familiar with the professional interpretation of them to understand the details of the publications.<sup>1</sup> The goal of the course to familiarize the students with the meaning and the professional interpretation of the most often used measures of quantitative veterinary epidemiology literature.

<sup>1</sup> A 21st century clinician who cannot critically read a study is as unprepared as one who cannot take a blood pressure or examine the cardiovascular system (Glasziou et al., 2008).



On slide 3-7 recently published Covid-19 examples represent some items of main topics will discussed during the course. As one may realize the cited abstracts contain quantitative information and terms should be understood by the readers. On slide 3, the summarized findings of Pollán et al. (2020) uses the term sensitivity, specificity of the applied tests and prevalence (with confidence interval) of the sero conversion. Slide 4 results use the odds ratio as a measure of risk comparison on different conditions (Mehra et al., 2020). Systematic review

based meta-analysis is a crucial element of evidence based (veterinary) medicine (EB(V)M), an example of this is shown on slide 5 (Bastos et al., 2020). In more and more papers may meet the reader with survival analysis results, just like on slide 6 (Del Valle et al., 2020). The EBM/EBVM is a widely accepted approach to improve the daily clinical practice, but as the showed paper (Greenhalgh, 2020) cited on slide 7 suggests that in the case of new diseases hardly applicable. Slide 8 shows an example of genomic epidemiology representing an emerging piece of quantitative epidemiology.

## *Epidemiology*

Due to the fact in epidemiology<sup>2</sup> the *disease in population* is the most focused moment, it is necessary the utilization methods of mathematics, statistics and computing science.

<sup>2</sup> Slide 9-10: the definition, the objectives and the detailed areas of epidemiology

### *Definition*

*Epidemiology* is the study of disease in populations and of factors that determine its occurrence; the key word being *populations* (Thrusfield et al., 2018).

Epidemiology is concerned with the prevention and control of disease in human and animal *populations*. Veterinary epidemiology additionally includes the investigation and assessment of other health-related events, notably *productivity* (Noordhuizen et al., 2001).

### *Objectives of epidemiology*

- determination of the origin of a disease whose cause is known
- investigation and control of a disease whose cause is either unknown or poorly understood
- acquisition of information on the ecology and natural history of a disease
- planning, monitoring and assessment of disease control programmes
- assessment of the economic effects of a disease, and analysis of the costs and economic benefits of alternative control programmes

### *Main areas of epidemiology*

- **Causality models**
- **Measures of health**
- **Diagnostic tests**
- **Measures of associations**
- **Evidence Based Veterinary Medicine (EBVM)**

The bold items will be discussed during the semester.

- Outbreak, spreading models
- Sample size estimations
- Monitoring & surveillance systems (MOSS)
- Spatial epidemiology

### *History of epidemiology*

Despite previous attempts, the founder of the discipline is considered to be the English surgeon-anesthesiologist, *John Snow*. In his article published in 1855, Snow presented an analysis of the 1854 cholera epidemic in Soho, London.<sup>3</sup> In this work, he conducted an analysis in which he plotted the number of deaths caused by cholera in each house.



The figure shows the reconstruction of the map of Snow with cumulative mortality data per building for the 1854 cholera epidemic in Soho, London. Each line represents a case of deaths. The heterogeneity in the cumulative mortality pattern shown on the map turned his attention to the public pump on Broad Street. As a result, he examined where the residents carry the water used in the household. (The houses did not have piped drinking water at that time.) He found that in the houses with a higher incidence, the water was carried from this public pump. The map shows that a significant number of deaths occurred in distant buildings as well, most of which showed a connection to the indicated public well. After the closure of the wells, the epidemic, which was already on a declining trend, ceased. Inter-

Slide 12

<sup>3</sup> <http://kora.matrix.msu.edu/files/21/120/15-78-52-22-1855-MCC2.pdf>

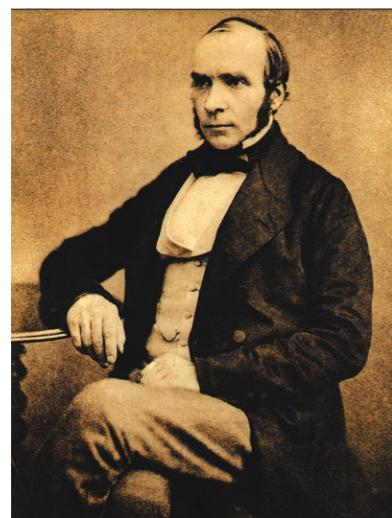


Figure 1: John Snow (1813-1858)

estingly, there was a brewery and a Dominican convent without any cases. One may speculate that the reason for this, there the workers and monks did not drink water at all; instead they only drank beer, and the infection did not affect them.

Snow's study and results at the time provoked many contradictions, e.g. the editor-in-chief of the Lancet medical paper strongly criticized the role of drinking water in the development of cholera. Even in this age, miasma and humoral theories dominated the determination of the pathology of diseases. Although *Jacob Henle* had already suggested that microorganisms cause disease as early as the 1840s, his disciple *Robert Koch* did not formulate the Koch postulates until 1884, which provided basic guidance for many decades in exploring the causal relationship between microorganisms and disease.

### *Causation*

Epidemiology studies the occurrence, associations of disease and factors, causes. The study of causality try to establish the connection of cause and consequence, in natural science or philosophy. In natural science the main goal to identify the causes, explaining certain natural phenomena.

Slide: 13-23

association or cause?

„*Fortis imaginatio generat casum*“  
Michel de Montaigne (1533–1592)

In epidemiology, the study of causality means identifying the triggers of a health-related event. In many cases, we cannot establish an exact causal chain, only coexistences of a hypothetical cause and effects. Nevertheless, they can also be a useful result, just like in Snow's example. In medical history, different models were developed for clarification the connection between the health-related event and its possible causes.

One of the first disease causality model was the postulates of *Koch* (Table 1). In the case of "simple cause" infectious disease those are sufficient to explain the causality. In the case of multifactorial diseases the Koch's model is not applicable.

An organism is causal if

- 1 it is present in all cases of the disease
- 2 it does not occur in another disease as a fortuitous and non-pathogenic parasite
- 3 it is isolated in pure culture from an animal, is repeatedly passed, and induces the same disease in other animals

Table 1: *Koch's postulates (1884)*

However, since the nineteenth century, the focus of human and animal medicine has been shifted from the single cause diseases to multifactorial ones. Table 2. shows the changes of the top ten death causes in England since 1860 to 1970. Although these are human health issues, the tendency is very similar in the animal health.

Accordingly, models have been developed to help elucidate the causes of multifactorial diseases, like the model of Hill and Evans, in Table 3 and 4, respectively. A common feature of both is the probabilistic approach opposing Koch's model, which is dichotomous in nature.

Year	Rank	Disease	Proportion (%)
1860	1	Tuberculosis	19.8
	2	Diarrhoea, enteritis	15.0
	3	Cholera	6.4
	4	Pneumonia/influenza/bronchitis	6.1
	5	Infantile convulsions	5.9
	6	Diphtheria, croup	2.7
		Disentery	2.7
		Stroke	2.7
	9	Scarlet fever	2.5
	10	Nephritis	2.4
1900	1	Pneumonia/influenza/bronchitis	14.4
	2	Tuberculosis	11.3
	3	Diarrhoea, enteritis	8.1
	4	Heart diseases	8.0
	5	Nephritis	4.7
	6	Accidents	4.5
	7	Stroke	4.2
		Diseases of early infancy	4.2
	9	Cancer	3.7
	10	Diphtheria	2.3
1970	1	Heart diseases	38.3
	2	Cancer	17.2
	3	Stroke	10.8
	4	Pneumonia/influenza/bronchitis	3.6
	5	Accidents, suicide	3.1
	6	Motor vehicle accidents	2.8
	7	Diseases of early infancy	2.3
	8	Diabetes	2.0
	9	Arteriosclerosis	1.7
	10	Cirrhosis	1.6

Table 2: The top ten causes of death in 1860, 1900, 1970 (Noordhuizen et al., 2001)

- 
1. Strength of association
    - strong associations are more likely to be causal
    - cannot infer that weak association is not causal
  2. Consistency
    - has the cause and effect relationship identified by a number of different researchers?
  3. Specificity
    - a single exposure should cause a single disease
    - when present, specificity does provide causality, but its absence does not preclude causation
  4. Temporality
    - cause must precede effect
  5. Dose-response relationship
    - as the level of exposure is increased, the rate of disease also increases
  6. Plausibility and coherence
    - biological mechanism
  7. Experimental evidence
  8. Analogy (BSE/scrapie)
- 

Table 3: *Hill's criteria for causation (1965)*

- 
- 1 the proportion of individuals with disease should be significantly higher in those exposed to the supposed cause than in those who are not
  - 2 exposure to the supposed cause should be present more commonly in those with than those without the disease, when all other risk factors are held constant
  - 3 the number of new cases of disease should be significantly higher in those exposed to the supposed cause than in those not so exposed, as shown in prospective studies
  - 4 temporally, the disease should follow exposure to the supposed cause with a distribution of incubation periods on a bell-shaped curve
  - 5 a spectrum of host responses, from mild to severe, should follow exposure to the supposed cause along a logical biological gradient
  - 6 a measurable host response (e.g. antibody) should appear regularly following exposure to the supposed cause in those lacking this response before exposure, or should increase in magnitude if present before exposure, this pattern should not occur in individuals not so exposed
  - 7 experimental reproduction of disease should occur with greater frequency in animals or man appropriately exposed to the supposed cause than in those not so exposed; this exposure may be deliberate in volunteers, experimentally induced in the laboratory, or demonstrated in a controlled regulation of natural exposure
  - 8 elimination (e.g. removal of a specific infectious agent) or modification (e.g. alteration of a deficient diet) of supposed cause should decrease the frequency of occurrence of the disease
  - 9 prevention or modification of the host's response (e.g. by immunization) should decrease or eliminate the disease that normally occurs on exposure to the supposed cause
  - 10 all relationships and associations should be biologically and epidemiologically credible
- 

Table 4: *Evans's postulates (1976)*

## *Reasoning*

Scientific conclusions are derived by two methods of reasoning: deduction and induction. *Deduction* is arguing from the general to the particular; that is, a general case is established, from which all dependent events are argued to be true. Thus, if one posits the truth of the general proposition 'all dogs are mammals', it follows by deduction that any particular example of a dog will be a mammal. *Induction* is arguing from the particular to the general. For instance, a dog may be vaccinated against distemper virus, and shown to be immune to challenge with the agent, from which the conclusion is drawn that the vaccine prevents distemper in all dogs.

One may accept (or reject) a causal hypothesis by four methods:

Slide: 24-28 (Thrusfield et al., 2018)

## *Tenacity*

Habit makes it easy to continue to believe a proposition and to offer a closed mind either to the opinions of others or to evidence that contradicts the proposition. Some people continued to believe that smoking was beneficial because it 'cleared the chest', even after Doll (1959) provided evidence that it induced lung cancer. The method of tenacity is unsatisfactory because it disregards the opinions of others, and, if they are considered, provides no framework for choosing between them.

## *Authority*

Source of authority can be baseless or well based. In the second case it may come from expert opinion, when the expert authority is generally acknowledged. This is a reasonable and widely-used approach. However, experts' opinions may vary, and such authority is only relatively final because opinions may be modified in the light of new knowledge or more convincing arguments.

## *Intuition*

Some propositions may be considered to be self-evident, without being sustained by evidence. Many veterinarians judged speed of slaughter of animals on infected premises to be crucial in the control of foot-and-mouth disease before firm evidence in support of this proposition was presented. Intuition may be moulded by training, experience and fashion. However, intuitive notions (e.g., that the Earth is flat) may subsequently be shown to be false. Therefore, intuitions need to be tested. "... *the intensity of the conviction that a hypothesis is true has no bearing on whether it is true or not.*" (Sir Peter Medawar, 1979)

## *Scientific inquiry*

- clarity, order and consistency
- independent of prepossession
- objective observations
- repeatability
- doubt (the scientist's prudent distrust of himself)
- science is progressive and is never too certain about its results

Differs radically from previous ones, which generally exclude the possibility of errors and have no provision for correcting them.

## *Models*

In everyday and scientific thinking, we use models, being simplifications of the complex real world. Due to its nature of this abstraction, none of the models represents the whole modelled peace of life. But it can be used in practice.<sup>4</sup>

Models constantly change over time. As long as the shortcomings of a model do not cause unacceptable consequences in practice, we tend to keep it. However, if its shortcomings are an obstacle, we will switch it to another model. The slides show how the Earth model has changed over history. Although Google Earth representation does not mean complete reality, it is a widely accepted model of the Globe today. We hope that the models will increasingly reflect reality so that their change is also an improvement.

In the example of Cusanus, the circle represents reality, and the square and polygon represent the model. As the number of angles of the polygon increases, it covers the circle more and more, and smaller and smaller parts of it remain uncovered.

Since Galileo Galilei (1564-1642), the paradigm that the physical rules of the physical world can be described mathematically has been prevalent in science. Among many examples, fractal geometry can capture complex natural patterns with simple mathematical formulas.

In the course, we will see that many mathematical, statistical models are used in epidemiology to quantify hypothetical rules. The vast majority of these models assume so-called linear relationships. However, there are many situations where they are unable to capture more complex biological patterns. With the development of computer science, the so-called deep learning methods are more and more widely used for complex biological questions. For example, a recent study claims that it can differentiate healthy people from Covid-19 patients through deep learning based on a photograph of the human eye.

Slide 29

<sup>4</sup> „All models are wrong but some are useful.,, (George E. P. Box, 1919-2013)

Slide 30-32

Slide 33-34

Slide 35

Slide 36

# Lecture 2

## Measures of health

Different measures are used to quantify health-related events in animal populations. We often use the risk and odds of the presence or appearance of a health-related event for this purpose. If in a population of ten animals there are three diseased animals (Fig 2), then the risk of the disease is  $3/10 = 0.3$ , while the odds of the disease is  $3/7 = 0.43$ . Note that the risk is a proportion; its values are ranging between 0 and 1, and can be interpreted as the probability of the disease occurrence.

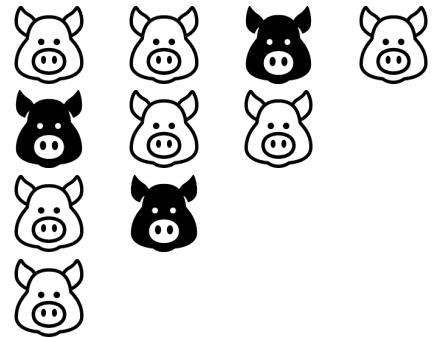


Figure 2: Three affected from ten animals (Slides 2)

## Incidence & prevalence

The two widely used measures to describe the disease occurrence in the populations are **incidence** and **prevalence**.<sup>5</sup> The purpose of the first measure is to quantify the number of new cases in the population within a certain time period. The number of **new cases** is the **incidence**. When our goal is to quantify the the number of affected animals in a population at a certain time point (or in a time period) then we use the **prevalence**. Usually the raw case numbers are normalized by any denominator to be comparable. In most of the cases the size of the **population at risk** is applied. The **population at risk** that part of the population that **might be affected**<sup>6</sup> by the given disorder.

The incidence has various types; in the veterinary literature, the most widely applied ones are the cumulative incidence and the incidence rate.

**Cumulative incidence**<sup>7</sup> (CI) in **closed** population:

$$CI = \frac{\text{number of new cases in the period}}{\text{size of the population at risk at the beginning of period}}$$

In **open** populations the size of the population is changing during the

<sup>5</sup> Slide 3

<sup>6</sup> In the COVID pandemic it is a fundamental problem of the correct risk assessment, the weakness of available data, eg. [COVID-19 Dashboard at Johns Hopkins](#). Since we at least suspect that the elderly are more likely to be fatally ill, it would be necessary to have age data in addition to the incidence data for an accurate analysis. But the age of tested, diseased and died persons are quite incompletely available, unfortunately.

<sup>7</sup> Slide 4-41

study period (selling, death, birth, import) the denominator:

- the population size at the half time of study period
- $N_{start} + \frac{1}{2}N_{new} - \frac{1}{2}N_{lost}$
- $N_{start} + \frac{1}{2}N_{new} - \frac{1}{2}N_{lost} + \frac{1}{2}N_{cases}$

The cumulative incidence is a proportion, without dimension, ranged between 0 and 1.

For instance, in a closed population with 100 animal population at risk, we've got the following occurrence table:

Week	New cases	CI
1	20	0.20
2	15	0.35
3	10	0.45
4	5	0.50
5	1	0.51

In the first week, we got 20 new cases, so the cumulative incidence for that time period is 0.2. In the next week, we found 15 new cases, so the cumulative incidence for the two first weeks is 0.35. For the 5 weeks period, the cumulative incidence is 0.51. At the end of the period, the probability of a randomly chosen animal is affected 51%. The **length** of the study period has strong influence on the cumulative incidence, the longer has higher.

On Slide 6, there is an example of the typical reporting form of cumulative incidence. But in other cases, the cumulative incidence is expressed as the number of new cases for fixed population size. The following slides (7-41) show an example of this type of reporting, based on yearly lung cancer cases in the USA.

The **incidence rate**<sup>8</sup> (IR) the number of new cases of disease that occur per unit of individual time at risk, over a defined follow-up period:

$$IR = \frac{\text{number of new cases in the period}}{\text{the sum of the individual time at risk}}$$

<sup>8</sup> Slide 42-45

The **time at risk** is the time spent by the individual without disease, i.e. the time to be in risk of disease. When an animal becomes ill, the rest of the period is not included into the calculation. E.g. if six healthy sows are studied for a year, and during this period none of them will be affected by the event, then the denominator for the calculation is "6 sow-year". The incidence rate can be calculated for /animal-week, /animal-year, /100 animal-week, /100 animal-year, etc.

Using the previous occurrence data, the calculation may be performed as follows. In the first week 20 animals became infected, so for them  $20 \times 0.5 = 10$  animal-week at risk, assuming that all became infected at the middle of the week. For the second week it

is  $15 \times 1.5 = 22.5$  animal-week. For the third  $10 \times 2.5 = 25$ , fourth  $5 \times 3.5 = 17.5$  and fifth  $1 \times 4.5 = 4.5$  animal-week at risk.

Week	New cases	Animal-week at risk
1	20	$20 \times 0.5 = 10$
2	15	$15 \times 1.5 = 22.5$
3	10	$10 \times 2.5 = 25$
4	5	$5 \times 3.5 = 17.5$
5	1	$1 \times 4.5 = 4.5$

The 49 remained healthy animals means  $49 \times 5 = 245$  animal-week. Summarizing  $10 + 22.5 + 25 + 17.5 + 4.5 + 245 = 324.5$  animal-week at risk, so the  $IR = 51/324.5 = 0.16/\text{animal-week}$ .

The incidence rate makes it possible:

- to quantify the disease occurrence in open population
- one individual may be affected more than one time

Difficulty is to record precisely the time at risk for all individuals. If this precise data recording is not applicable, there are some approximations for the denominator calculation:

- the population at risk at middle of period  $\times$  the length of study
- $(N_{\text{start}} + \frac{1}{2}N_{\text{new}} - \frac{1}{2}N_{\text{lost}}) \times$  the length of study
- $(N_{\text{start}} + \frac{1}{2}N_{\text{new}} - \frac{1}{2}N_{\text{lost}} + \frac{1}{2}N_{\text{cases}}) \times$  the length of study

On Slide 45, there is an example of the typical reporting form of incidence rate. On the slide 46, we see an example where the authors mistakenly refer to prevalence as the incidence rate in the abstract. To their advantage, the concept of prevalence is used correctly in the article itself, already. To demonstrate, to how often differs the abstract from the full text of articles slide 47 shows an example. Among others, the authors highlight that at least 1 inconsistency was found in 75% of studies.

Following the logic of incidence, similar measures can be constructed when the observed outcome is not the infection, disease, but death.

**Cumulative mortality**, CM is estimated as cumulative incidence:

$$CM = \frac{\text{number of died animals in the period}}{\text{size of the population at risk at the beginning of period}}$$

**Mortality rate**, M is calculated as the incidence rate:

$$M = \frac{\text{number of died animals in the period}}{\text{the sum of the individual time at risk of death}}$$

**Fatality rate**:

$$CF = \frac{\text{the number of fatal cases of a disease}}{\text{all individuals who contract the disease}}$$

## Survival analysis

Survival analysis has emerged as a special area related to the incidence of deaths in a population. It is important to note that in survival studies, the observed outcome is not necessarily the death.

In survival studies the time spent until a certain outcome from a starting point<sup>9</sup> is the base of health realted event quantification. As outcome different events might be considered (e.g. death, progression). In the studies the precise survival time is known just for subjects having the **end point** event occurrence.

A key question in survival studies is how likely<sup>10</sup> a patient survives a given period. During the times, several methods have been developed to estimate this. In the simplest case:

$$\text{survival rate}(t) = \frac{\text{No. of subject having survival time } \geq t}{n},$$

where  $t$  a certain length of a time period,  $n$  is the size of the population at risk at the starting of the study period. In a numerical example be  $n = 23870$

### 1. First month

- The number of death until the end of period: 591
- The rate of the survivors of this period:  $\frac{23870 - 591}{23870} = 0.975$
- Survival rate: 97.5%

### 2. Second month

- The number of death until the end of period: 591 + 1211
- The rate of the survivors of this period:  $\frac{23870 - (591 + 1211)}{23870} = 0.925$
- Survival rate: 92.5%

### 3. and so on

If we do this calculation for each month and plot the survival rates as a function of the elapsed time, we get the survival curve presented on slide 5. From this curve, various measures can be estimated. For example, the 5-year survival rate,<sup>11</sup> which shows us the proportion of the study population surviving five years. Another the median survival time,<sup>12</sup> giving the length of the time which was survived by the half of the study population.

There are always subjects having no end point event occurrence during the study period, or the progress of the disease is not followable from a certain time point for them. The events belonging these subjects are handled as **censored**.<sup>13</sup> The different versions of censored events are presented on slide 54. Nevertheless in the survival studies the time to censored events and the censored subject means also important information. Namely because of at least it is known about that the end point event has not been occurred until the censoring time point.

<sup>9</sup> time to event, survival time

<sup>10</sup> survival rate

<sup>11</sup> Slide 52

<sup>12</sup> Slide 53

<sup>13</sup> censoring

To construct survival plots using based on subjects with the observed endpoint and censored data, the Kaplan-Meier method can be applied.

An example of this is shown on slides 55-63. Imagine five patients. During the first month, no one had the event, so at the end of the first month, the survival rate was 100%. No event will occur during the second month, so the survival rate will remain 100%. During the third month, the event occurs in patient two, reducing the survival rate to 80%. From the end of the fourth month, we have no further information about the patient three, i.e. it has become censored. It is important to notice this change does not reduce survival curve level; only a censored mark appears on it. In the fifth month, the event occurs in the patient five, reducing the survival rate to 53%. There will be no event in the following months, so the survival curve will not change either. Analyzing with this method, the data of the example presented on slides 50-53, we get the figure of slide 64. If we read out the five-year survival rate and the median survival time, and compare them with the previous ones, we find the latter ones give much better prognostic values.

Slide 66 shows an example where the authors applied survival analysis on reinfection of PRRS free herd as endpoint. On the plot, one may observe that the two curves differ strongly. But in numerous studies, the question is the size of the difference between two survival curves. Since to answer this question we would need so-called parametric curves, but the Kaplan-Meier is not a parametric one, in practice, we approximate it by parametric models. From the first lecture were we may recall Cusanus's theory on the relationship between reality and models (slides 67-68). The following slides the Kaplan-Meier curve is approximated by Weibull (slide 70), Exponential (slide 71) and Logit (slide 72) parametric models. One may found the precision of different parametric models comparing to the Kaplan-Meier curve varies from point to point. It is important to realize the Kaplan-Meier is also a model of reality only.

### *Adjusting measures*

On slide 73, we can see the first figure of the series of yearly lung cancer incidence in the USA. It is striking that the risk of disease varies in different age groups. If we were to calculate the cumulative incidence from the same data set by ignoring the age groups, we would be ignoring this important internal structure. And then if we were to compare this crude incidence with a similarly calculated incidence in a country where the expected age is significantly lower, we could

draw erroneous conclusions. Therefore, if possible, it is worth to use adjusted measures besides the crude ones.

On slides 74-80 following Thrusfield's example, a direct and an indirect adjustment are presented (Thrusfield et al., 2018). In the example, dogs were tested for leptospirosis in two Scottish cities. They resulted in slightly different seroprevalence (slide 74). There is the knowledge that infection is more common in male dogs. If the sex ratio of the sampled individuals differs in the two cities, this may be the only reason for the difference in seroprevalences. If sex data are also available, this can be checked by indirect correction. The second table on slide 75 shows data and seroprevalences broken down by sex. If we recalculate the prevalences by city using the sex weighted formula shown on the slide 76, we can see that the probability of infection does not differ between the two cities.

In many cases, data are not available to stratify subjects. In those cases, indirect correction can be used whenever we express the risk within each group relative to the risk to the entire population. On the slide 78, we calculate the overall seroprevalence and then multiply this value by the sample size per city to obtain the number of expected positive cases per city (slide 79). Dividing the number of observed cases by the expected case number the standardized morbidity (mortality) ratio is obtained as a result, which can also be interpreted as the relative risk of infection by cities (slide 80).

# Lecture 3

## Diagnostic Tests

One of the important moments of veterinary activity is the establishment of diagnoses. In doing so, we take deviations from the normal condition into account and evaluate them according to their importance. However, in most cases, we may have uncertainties when making a diagnosis. To reduce these uncertainties, we use diagnostic tests. In a broader sense, diagnostic tests include questions asked during medical history, medical imaging procedures, laboratory tests, and autopsy findings.

In addition to using diagnostic tests to assess the condition of a particular individual, it can be used to examine the frequency of a disease or infection in a population or to operate screening, monitoring, and surveillance systems.

If we use a perfect diagnostic test, we can be sure that it will distinguish between a sick and a healthy animal without error. However, the vast majority of tests is imperfect. This error may be due to the imperfection of the applied technology or method. But it can also come from an uncertain case definition. The magnitude of this error can be determined by comparing the classification results of the certain test with the classification results of a diagnostic procedure that can be considered a gold standard.

As in Table 5, it is customary to compare the actual health status of the animals tested and their classification according to the tests performed with the particular test. The number of infected individuals found to be positive by the test gives the count of true positives; the number of uninfected individuals found to be negative by test gives the count of true negatives. The number of misclassified individuals are counted in false positive and false negative cells. Based on this type of contingency tables, one can evaluate the performance of the diagnostic tests.

<sup>13</sup> slide 2-4

	Infection +	Infection -
Test +	True positive <i>a</i>	False positive <i>b</i>
Test -	False negative <i>c</i>	True negative <i>d</i>

Table 5: Contingency table to compare the classification results of a certain test with the state of the individuals

## Test evaluation

Take an example in which we have a pig herd of 100 animals with 4 infected animals, i.e. a prevalence is 4%. Let's say our diagnostic test detects 3 of the 4 infected animals as positive. However, an additional seven of the 96 uninfected animals are also identified as positive. All other animals, 89 uninfected and 1 infected, were identified as negative by the test.

If we arrange the true positives/negatives and the false positives/negatives into a table (Table 6), then we can calculate measures to evaluate the performance the test was applied. The two most widely used measures for test evaluation is the sensitivity and the specificity. Both of them may range between 0 and 1. A higher value means better performance.

Sensitivity is the proportion of infected animals identified as positive by the test. Or the probability that a truly diseased animal will be classified as diseased. Sensitivity (Se) is calculated by dividing the number of true positives by the number of all infected animals.

	Infection +	Infection -	Total
Test +	True positive $a$	False positive $b$	Test positive $a + b$
Test -	False negative $c$	True negative $d$	Test negative $c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

$$Se = \frac{a}{a + c} = \frac{3}{3 + 1} = 0.75$$

Specificity is the proportion of uninfected animals identified as negative by the test. Or the probability that a truly non-diseased animal will be classified as non-diseased. Specificity (Sp) is calculated by dividing the number of true negatives by the number of all uninfected animals.

	Infection +	Infection -	Total
Test +	True positive $a$	False positive $b$	Test positive $a + b$
Test -	False negative $c$	True negative $d$	Test negative $c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

$$Sp = \frac{d}{b + d} = \frac{89}{7 + 89} = 0.93$$

In this example, the output of our diagnostic test can clearly be one of two values, i.e., the test gives either a positive or negative result.

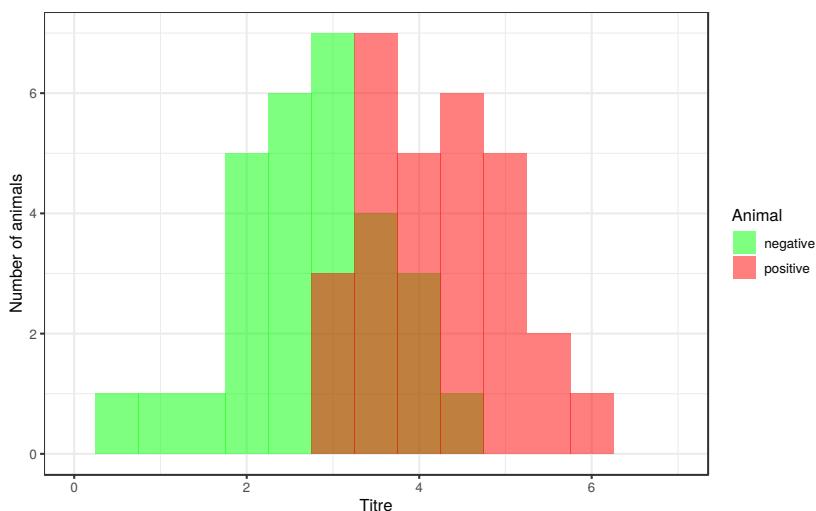
<sup>13</sup> slide 5-15

	Infection +	Infection -	$\Sigma$
Test +	3	7	10
Test -	1	89	90
$\Sigma$	4	96	100

Table 6: Contingency table of test based classification of the pigs.

However, such binary output diagnostic tests are less common. Most tests are based on measurements that produce values on a continuous scale. To obtain dichotomous outputs (positive/negative) by the application of such tests, it is necessary to determine a threshold. Below this limit, we consider the individual as a negative, while an animal with a value above the cutpoint is positive.

In the following example (Noordhuizen et al., 2001), infected and uninfected animals were tested by a serological test providing a continuous titer value (Table 7, Figure 3).



Titre	Number of animals negative	Number of animals positive
0.5	1	0
1.0	1	0
1.5	1	0
2.0	5	0
2.5	6	0
3.0	7	3
3.5	4	7
4.0	3	5
4.5	1	6
5.0	0	5
5.5	0	2
6.0	0	1

Table 7: The number of infected and uninfected animals by titer values.

Figure 3: The histogram of titer values in positive (infected) and negative (uninfected) animals. The same data presented by density plot on slide 22.

It can be seen that there are titers that were measurable only from infected or only uninfected animals. However, some titers occurred in both groups. Thus, depending on the cut-off value chosen to dichotomize our continuous titers, we construct a test with different sensitivity and specificity.

It is important to note that as the value of sensitivity increases, the specificity decreases, and vice versa.

If we modify the cut-off to increase the sensitivity, we increase the number of test-positive individuals, both true and false positives. If we change the cut-off value to increase the specificity, we increase the number of animals with a negative test result, both the number of true and false negative individuals.

Based on these, the question may arise as what threshold value may be optimal in diagnostic work? The answer is that it depends on the purpose of our testing.

If we increase the sensitivity of the test we may find more (or all) infected animals. On the other side it increases the probability that test-negative animals are not infected, indeed. When screening

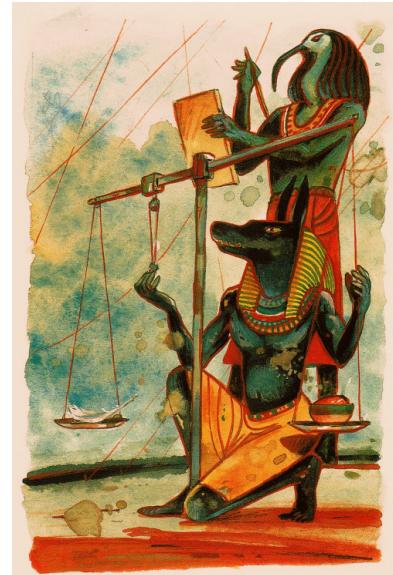


Figure 4: By Maat's scale Anubis evaluates the weight of hearts.

breeding pigs for purchase into a herd, a false-positive result would be much less harmful to a client than a false negative, which might allow infected pigs to enter a noninfected herd (Gardner, 2012).

If the goal is to make sure the test positives are truly infected, we need to increase the specificity. Fewer infected animals will be test positive, and parallel fewer uninfected animals will be test positive too. A veterinarian relying on the results of a test to decide to cull a sow probably wants to minimize the chance of a false-positive diagnosis by using a highly specific test, especially if the sow is asymptomatic and pregnant and there are no other reasons for culling (Gardner, 2012).

Slides 31-42 show how the contingency table and sensitivity/specifity changes according to cutpoint value increase. Slides 43-53 show how the values of sensitivity and specificity change as a function of the titer cut-off. On slide 54-64, the sensitivity values calculated at each limit are presented as a function of the associated specificity values. The use of the Receiver operating characteristic (ROC) curve to evaluate the overall performance of diagnostic tests is widespread (Figure 5). In that curve, the sensitivity is plotted as a function of 1-specificity (slide 65-76). Using ROC curve, it is possible to evaluate the general performance of a diagnostic test by one value. The closer the value of the area under the curve (AUC) to one, the better the performance of the diagnostic test.

### Predictive values

Since as we have seen tests might be imperfect, some questions arise. What is the probability that an individual is infected if its test is positive? What is the probability that an individual is not infected if its test is negative? The predictive values of the test give us the answer. Let's see an illustrative story (Woodworth, 2004).<sup>14</sup>

*Andrew got a tattoo. Two months later he was refused as a blood donor. The phlebotomist explained that he had to wait a year to make sure he didn't get hepatitis B from the tattoo. That got him worried, so he ordered a home test kit for hepatitis B virus (HBV) from a website. The website said that the sensitivity of the test was 0.99 and the specificity was 0.995.*

*Hepatitis B is rare among those who are not intravenous drug users – about 2 cases per 100,000 people. Studies suggest that getting a tattoo from an operator who follows accepted hygiene standards does not greatly increase the risk. Let's assume that Andrew believed that his risk was about 3 in 100,000.*

*If Andrew expect 10 million people as population at risk, then about 300 would have HBV, and the rest would not. As we know HBV test has 99% sensitivity, which means that it will catch 99% of the HBV cases (297 of the 300 cases) and miss the rest.*

*The test has 99.5% specificity, which means that 99.5% of the noninfected people will test negative, but 0.5% of them will be false positives.*

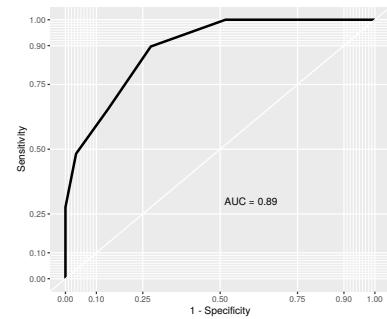


Figure 5: Receiver operating characteristic (ROC) curve.

<sup>13</sup> slide 77-98

<sup>14</sup> Another example: <http://yudkowsky.net/rational/bayes>

	HBV +	HBV -	$\Sigma$
Test +	297	49,998	50,295
Test -	3	9,949,702	9,949,705
$\Sigma$	300	9,999,700	10,000,000

Suppose Andrew tests negative. There are 9,949,705 people like him – negative. Of these only 3 have HBV, so there are 3 chances in 9,949,705 (about 1 in 3.3 million) that a person who tests negative actually is infected.

On the other hand, suppose Andrew tests positive. There are 50,295 people like him – positive. Out of this group, only 297 really do have HBV (about 1 of 170). That means that even if Andrew tests positive, there is still only about 0.6% chance that he is actually infected (Woodworth, 2004).

The predictive value of a test with other word is the post-test probability of the affected/unaffected state of the subject. To estimate this probability besides the sensitivity and specificity, the so-called pre-test probability is also utilized. Pre-test probability is the probability to be an individual affected before testing the individual. In the story above this pre-test probability of infection is the prevalence.<sup>15</sup>

Positive predictive value is the proportion of animals tested positive while they are truly infected. Or the probability that animals with a positive test result truly infected.

	Infection +	Infection -	Total
Test +	True positive $a$	False positive $b$	Test positive $a + b$
Test -	False negative $c$	True negative $d$	Test negative $c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

$$PPV = \frac{a}{a + b} = \frac{3}{3 + 7} = 0.30$$

Negative predictive value is the proportion of animals tested negative while they are truly uninfected. Or the probability that animals with a negative test result truly uninfected.

	Infection +	Infection -	Total
Test +	True positive $a$	False positive $b$	Test positive $a + b$
Test -	False negative $c$	True negative $d$	Test negative $c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

$$NPV = \frac{d}{c + d} = \frac{89}{1 + 89} = 0.99$$

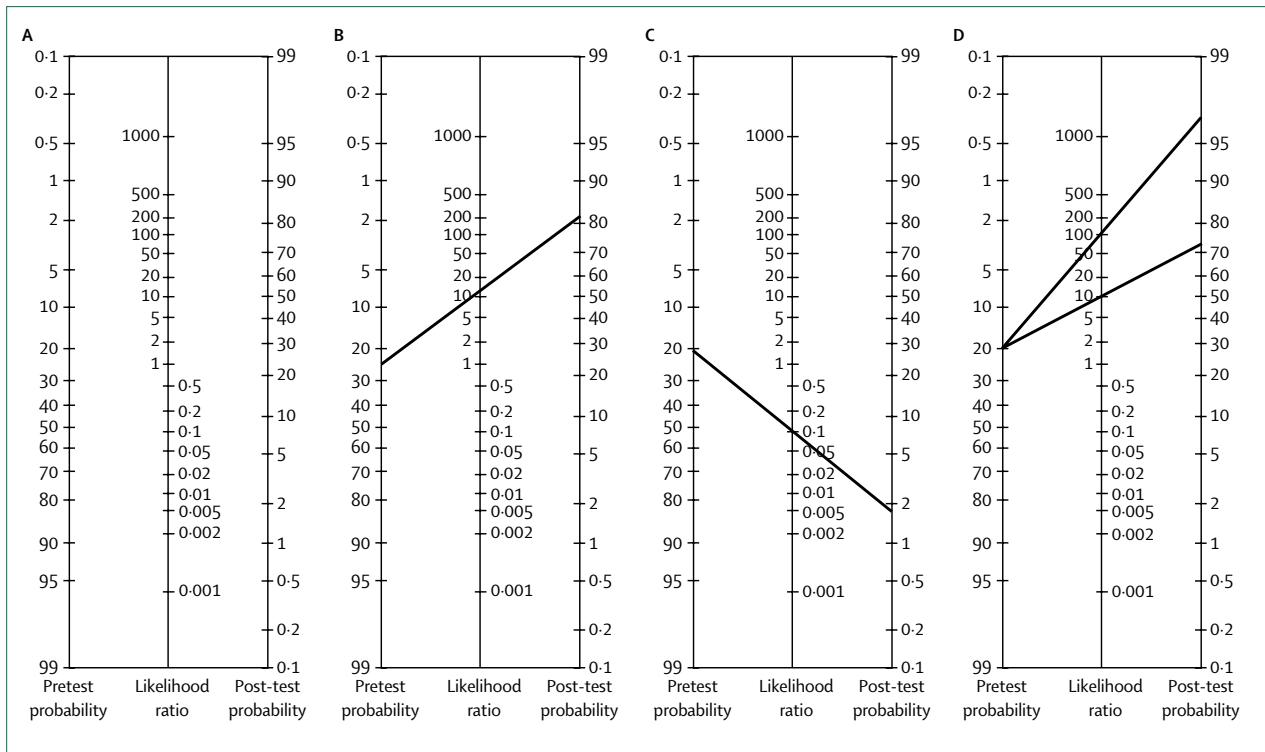
<sup>15</sup> The predictive values can be calculated without the contingency table, using only the prevalence (P), sensitivity (Se) and specificity (Sp) values:

$$PPV = \frac{P \times Se}{P \times Se + (1 - P) \times (1 - Sp)}$$

$$NPV = \frac{(1 - P) \times Se}{P \times (1 - Se) + (1 - P) \times Sp}$$

As the example on slide 86-94 shows, with the decreasing of prevalence (pre-test probability), the PPV decreases regardless of sensitivity and specificity of the test.

To estimate the post-test probability based on diagnostic test outcome, a further opportunity is the likelihood of ratios utilization. From sensitivity and specificity, one can calculate likelihood ratios for positive or negative test outcome. Using the Bayes' theorem the compilation of pre-test probability and likelihood, the post-test probability can be obtained. On slide 97 an example shows that if in a population with 5% prevalence a dairy cow has a positive test result by the California Mastitis Test, then the post-test possibility being the animal affected by mastitis is 11%. With less calculation the nomogram helps the user to read out the post-test probability when the pre-test probability and the likelihood ratio is known.



An interactive nomogram is available at [https://jamaevidence.mhmedical.com/data/calculators/LR\\_nomogram.html](https://jamaevidence.mhmedical.com/data/calculators/LR_nomogram.html).

Figure 6: Nomograms for probabilities and likelihood ratios (A) Nomogram reprinted from reference 13 with permission of the Massachusetts Medical Association. (B) Straight edge applied for pretest probability of 0.25 and likelihood ratio of 13. (C) Straight edge applied for pretest probability of 0.20 and likelihood ratio of 0.1. (D) Effect of likelihood ratios of 10 and 100 on pretest probability of 0.2 (Grimes & Schulz, 2005)

# Lecture 4

## *Multiple diagnostic tests interpretation*

To improve diagnostic accuracy, tests can be repeated, or additional tests may be involved. In fact, most diagnoses are based on multiple tests (e.g., medical history, physical examination, laboratory tests). Multiple tests can be applied simultaneously or consecutively, and the results can be interpreted in parallel or serial. Sensitivity and specificity values of test combinations differ from the sensitivity and specificity of individual tests. In interpreting results from a combination of tests, it is a fundamental assumption the tests must be independent of each other. If this independence is not satisfied, then the accuracy improvement will be lower than theoretically expected.

Such correlated results are expected when the combined tests measure the same/similar characteristics of the sample, but less likely when different biological responses (e.g. histopathological and serological testing) are the target of the tests. In parallel<sup>16</sup> testing, the sensitivity will be higher than the sensitivity of any individual test used. In serial<sup>17</sup> testing, the specificity will be higher than the specificity of any individual test used.

If two tests are used, one of the following four results is possible: both are positive; both are negative; the first test is negative and the second one is positive or the first test is positive and the second one is negative.

In a parallel interpretation, an animal is considered positive if one test is positive - this increases the sensitivity of the combined tests, but reduces its specificity. This parallel testing strategy is useful when none of the tests has a particularly high sensitivity, but they can detect different types of the disease (e.g. early - late, fast-slowly progressing). Culturing may be more sensitive than serological tests in the early stages of infection. Still, serology may be more sensitive in a later stage of it when the amount of pathogens is lower.

In serial testing, both consecutive tests must be positive to identify the animal as positive - this increases the specificity of the combined tests, but reduces its sensitivity. The first test can be high sensitivity

<sup>15</sup> slide 2-6

<sup>16</sup>

$$Se_{par} = 1 - (1 - Se_1) \times (1 - Se_2)$$

$$Sp_{par} = Sp_1 \times Sp_2$$

<sup>17</sup>

$$Se_{ser} = Se_1 \times Se_2$$

$$Sp_{ser} = 1 - (1 - Sp_1) \times (1 - Sp_2)$$

and inexpensive; the result can be followed by a high-specificity test to determine false positives. It is a cost-effective approach if the first test negatives are not tested by the second test. This strategy allows the vet to use fewer tests to rule out the disease; however, it is more time-consuming. When both tests are positive, for the estimation of disease probability, the first test positive predictive value will be the pre-test probability of the second test.

For example, for a test A, the positive predictive value was 67.9% applied in a herd with prevalence 20%. If an animal tested by A was positive and was retested with another test B (sensitivity of 45.9%, specificity of 96.9%) the positive predictive value of 67.9% is considered as the pre-test probability for B-test. Using the Bayes theorem, the positive predictive value after the application of the second test will be 96.9%, assuming that the results of A and B are not correlated. If the condition of uncorrelation of the tests can be maintained, then the two positive values obtained by using A and B together are a stronger indicator of infection than it was predicted by test A alone (Gardner, 2012).

Occasionally, the question arises to quantify the agreement<sup>18</sup> in the classification results of different tests used in the diagnosis of the same disease. To investigate this, various statistical procedures can be found in the literature. One commonly used statistic is Cohen's kappa (Cohen, 1960).

So far, we have seen that the bias of the tests causes some degree of diagnostic uncertainty. The size of this can be further enhanced if we take into account that the estimates of sensitivity and specificity of the same test from different studies may differ significantly.<sup>19</sup>

It is also important to see that while so far only one value (point estimate) has been used to quantify sensitivity and specificity, these values themselves have uncertainties that can be characterized by the width of the confidence interval of them.<sup>20</sup> The wider the confidence interval, (i.e. the larger the difference between the lower and upper values) the less certain the value estimated (point estimate) for sensitivity and specificity.

### *Prevalence adjustment*

If the applied diagnostic test is biased, the number of cases we count by them does not mean the number of individual actually affected.

Prevalence represents the fraction of existing cases in a population, i.e. the ratio between the number of diseased animals and the total number of animals at risk, or the probability that a randomly-chosen animal is diseased.

<sup>17</sup> slide 7

<sup>18</sup> „In the comparison of IFAT and PCR or cELISA and PCR, weaker agreements ( $\kappa = 0.83$  and  $\kappa = 0.80$ , respectively) were obtained.” (Farkas et al., 2013)

<sup>19</sup> slide 8

<sup>19</sup> see Elanco Keto-Test table

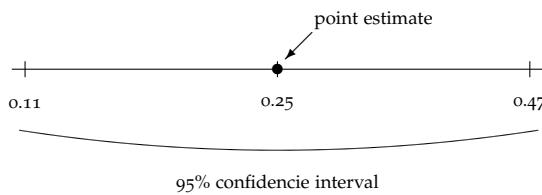
<sup>20</sup> see IDEXX Milk Pregnancy Test, IDEXX SNAP BVDV Ag Test examples

<sup>20</sup> slide 9-24

Point prevalence is the proportion of infected individuals in a defined population at a given time point. Period prevalence is the proportion of infected individuals in a defined population found over a specified time period.

If the case definition is based on an imperfect test, the test bias should be taken into account in prevalence estimation.

Accordingly, we distinguish between apparent ( $P_A$ ) and true ( $P_T$ ) prevalence. In the former case, we do not take into account that our test is biased. While for the latter, the apparent prevalence is adjusted by the test bias (sensitivity, specificity). In the literature, in addition to presenting the prevalence (either apparent or true) as a point estimate, it is common to give a 95% confidence interval (CI) to demonstrate the reliability of point estimate.<sup>21</sup>



<sup>21</sup> The reliability of a confidence interval only represents a level of confidence in its generation. Thus, it does not mean that the unknown studied population parameter (here the prevalence) is fallen within the given range with a 95% probability. Instead, it means that 95 of the 100 estimated intervals taken from the same population by the same sample size, will contain the true value of the parameter of interest.

As we increase the sample size to estimate prevalence, we increase the accuracy of our point estimate, which is indicated by a decrease in the width of the confidence interval (slide 16).

The Rogan-Gladen estimator is a widely used formula for estimating true prevalence:

$$P_T = \frac{P_A + Sp - 1}{Sp + Se - 1}$$

The slide 21 shows how differs the true prevalence (calculated from the  $P_A$ , Se and Sp) comparing to the apparent prevalence estimated by a test used in the diagnosis of bovine paratuberculosis (Messam et al., 2008). On the same slide, the second case shows that if prevalence is low, the Rogan-Gladen estimator may result in negative prevalence, which is not meaningful, erroneous. To perform similar adjusted prevalence estimation different tools are available (slide 22).

In Bayesian estimation the most widely used formulation:

$$x|P_A, Se, Sp \sim binomial(n, P_T Se + (1 - P_T)(1 - Sp))$$

Applying this estimation on the previous example, we get 0.02 as the true prevalence. Besides this result from the posterior distributions provided by the Bayesian estimation, further inferences can be drawn. This approach needs more computations skills but to help vets, numerous resources are available to do that (slide 24).



Thomas Bayes (1702 – 1761)

## Herd level interpretation of diagnostic test results

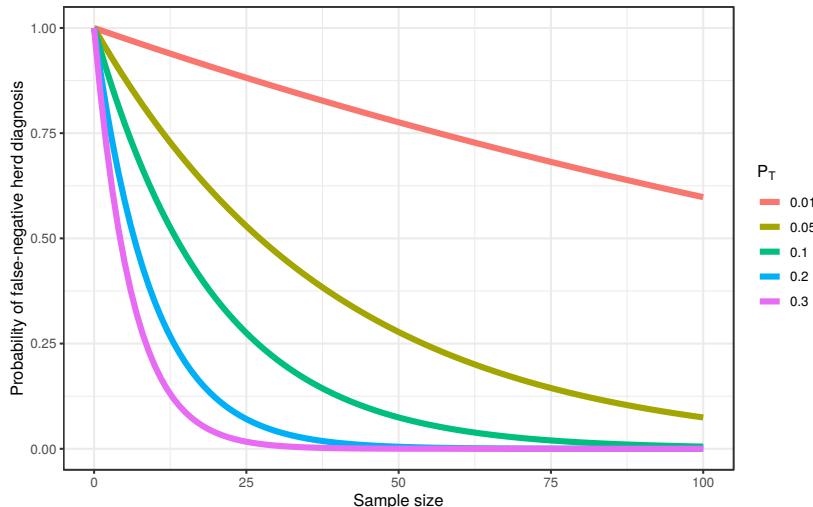
In many cases, the health state of a population unit (farm, barn, litter or other groups) is the subject of study, not the state of the individuals. It is not commonly known the heard-level tests should be interpreted differently than individual tests. Herd-level interpretation of tests is often more complex, especially if the tests used have no perfect specificity. As in the case of individual test results, for the correct decision on the state of a herd, the herd-level sensitivity and specificity of applied tests must be known. Usually, the most likely performance of herd tests is estimated based on the individual level sensitivity and specificity of applied tests.

Herd-level sensitivity (HSE) is the probability that an infected herd will be detected as positive by the test. Herd-level specificity (HSP) is the probability that an uninfected herd will be detected as negative by the test. Herd-level sensitivity and specificity besides the individual level sensitivity and specificity of the applied test depend on further factors, like the sample size, the prevalence within the farm and the critical number.<sup>22</sup> Critical number is the number of the animals should be positive in the population to consider the herd as positive.

<sup>21</sup> slide 25-30

<sup>22</sup> The slide 29 and 30 shows the calculation formulas of HSE and HSP with different critical numbers.

Figure 7: The dependency of false-negative herd diagnosis on sample size and prevalence.



From Figure 7 we may draw conclusions. With the increase of sample size the herd-level sensitivity increases, while the probability of the false-negative herd detection decreases. Observing the effect of prevalence, we can find that the probability of the false-negative herd-detection decreases faster when the within-herd prevalence is higher. Also, it is easier to distinguish between infected and uninfected herd by the same sample size when the within-herd prevalence is higher.

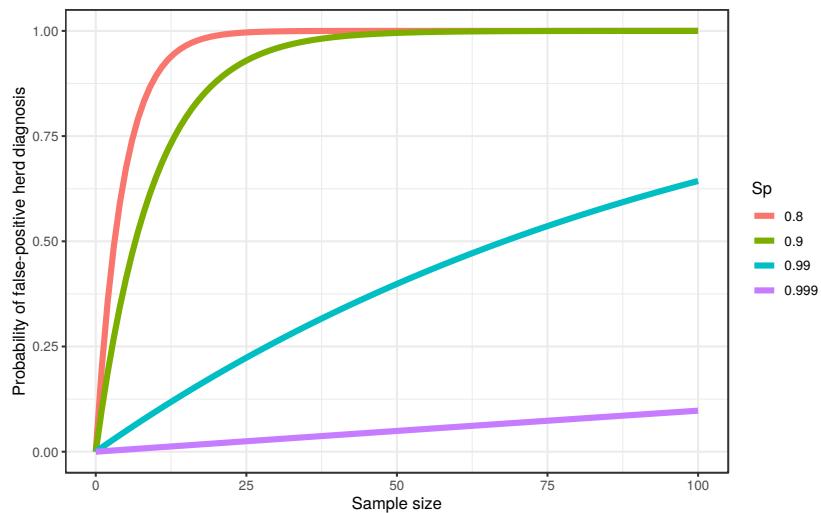


Figure 8: The dependency of false-positive herd diagnosis on sample size and specificity.

Based on Figure 8 we can find that by the increase the samples size the probability of finding at least one false positive animal is increasing. This increase is faster with lower specificity and slower with higher sensitivity. Considering that proportionally more samples are usually taken from larger flocks, we may conclude that larger herds will be more often false positive.

### *Sample size for disease detection*

To gain insight into the health status of a herd, we need to gather data. If we collect data on all animals, we perform a census. If a survey is designed well, then a reasonably accurate and acceptable estimate of a variable can be made by examining some of the animals in the relevant population; that is, a sample.<sup>23</sup> The target population is the total population about which information is required. The study population is the population from which a sample is drawn. The study population consists of elementary units, which cannot be divided further. A collection of elementary units, grouped according to a common characteristic, is a stratum.

Veterinarians often need to determine whether an infection is or has ever been present in the herd or a subpopulation of the herd.<sup>24</sup> For tests of 100% specificity, a single positive is usually considered sufficient to class the herd as positive, although for serological tests of imperfect specificity, more than one positive might be necessary. To estimate required numbers to detect infection, the following values are necessary: the required level of confidence, usually 95%, the likely

<sup>22</sup> slide 31-36

<sup>23</sup> For sample size estimation for various scenarios numerous tools are available (slide 34-35)

<sup>24</sup> If a veterinarian's only goal is to detect infection, sampling does not need to be random but can be directed to higher-prevalence groups, for example, different age groups when there is an age-related risk of infection or clinically affected versus otherwise healthy pigs.

prevalence of infection in the herd or in the specific group of pigs being evaluated, the population size.

The selected prevalence value should be realistic, but if there is doubt, erring toward a lower prevalence is preferable to ensure that adequate numbers of pigs are sampled. Figure 9 shows the relationship between required sample size and herd size, expected prevalence for different PRRS tests (Table 8).

Table 8: The sensitivity and specificity of various PRRS tests.

Test	Sample	Se	Sp
FMIA-Se	serum	73.3%	73.3%
HIPRA-OF	saliva	20.0%	100.0%
HIPRA-Se	serum	66.7%	93.3%
IDEXX-OF	saliva	86.7%	100.0%
IDEXX-Se	serum	100.0%	100.0%
IDEXX-SO	saliva	60.0%	93.3%

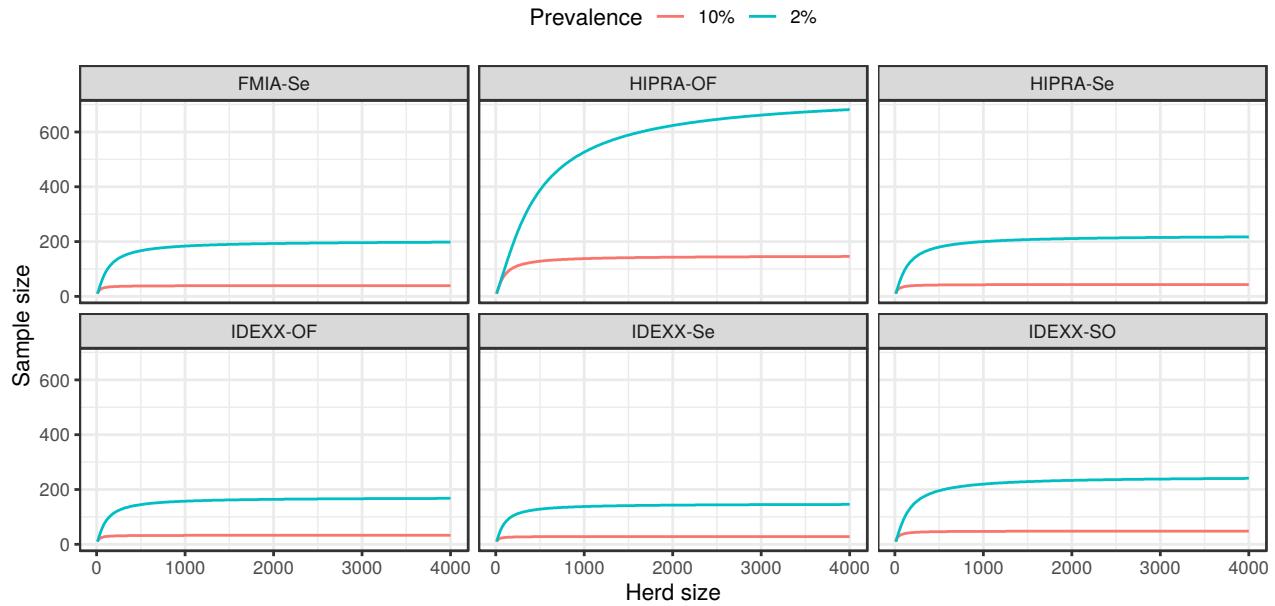


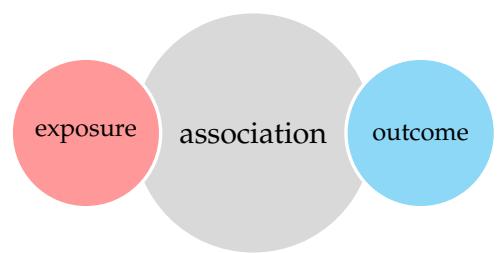
Figure 9: The dependency of sample size on herd size and expected prevalence.

# Lecture 5

## Analytic epidemiology

The topics discussed in previous lectures are the so-called descriptive epidemiology. Another major area of epidemiology is analytical epidemiology. Within this, the task is analysing the association between the suspected causal factor and the disease in the population. Absolute and relative association measures are used to express the extent of the relationships identified between exposures and outputs.

slide 2



## Study designs

Different study designs are used to examine the relationships between putative risk factors and diseases. In clinical trials, individuals are actively classified into a treatment or exposure group. In observational studies, the person performing the study does not actively allocate individuals into groups, but the subject of the study belongs to a group based on observation.

slide 3-7

In case reports and case series, it is important to note that they are not suitable for the study of causal relationships because no control individuals are involved. In case studies, information collected about a single diseased, affected individual is used to understand the nature of the disease. The cases series can give an insight into the variability of the disease in different individuals, but still without controls.

In cross-sectional studies, a random sample is taken from the target population, and the individuals in the sample are examined to determine whether or not they are affected by the disease. Then we collect data for each individual on whether or not they were exposed to the putative risk factors. Finally, we analyse the relationship between exposure and outcome data. This last step is the same in all following designs.

In case-control studies, individuals are selected for investigation based on their health status, into case and control groups. Then, similarly to cross-sectional studies, we collect data for each individual on their exposure state.

In cohort studies, individuals are selected for comparison according to whether or not they have been exposed to the putative risk factor. We then either collect data retrospectively on whether the disease of interest appeared in them or observe them prospectively for the development of the disease.

In randomised controlled (clinical) trials, the target population consists of individuals with the same health state. Of these, individuals are randomly allocated to one or the other treatment group. In addition to randomisation, it is also important that treatment groups are balanced in terms of other factors influencing disease outcome (e.g. age, sex). Patients are monitored continuously during treatment. At the end of the study, the effect of the treatments on health is analysed and evaluated.

### *Association of disease and exposure*

The slide 8-9 show some published examples of association studies with the main questions asked in them. In the questions, exposure and outcome moments are highlighted. If we can identify factors that can be related to the outcome, it can be a base of controlling the disease, even if we don't know the exact causal chain.

The simplest case of association studies is when both health and exposure status can take only two values. For example, diseased/not diseased and exposed/not exposed. The previously used  $2 \times 2$  table helps to analyse these situations (Table 9).

slide 12-16

	Disease+	Disease-	$\Sigma$
Exposure+	$a$	$b$	$a + b$
Exposure-	$c$	$d$	$c + d$
$\Sigma$	$a + c$	$b + d$	$a + b + c + d$

Table 9: Contingency table for calculation of association measures

Calculating various risks from the table, those can be used to calculate association measures:

- risk of the disease in the exposed group:  $R_{E+} = \frac{a}{a+b}$
- risk of the disease in the unexposed group:  $R_{E-} = \frac{c}{c+d}$
- risk of the disease in the whole population:  $R_T = \frac{a+c}{a+b+c+d}$

Association measures are often grouped as measures of effect, total effect, therapeutic effect and strength.

Using data published by Nakashbandi et al. (2020) we go through the calculation and interpretation of some more frequently used measures. In their work, by a cohort design, they studied the effect of

slide 17-19

obesity on the recovery of COVID-19 patients. They found that 51 died of the 139 normal-weight patients, while 81 died of the 150 over-weighted patients. The first question arises, is there an association between death risk and obesity? And if there is, how strong is it? Doing calculations based on the table we get the incidence risk<sup>25</sup> among the normal-weight patients  $R_{E-} = 51/139 = 367$  cases/1000, among the overweight patients  $R_{E+} = 81/150 = 540$  cases/1000 and in the whole population  $R_T = 132/289 = 457$  cases/1000. Using these values we calculate association measures.

<sup>25</sup> cumulative incidence, cumulative mortality

### Measures of effect

The *attributable risk* (AR) is the incidence risk of disease in the exposed group that is attributable to exposure. It describes the absolute frequency of disease associated with the exposure.

slide 20-21

$$AR = R_{E+} - R_{E-} = 540 - 367 = 173 \text{ cases/1000}$$

From the example the incidence risk of death in overweight COVID-19 patients attributed to overweight is 173 cases/1000.

The *attributable fraction* (AF) is the proportion of disease in the exposed group that is due to exposure. The proportion of disease in the exposed that is due to exposure.

$$AF = \frac{R_{E+} - R_{E-}}{R_{E+}} = \frac{540 - 367}{540} = 0.32$$

In our example, 32% of death in overweight COVID-19 patients is attributable to overweight (AF=0.32)

### Measures of total effect

The *population attributable risk* (PAR) is the incidence risk of disease in the population attributable to exposure.

slide 22-23

$$PAR = R_T - R_{E-} = PAR = 457 - 367 = 90 \text{ cases/1000}$$

So, the risk of death in COVID-19 patients that may be attributed to overweight is 90 cases/1000 (PAR=0.09).

The *population attributable fraction* (PAF) is the proportion of disease in the population that is due to exposure.

$$PAF = \frac{R_T - R_{E-}}{R_T} = \frac{457 - 367}{457} = 0.20$$

That is, 20% of death in COVID-19 patients is attributable to overweight (PAF = 0.20).

## Measures of therapeutic effect

A special approach when we are interested in the effectiveness of the treatment. To quantify the treatment effectiveness, we go through some more frequently readable measures. Here we introduce the concept of event rate, which is calculated as the incident risk or cumulative incidence used so far. The *event rate* is the number of individual experiencing an event as a proportion of the number of individuals in the population. The *absolute risk reduction* (ARR) is the absolute difference between the untreated and treated group event rates. The *absolute risk increase* (ARI) is the absolute difference between the treated and untreated group event rates. The so-called *relative risk reduction* (RRR) is the difference in event rates between 2 groups, expressed as a proportion of the event rate in the untreated group. The *number needed to treat* (NNT) is the number of patients who would have to receive the treatment for 1 of them to benefit ( $NNT = 1/ARR$ ). And the *number needed to harm* (NNH) is the number of patients who would have to receive the treatment for 1 of them to experience an adverse effect ( $NNH = 1/ARI$ ).

Suppose we study the effect of a new drug on COVID-19 patients with high-risk (e.g. older men). The risk of death is 40% and 30% in the untreated and the treated group, respectively. Based on these risks the absolute risk reduction of the new drug is  $0.4 - 0.3 = 0.1$ , while the relative risk reduction is  $(0.4 - 0.3)/0.1 = 0.25$ .

When we used the same medicine in a lower risk group (e.g. young women). In that study the risk of death is 100% and 7.5% in the untreated and the treated group, respectively. So, the  $ARR = 0.1 - 0.075 = 0.025$  and the  $RRR = (0.1 - 0.075)/0.1 = 0.025/0.1 = 0.25$ .

The absolute risk reduction becomes smaller when event rates are low, whereas the relative risk reduction, or “efficacy” of the treatment, often remains constant.

Continuing the example, while in the high-risk group the number needed to treat is  $1/(0.4 - 0.3) = 10$ , in the low-risk group it is  $1/(0.1 - 0.075) = 40$ .

If the NNT for treatment is 10, the practitioner would have to give the treatment to 10 patients to prevent 1 patient from having the adverse outcome. Another conclusion is that each patient who received the treatment would have a 1 in 10 chance of being a beneficiary. If the absolute risk reduction is large, you need to treat only a small number of patients to observe a benefit in at least some of them. If the absolute risk reduction is small, you must treat many people to observe a benefit in just a few.

slide 29-44

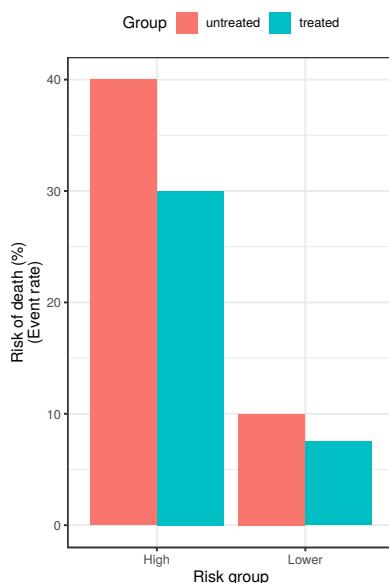


Figure 10: Effectiveness of treatment in groups of different baseline risks.

### Measures of strength

The *relative risk* (RR) is the incidence risk of disease in the exposed ( $R_{E+}$ ) divided by the incidence risk of disease in the unexposed group ( $R_{E-}$ ). For the data of Nakeshbandi et al. (2020):

$$RR = \frac{R_{E+}}{R_{E-}} = \frac{0.54}{0.367} = 1.47$$

We can interpret the result as the incidence risk of death in overweight COVID-19 patients is 1.47 times higher than the incidence risk of death in normal-weight COVID-19 patients. In the scientific literature usually the relative risk is reported with its 95% confidence interval, e.g. RR: 1.47, 95%CI: 1.13 – 1.92.

When the time at risk is available, the incidence rate (IR) ratio is the incidence rate of disease in the exposed divided by the incidence rate of disease in the unexposed.

The relative risk<sup>26</sup> provides an estimate of how many times more likely exposed individuals are diseased compared with non-exposed individuals. If  $RR = 1$  then the risk of disease in the exposed and non-exposed groups are equal. If  $RR > 1$  then exposure increases the risk of disease with greater departures from 1 indicative of a stronger effect. If  $RR < 1$  then exposure (e.g. treatment) decreases the risk of disease, so it is a preventive factor.

Whereas in case-control studies, incidence risk can't be calculated among controls, therefore no relative risk can be calculated. In those studies, the odds ratio<sup>27</sup> provides an alternative to express the strength of association.

The *odds ratio* (OR) is the odds in one group divided by odds in another group. In the cohort and case-control studies, the meaning is altered (Table 10).

For cohort studies the OR is the odds of disease in the exposed group ( $O_{E+}$ ) divided by the odds of disease in the unexposed group ( $O_{E-}$ ). Using the cell annotations of Table 9 the odds in exposed group  $O_{E+} = a/b$  and in unexposed group  $O_{E-} = c/d$ . With the data of Nakeshbandi et al. (2020) the odds ratio is calculated as:

$$OR = \frac{O_{E+}}{O_{E-}} = \frac{a/b}{c/d} = \frac{81/69}{51/88} = \frac{1.17}{0.58} = 2.017$$

That can be interpreted, the odds of death in overweight COVID-19 patients is 2.02 times higher than the odds of death in normal-weight COVID-19 patients. The confidence interval for the point estimate is 95%CI: 1.26 – 3.24. If  $OR = 1$  then the odds of disease in the exposed and non-exposed groups are equal. If  $OR > 1$  then exposure odds of disease with greater departures from 1 indicative of a stronger effect.

slide 29-44

<sup>26</sup> incidence risk or incidence rate ratio

<sup>27</sup> one may recall Figure 2 to the difference between odds and risk

If  $OR < 1$  then exposure (e.g. treatment) decreases the odds of disease, so it is a preventive factor.

On slide 38, the relationship between the odds ratio and the relative risk is presented. If the overall risk ( $p$ , the colours of the figure) is low, then the numerical value of OR and RR is close, but as the overall risk increases, their values more and more differ.

For case-control studies the OR is the odds of exposure ( $O_{D+}$ ) in the diseased divided by the odds of exposure in the undiseased ( $O_{D-}$ ). Using the cell annotations of Table 9 the odds in diseased group  $O_{D+} = a/c$  and in undiseased group  $O_{D-} = b/d$ .

$$OR = \frac{O_{D+}}{O_{D-}} = \frac{a/c}{b/d}$$

Cohort study	Case-control study
<i>Sequence</i>	
1. Define exposure status 2. Define disease status	1. Define disease status 2. Define exposure status
<i>Measure of strength</i>	
RR or OR	OR
<i>Interpretation of odds ratio</i>	
Odds of disease in exposed, compared with odds of disease in unexposed	Odds of exposure in diseased, compared with odds of exposure in undiseased

Table 10: Measures of strength for cohort and case-control studies (Stevenson, 2012)

To present more than one OR (RR, etc.), the forest plot is a popular visualisation method. As on slide 43 is presented for each factor, there is plotted a point estimate with a confidence interval. As it was mentioned above if the OR/RR is equal to 1, then the odds or risk in the two studied groups is equal. This is why we call the value 1 of these measures as no effect value. To help the readers to evaluate the effect of factors, a vertical line is drawn at the no-effect value.

Some statements of Cuzick (2005) might be useful to consider in the evaluation of forest plots. Confidence intervals in subgroups are always wider than those for the main effect because of smaller numbers. If the interval for a subgroup crosses the no effect point, this is widely misinterpreted as a lack of effect in the subgroup even where the overall effect is significant. Interpretation of subgroup effects would be helped if this line was deemphasised or omitted and replaced by a bold vertical line at the overall treatment effect level, making it easier to see if a subgroup confidence interval differed significantly from the overall effect.

### *Uncertainty in associations*

As mentioned above, besides odds ratio or relative risk, their confidence interval is reported as well. Also, the p-value is frequently reported for those measures. Both the confidence interval and the p-value are used to quantify the uncertainty of our estimate.

Lakos et al. (2012) studied whether the risk of adverse pregnancy outcomes among women with Lyme borreliosis is associated with antibiotic treatment? The usual arrangement of the exposure and outcome data is shown in Table 11.

		Adverse outcome		$\Sigma$
		+	-	
No treatment	6	4	10	
	Per os AB	6	13	19
$\Sigma$	12	17	29	

From Table 11 the estimates are OR: 3.11, 95%CI: 0.51 - 21.60, p=0.24. It can be interpreted as in the no-treatment group, the odds of an adverse outcome is 3.11 times higher than in the per os antibiotic treatment group. A practical interpretation is our estimate is not significant **STATISTICALLY**, because  $p > 0.05$  and/or 95%CI contains the value 1.0. A correct statistical interpretation of both values is far from everyday thinking, but some remarks must be mentioned. The p-value expresses the probability that the value of OR is at least 3.11 if the true odds ratio is equal to 1. The p-value is not the probability of  $H_0$  is true (false); it expresses only the extremity of observed values assuming  $H_0$ .

It is essential taking into account the p-value and width of confidence interval depends on the sample size. Imagine doubling the number of women in the study, but without changing the proportions in the groups (Table 12).

From Table 12 the estimates are: OR: 3.18, 95%CI: 0.92 - 11.70, p=0.051. As we can see, doubling the sample size with the same risks, the OR was changed slightly, the confidence interval became narrower, and the p-value decreased notably. But the result is still not statistically significant, even if it is only 0.01 above the decision threshold. And here is a property of statistical significance decision limit what is a subject of critics, that is a negligible difference of p values may have a meaningful consequence, while  $p = 0.051$  is statistically not significant,  $p = 0.049$  is already statistically significant.

If we triple the number of women in the study, without changing the proportions in the table, the estimates are OR: 3.20, 95%CI: 1.18 - 9.04,

slide 45-50

Table 11: Adverse outcomes and antibioticum treatment contingency table (Lakos & Solymosi, 2010). Adverse outcomes: cavernous hemangioma, cerebral bleeding, dysplasia coxae, hypospadias, muscular hypotonicity, neonatal jaundice requiring exchange transfusion, papulovesicular eruption at birth, premature birth pyloric stenosis, skeletal anomaly, small for dates, spontaneous abortion, stillbirth.

		Adverse outcome		$\Sigma$
		+	-	
No treatment	12	8	20	
	Per os AB	12	26	38
$\Sigma$	24	34	58	

Table 12: Adverse outcomes and antibioticum treatment contingency table for doubled sample size.

$p=0.013$ . Since the 95%CI does not contain value 1, and the p-value is below 0.05, our effect size estimate became statistically significant.

The example of Spiegelhalter et al. (2004) shows a similar property of p-value. Imagine that patients got A and B treatment. After that we asked them what treatment would they prefer. Table 13 shows the results of the hypothetical study with changing sample size.

Number of patients receiving A or B	Numbers preferring A : B	% preferring A	two-sided p-value
20	15 : 5	75.00	0.04
200	115 : 86	57.00	0.04
2 000	1046 : 954	52.30	0.04
2 000 000	1 001 445 : 998 555	50.07	0.04

As the sample size is increased with unchanging statistically significant p-value, the preference difference is decreasing. At the largest sample size, a completely irrelevant preference difference is still statistically significant.

The p-value is an important aid in assessing the reliability of our estimate. Still, the rigid application of its decision limits leads to erroneous professional conclusions or unpublished important scientific results.

Amrhein et al. (2019) and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects. As they emphasised: “*Eradicating categorization will help to halt overconfident claims, unwarranted declarations of ‘no difference’ and absurd statements about ‘replication failure’.*”

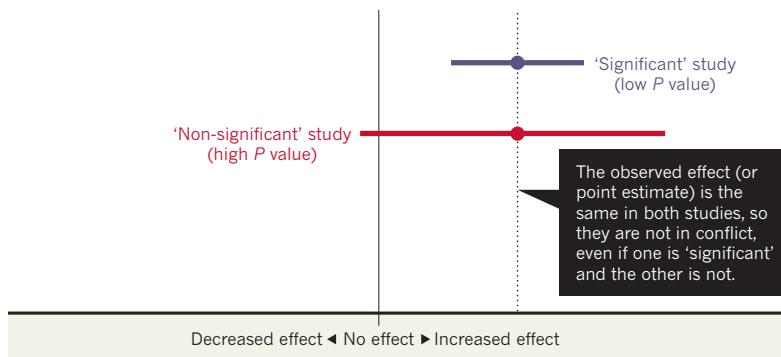


Table 13: Hypothetical study on treatment preference (Spiegelhalter et al., 2004)

Figure 11: BEWARE FALSE CONCLUSIONS. Studies currently dubbed ‘statistically significant’ and ‘statistically non-significant’ need not be contradictory, and such designations might cause genuine effects to be dismissed (Amrhein et al., 2019)

# Lecture 6

## *Association with more than 2 levels of categories*

So far, we have dealt with association studies in which there were two categories in terms of both exposure and outcome. In the previously presented risk assessment of treatments among women with Lyme borreliosis on adverse pregnancy outcomes; actually, there were three treatments (Lakos & Solymosi, 2010). That is, more than two. In such cases, the effects of the treatments can also be compared pairwise, as described earlier. However, it is common to analyze the relationship between outputs and exposures at the same time. In its simplest approach, this is done by any independence test (Chi-square, Fisher exact test). In general, the associations are evaluated based on the p-value given by the test to learn whether the output is independent of exposure. That is, if the p-value is less than 0.05, then the output is not independent of the treatment, i.e., we can interpret it as a statistical relationship between them. However, this analysis says nothing about which treatment and how it affects the outcome.

We can use the association plot to analyze this. This depicts Pearson's residuals expressing the difference between observed and expected case numbers. Expected values are case numbers that would occur if there were no relationship between treatments and outcomes. These can be calculated by fixing the column and row sums. The association figure is interpreted as evaluating the deviation of the coloured bars from the horizontal dotted lines.

For example, in the patients with adverse outcome of NO treatment group, the bar is directed upward from the dotted line. This means that we can observe more cases among the untreated ones than we would expect in the case of independence. The higher the column, the more the observed number of cases differs from the expected one.

In the PAR treatment group, we can see that the bar deviates downwards from the dotted line, i.e., the number of cases in this treatment group is less than would be expected in the case of independence. Among mothers who did not have an adverse outcome, the effects of the treatments were seemingly opposite to those who had an adverse

slide 2-5

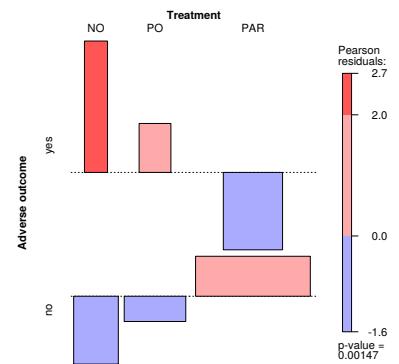


Figure 12: Association plot with  $2 \times 3$  categories

outcome. The width of the columns is proportional to the expected number of cases.

### *Relationship between continuous & continuous variables*

We often come across studies that analyze the relationship between continuous variables. In the example shown, the ambient temperature inside and outside of a barn was measured.

In studies like this, a correlation analysis is performed on the question of whether there is a relationship and how strong it is between the two variables. The resulting R-value ranges from -1 to 1. The more it deviates from zero, the stronger the relationship. If it has a negative value, then an increase in one value is co-occurred by a decrease in the other, and vice versa. If its value is positive, then an increase in one is co-occurred by an increase in the other, and vice versa.

If we are looking for an answer to how much a change in one continuous variable results a change in another continuous variable, then linear regression analysis can be performed. In our example, we examine the dependence of the internal temperature on the outside temperature. From the regression results, it is important to know the interpretation of the regression coefficient ( $\beta$ ). It expresses how much an increase in the explanatory variable (outdoor temperature) by unit 1 results a change in the value of the dependent variable (indoor temperature). In this example, this 0.806852, i.e. an increase in the outside temperature of 1 degree Celsius increases 0.81 degrees Celsius of the internal temperature.

slide 6-7

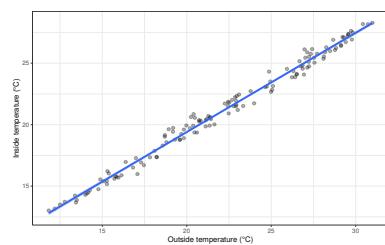


Figure 13: Internal temperature as a function of the outside temperature, with linear trend line

### *Logistic regression*

Generalized linear models are an extension of linear regression models. These allow us to perform regression analyzes on variables other than continuous ones. One such case is when the dependent variable is dichotomous. That is, its value is 0/1, healthy/unhealthy, negative/positive, etc. In the analysis of such data sets, we can examine the relationship between the dependent and independent variable(s) by logistic regression. The independent variables used in the model can also be categorical (dichotomous) or continuous.

slide 8-11

### *With dichotomous independent variable*

The data presented on page 37 were re-analyzed here by logistic regression. Our dichotomous dependent variable was the adverse outcome of pregnancy, and our dichotomous explanatory variable was treatment. As in the case of linear regression, we must emphasize

the interpretation of the coefficient. The estimated coefficient for the explanatory variable means the extent to which an increase in the explanatory variable by one unit increases the occurrence probability of a 1/ unhealthy/positive event. If we take the exponential of the value of the coefficient, we get the estimate of the odds ratio. If our explanatory variable is also dichotomous, then the coefficient expresses the effect of the change between its two values (treated/untreated). In our example, the exponential of 1.1787 is 3.25, which means that a lack of treatment increases the odds of adverse outcomes by 3.25-fold compared to antibiotic treatment.

#### *With continuous independent variable*

In the following example, we see a logistic regression in which the explanatory variable is continuous. Lakos et al. (2012) examined how the odds of seropositivity of Lyme borreliosis in foresters change by age. The resulting coefficient is to be interpreted similarly to the previous ones, i.e., one year of ageing results a 1.02-fold increase of the seropositivity odds.

Multiple explanatory variables can be used in both linear and generalized linear models. They adjust each other's effect, i.e. their coefficients by the joint fitting.

#### *Associations in survival studies*

Survival studies usually do not just analyze one survival curve but compare two or more. Usually, to gain knowledge about which treatment has a more advantageous effect. As an example, we present a vaccination study to prevent respiratory diseases. The cattle were divided into two groups, one receiving a placebo and the other receiving a vaccine. The endpoint examined was the onset of respiratory symptoms, and the time to endpoint was recorded. The survival data of the two groups were plotted with a Kaplan-Meier plot, and their difference was tested by a non-parametric test so-called log-rank test. Based on this result, we can say that the two treatments show a significant ( $p = 0.038$ ) difference. And from examining the curves, we can see that the results were better in the vaccinated group, as that curve placed above the curve in the placebo group at most time points.

In survival studies, another measure, the hazard function is used to quantify risk and risk differences too. The hazard function gives the probability that the event will occur in the next period of time, given that the individual has not yet shown the event. Since the hazard function shows significant variation per time unit, it is not used for plotting, but rather the cumulative hazard. Based on the survival

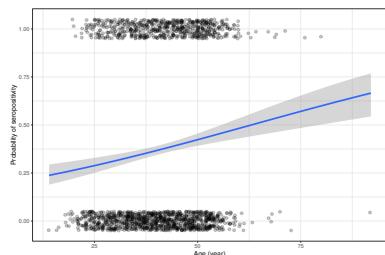


Figure 14: Age dependence of seropositivity in hunters by logistic regression

slide 11

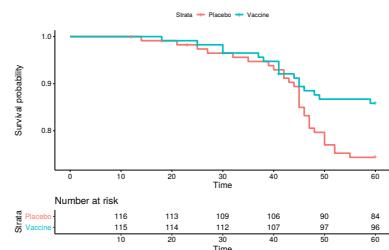


Figure 15: Kaplan-Meier plots of vaccination trial

data obtained in the vaccination study, cumulative hazard curves were plotted by groups. In this form, the hazard values cannot be compared by so-called parametric tests. Therefore, it is conventional to approximate these non-parametric curves with some parametric distribution. If the curves created in this way satisfy certain conditions, they can be compared with parametric tests. From these comparisons, we obtain the hazard ratio.

The hazard ratio (HR) is an expression of the hazard or chance of events occurring in the treatment arm as a ratio of the hazard of the events occurring in the control arm. A hazard ratio of 2 means that a treated patient who has not yet an event by a certain time has twice the chance of having an event at the next point in time compared with someone in the control group. If  $HR = 1$ , then the factor does not affect; if  $HR > 1$ , then the factor is positively associated with the event; if  $HR < 1$ , then the factor is negatively associated with the event. In publications, it is reported with a confidence interval, or in case of several comparisons, those are presented by forest plot.

### *Evidence Based Medicine*

"Evidence based medicine (EBM) is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of EBM means integrating individual clinical expertise with the best available external clinical evidence from systematic research."(Sackett et al., 2000)

The practise of Evidence-Based Veterinary Medicine (EBVM) is the use of best available scientific evidence, in conjunction with clinical expertise and consideration of owner and patient factors, to make the best clinical decisions for patients. "The primary difference between evidence-based medicine and evidence-based veterinary medicine is that, in the latter, the emphasis must be necessarily placed on poorer sources of evidence."(Kastelic, 2006) As we will see, the pieces of evidence come mainly from scientific literature sources. Over recent decades, there have been massive increases in the availability of information, both in the medical and veterinary literature, but also in mainstream media. Which complicates the information filtering of the clinicians. "Clients have access to many of the same resources that veterinary professionals do, but some will lack the clinical knowledge and judgement to assess whether the advice they find online is sensible. They may have attempted diagnosis, and even worse, treatment, before seeking veterinary advice, and the veterinary surgeon now has an important role in educating owners and debunking myths."

The EBVM is, in fact, a structured, well-defined process of knowledge update, helping the veterinarian to base the practical decisions

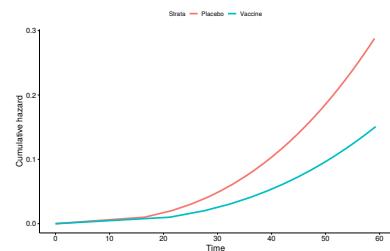
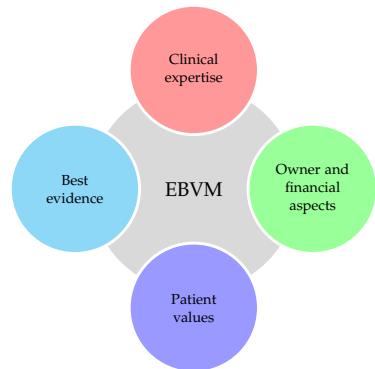


Figure 16: Weibull approximation of vaccination trial cumulative hazard curves

a large extent of the section is an excerpt of <http://www.ebvmlearning.org/> slide 21-54



on the state of the art information. The main elements of the EBVM process are the followings. Ask: defining a clinical question that is of interest and (hopefully!) answerable. Acquire: finding the best available evidence to answer the question. Appraise: assessing the quality of the relevant evidence found. Apply: implementing the evidence into clinical practice where appropriate. Assess: evaluating the impact of the implementation and changes in clinical practice (Fig 17).

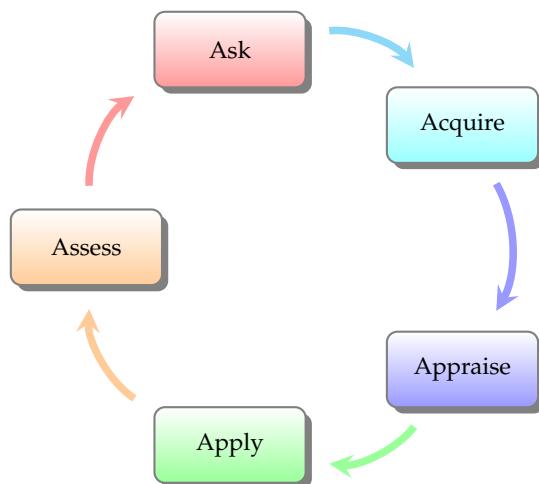


Figure 17: The main elements of the EBVM process

## Ask

Formatting your question correctly is important in ensuring that your search for evidence is structured, systematic and complete. Various question structuring approaches can be applied. The most common way to format a question is to use the PICO system (Fig 18).

Clinical questions can be divided into five main topic areas: Treatment; Prognosis and Incidence; Aetiology and Risk; Diagnosis; Prevalence.

**Questions on treatment** refer to treatment choices a veterinarian would need to make in order to achieve a desirable outcome. These choices can include drugs or medicines to be used, surgical methods, changes in diet or management, and many more. These types of questions are best answered by randomised controlled trials when they are available.

*Example question:* Which diet is best to feed cats with chronic renal disease?

*PICO structured question:* In [P: cats with chronic renal disease] does [I: feeding a renal prescription diet] compared with [C: not feeding a renal prescription diet] impact on [O: survival time]?

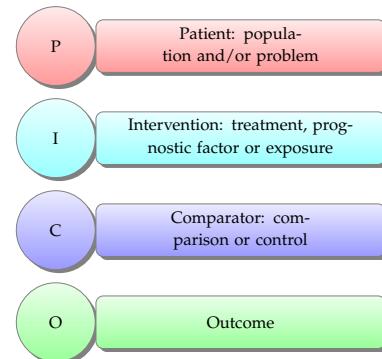


Figure 18: PICO system (<https://pico.vet/>)

**Questions on prognosis and incidence** relate to the likelihood of disease or the progression of disease over time. These questions are best answered by cohort studies.

*Example question:* Does sex affect survival in flat-coat retrievers with cancer?

*PICO structured question:* In [flat-coated retrievers with cutaneous lymphoma], does [being a male] compared with [being a female] affect [average life expectancy]?

**Questions on aetiology and risk** investigate the origin of disease or the factors influencing development of a certain condition or disease. These questions are best answered by cohort studies, case-control studies or cross-sectional studies.

*Example question:* What are the risks of general anaesthesia in ferrets?

*PICO structured question:* In [ferrets], is [general anaesthesia by triple injectable agent] compared with [general anaesthesia by induction and inhalational agent] associated with [an increased risk of death]?

**Questions on diagnosis** involve identification of a disorder based on the animal's presenting signs. These questions are best answered by diagnostic test validation studies (also known as diagnostic evaluation studies).

*Example question:* Which diagnostic test is most reliable for diagnosing fascioliasis in dairy cattle?

*PICO structured question:* In [lactating dairy cattle] does [milk ELISA] compared with [serum ELISA] have [a better sensitivity and specificity for diagnosing fascioliasis]?

**Questions on prevalence** consider the frequency of disease at a certain point in time, and are best answered by cross-sectional studies and surveys.

*Example question:* What is the prevalence of cardiac disorders in Welsh Section A mountain ponies?

*PICO structured question:* In [horses], does [being a Welsh Section A mountain pony] compared with [being any other breed] increase the [prevalence of cardiac disorders]?

## Acquire

Ideally, clinical decisions will incorporate the most current and relevant scientific research, but where is the best place to search for the evidence base for veterinary medicine? Unfortunately, there is no "one-stop-shop", and so a variety of search tools, databases and methods must be used. To be able to track, check, and repeat the results of our acquire, it is important to document it. PRISMA provides a method for that purpose.<sup>28</sup>.

<sup>28</sup> <http://prisma-statement.org/PRISMAStatement/>

## *Appraise*

We must evaluate the collected literature data according to their quality and relevance. It is essential to take into account that different types of studies can provide evidence at different level. Thus, evidence can be hierarchized based on the type of study they come from. This hierarchy is usually represented as a pyramid (Fig 19).

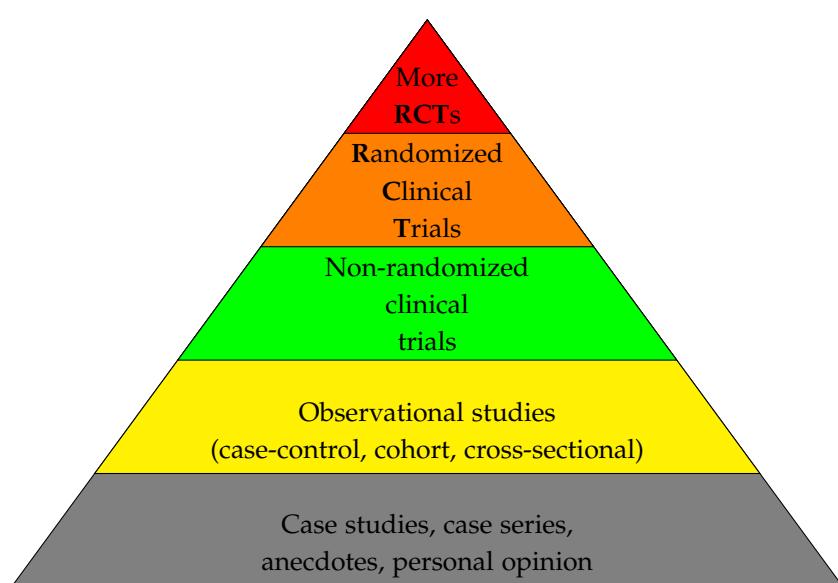


Figure 19: The evidence hierarchy pyramid

As one can see, the least grounded knowledge comes from case studies, case series, anecdotes. In contrast, the most grounded knowledge at the top of the pyramid stems from the synthesis of several RCTs. On slide 41 and 42 there is an example for anecdote and non-randomized clinical trial. The presented hierarchy is just one of the possible orders, but widely applied. Another hierarchy is shown on slide 43.

## *Apply*

Based on the evaluated, filtered information, we modify our previous clinical practice. Applying the new practice, it is necessary to clarify on who, when, what and how we apply it. If you're still not convinced that the evidence fully pertains to you, one way to start putting evidence into practice might be to run a sort of 'pilot' to assess the possible effects of your new approach in the context of your practice.

## *Assess*

The only way we can establish if the care of patients is improved by the application of evidence to practice is to measure the effect of the apply stage. It is vital to assess what we do in practice in order to ensure our practice is moving with the times and adapting and responding to the advances in the profession. A simple way of assessing your own performance as an EBVM practitioner is to ask yourself some questions, and to provide truthful answers! Some suggested questions are:

- Do I identify and prioritise problems to be solved (specifically in relation to what information I need to make my best decisions)?
- Do I perform a competent and complete examination of each animal, in order to establish the likelihood of alternative diagnoses?
- Do I have an accurate knowledge of disease manifestations, the sensitivities and specificities of the clinical signs I am looking for, and the frequency of occurrence of different combinations of clinical signs within a disease?
- Do I search for missing information when I know I am lacking it?
- Do I appraise information I am given in terms of scientific validity?
- Do I understand terms such as specificity and sensitivity, which enable me to interpret important information in my daily practice?
- Do I have the resources to access the Internet and use these to the best of my ability?
- Am I aware of the veterinary information databases?
- Do I actively consider if the application of new information I am given is scientifically justified and sensible for the situation to which I might apply it?
- Do I explain the pros and cons of the different options to owners, taking into account and making clear their different utilities?

## *Evidence synthesis*

As we have seen, the information gathered is synthesized, and usable knowledge can be deduced from its unity. Regarding the collection of evidence, we distinguish between two types of literature review: the so-called narrative reviews and systematic reviews. An example of the former is when we collect literature data for our thesis or other manuscripts. In such cases, we generally do not seek to collect all relevant data. For this reason, they tend to cover a subset of studies based on availability or author selection. It can introduce an element of selection bias.

In contrast, systematic reviews employ standardised and rigorous methodologies to review scientific literature, with a view to minimising bias. They conduct a comprehensive literature search to identify, appraise, and synthesise all the relevant studies on a particular topic. They will formally and openly report (see PRISMA) the sources they use as well as the search strategies used to find those sources so that searches can be peer-reviewed and replicated.

The synthesis of the collected, filtered information can be accomplished in two ways. By meta-synthesis, we compile the data, information into a qualitative summary.

By meta-analysis, the data obtained from various sources are re-analysed by a unified mathematical-statistical method. Various methods (fix, random or mixed effect models) can be found in the literature to perform meta-analyses.

On slide 49-50 an example shows a meta-analysis study on vaccine efficacy to prevent urinary shedding of *Leptospira* serovar Hardjo (Hardjo) in cattle. On slide 49, the PRISMA flow diagram represents the process of data inclusion and exclusion. It is worth observing Sanhueza et al. (2018) collected more than 1500 articles of the topic, but finally, they found only 8 to be analysable correctly.

The results of meta-analyses are usually summarised by forest plot. The estimates of the individual studies included in this analysis (e.g. OR, RR) and the estimated total impact from their combined analysis are presented. The overall effect is usually marked with a diamond, the higher and narrower, the more accurate the estimate.

On slide 51, a foot and mouth disease vaccination example showing a forest plot of the relative risk (RR) of the clinical disease in vaccinated cattle for each of the 13 included studies, the pooled RR (PRR) per virus serotype and the overall PRR together with the 95% CI (Halasa et al., 2011).

The example on slide 52-53 shows our study (Patai et al., 2017) aimed to evaluate all prospective trials published in full text that studied the efficacy of diclofenac or indomethacin and were controlled with placebo or non-treatment for the prevention of PEP in adult patients undergoing ERCP. One can find the PRISMA flow diagram and the forest plot of the results.

The slide 54 useful EBM and EBVM links are listed for further studies on the topic.



# Lecture 7

## Spatial epidemiology

Spatial epidemiology is the branch of epidemiology that deals with the geographical processing and analysis of health-related data. Main areas:

- disease mapping
- spatial pattern analysis
- ecological analysis

## Disease mapping

The purpose of the geographic representation of health-related data is to map the risk. Risk mapping provides significant additional information compared to simple tabular representation of the occurrence data. Several methods are used in epidemiology for disease mapping. The simplest method of mapping is to mark the location of the disease with a dot on the map (Fig 21). Health-related information in most of the cases has point-based geographical reference in its raw form (e.g., GPS coordinate, home address, location). However, in most cases, limited information can be drawn from point-based mapping about the geographical pattern of risk. Impressions based on point maps can be significantly influenced by the resolution used in mapping, the size, shape, and colour of the points. The weakness of point maps is that we cannot display the case number with them, only the positive/negative positions.

By the so-called bubble maps, it is possible to show how many cases were found in a given place or settlement in the given time period (Fig 22). The size of the mark used on the map (in our case, a circle) shows the number of foxes diagnosed with rabies. Here, in addition to the technical factors mentioned in the point maps, another distorting moment is that if there is a settlement in which there have been many cases, the positive place in its neighbourhood might be overlapped by the sign showing the number of cases.

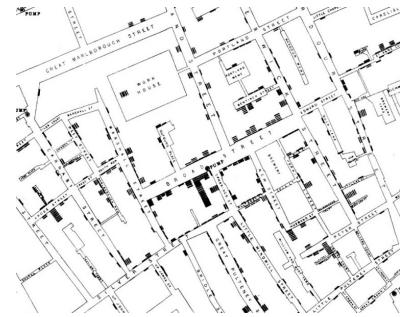


Figure 20: John Snow's cholera map the beginning of spatial epidemiology

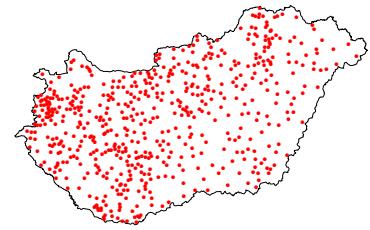


Figure 21: Dot map. The red dots indicate the settlements from which red foxes were diagnosed with rabies during 1990.

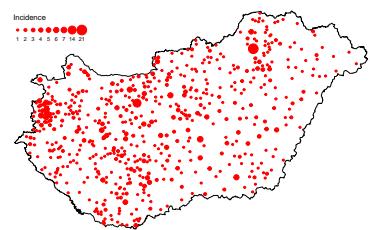


Figure 22: Bubble map of rabies incidences during 1990.

Additional possibilities in risk mapping are the so-called choropleth maps. The name is of Greek origin, khora means the region of the settlement, and plethos means the people. In disease mapping, choropleth maps accordingly represent the risk of occurrence aggregated by geographical units, regions.

The basis for aggregation can be arbitrary, although it can significantly influence our conclusions drawn from the map. The map of Fig 23 shows county-level aggregation of rabies cases during 1990. The same rabies cases aggregated on administrative subregions of Hungary are mapped on Fig 24.

From a technical (Geographic Information System, GIS) point of view, the basic units of aggregation and representation on choropleth maps are polygons. These polygons are in close contact with their neighbours and fill the study area. This avoids overlapping the adjacent data mentioned in the bubble map. Furthermore, this allows us to represent our discontinuous data on a continuous surface, which makes it easier to assess the risk in space. Based on the map, subregions with larger area may appear having a higher risk. Before we accept this, let us think that if the rabies risk of foxes were homogeneous in the country, we would have more cases in larger subregions.

Because reliable population data for foxes are not available, we use the size of the subregions instead of the population to normalize the number of cases (Fig 25). Based on this, it seems that there was a very high incidence in the Szombathely subregion in 1990, and some subregions with moderated risk in Transdanubia and Miskolc environment. The evaliability of choropleth maps can be significantly influenced by the colour scale used painting territorial units, and the choice of value ranges for categories.

Further spatial risk visualization opportunity is the so-called isopleth mapping. When creating isopleth maps, the point-like source data are interpolated in the two-dimensional space by some smoothing method (Fig 26). The smoothing method used to create isopleth maps and its parameterization can greatly influence the risk map, and thus our epidemiological inferences.

Disease mapping procedures may be combined or supplemented with other information (slide 24). Most often, perhaps, maps showing prevalences and relative risks can be found in the literature (slide 25-26).

The risk mapping approaches so far, except the two-dimensional smoothing, are not real spatial models because they do not involve the geographical distance/proximity of events from each other. So far, maps have been used for visualisation only, but their structure did not affect the estimates.

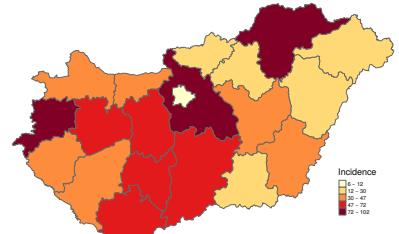


Figure 23: Choropleth map of rabies incidences during 1990 aggregated on county level.

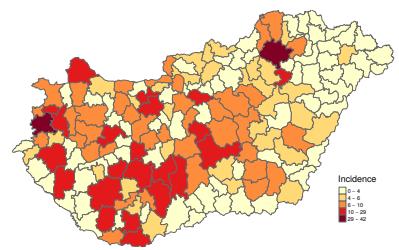


Figure 24: Choropleth map of rabies incidences during 1990 aggregated on sub-region level.



Figure 25: Choropleth map of rabies incidences during 1990 aggregated on sub-region level and normalized by the district area.

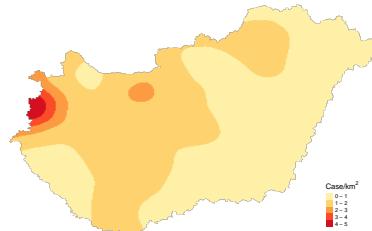
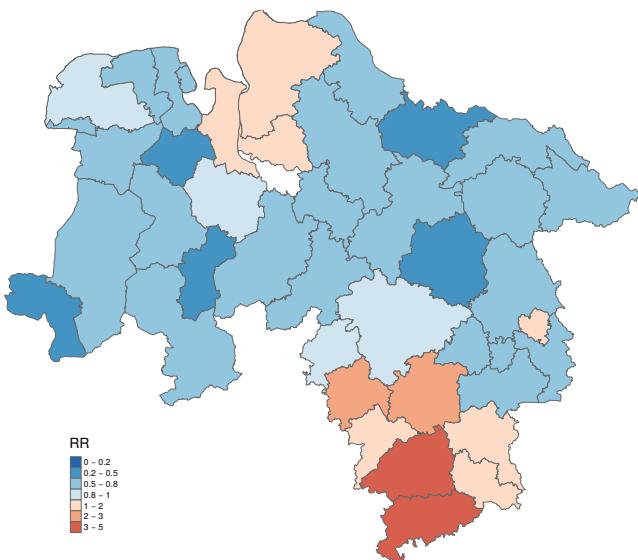


Figure 26: Isopleth map of the data shown on Fig 25.

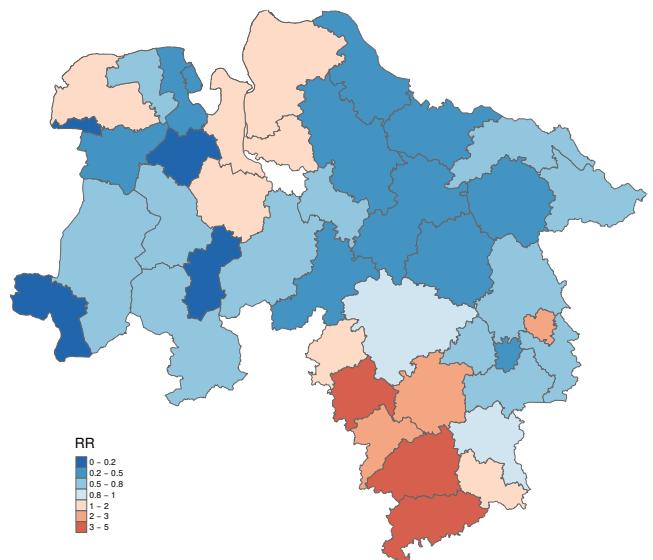
However, it is easy to recognise that the risks of *E. multilocularis* infection observed, at least in the direct neighbourhood, may not be completely independent of each other. Just keep in mind that administrative boundaries, in most cases, do not constitute a physical barrier or match with natural boundaries that would block the movement of foxes.

Thus, the territory of a fox shot in one district may overlap another district, or it may have moved from one district to another for some exceptional reason. Alternatively, the origin of infection may not be in the district in which the case was registered. According to these, our risk estimates may be biased (ecological bias). Fig 28 demonstrates the differences when the model incorporates the neighbourhood effect or does not.

a)



b)



Using more complex models, the risk of district can be broken down into parts attributable to the neighborhood and non-neighbourhood effects (Besag et al., 1991). Example of Besag-York-Mollie model is shown on slide 30-32.

### *Spatial pattern analysis*

From the disease mapping examples, we could see that the map appearance can be strongly influenced by several factors (e.g. point size, colour scale, smoothing procedure, and its parameterization). Therefore, a purely visual assessment of the geographical pattern of risk may lead to subjective conclusions. The same risk map may be in-

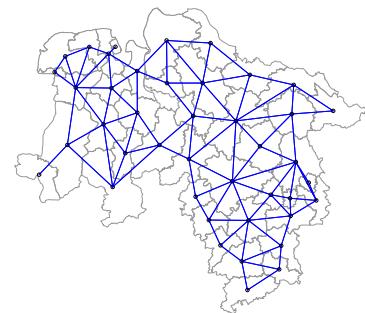


Figure 27: Neighbourhood map. Blue lines between the centres of the polygons indicate that the polygons in the neighbourhood structure are adjacent.

Figure 28: Log-normal model based estimate of relative risk of *E. multilocularis* infection, with (a) and without (b) neighbourhood structure.

terpreted differently by different people. Therefore, several methods have been developed that assess spatial risk more objectively. The spatial epidemiology area for describing the risk quantitatively in space called spatial pattern analysis. In addition to the most commonly used spatial cluster analyzes in epidemiology, barrier detection is presented below.

In geostatistics, three types are usually distinguished for the distribution of points in two-dimensional space: regular, random, and aggregated.<sup>29</sup>

<sup>29</sup> slide 34-36

### *Spatial clustering*

The spatial cluster analysis aims to examine the geographical aggregation, clustering of cases and risks. According to the definition of CDC (1990), a cluster is an apparent or real aggregation of health-related events in space and/or time. Several methods of spatial cluster analysis can be found in the veterinary epidemiological literature (Ward & Carpenter, 2000a,b). Methods are usually grouped according to whether they are point-based or area-based. They are distinguished also as global, local or focused cluster analysis methods. With global cluster analysis methods, we can obtain a test statistic value for the spatial aggregation of cases. This value indicates whether the occurred cases (or risk) in the study region were aggregated in space or not. Local clustering methods make it possible to identify the position of aggregation in space. Using focused clustering methods, we obtain aggregation information for a predefined geographic location.

### *Global clustering*

The two most frequently used methods of global cluster analysis are the autocorrelation based Moran I and Geary c. Autocorrelation methods analyze the neighbourhood relation between the observation units and the similarity in the variable of interest (e.g. risk). Table 14 helps to interpret the results of the two methods. Both global clustering methods give us results<sup>30</sup> supporting spatial clustering of *E. multilocularis* infection. But as these tests are global ones, they can't localize the clusters.

Table 14: Summary of the meaning of Moran I and Geary c test statistic.

Spatial pattern	Geary c	Moran I
Aggregated	$0 \leq c < 1$	$I > 0$
Regular	$1 < c < 3$	$I < 0$
Random	$c = 1$	$I = 0$

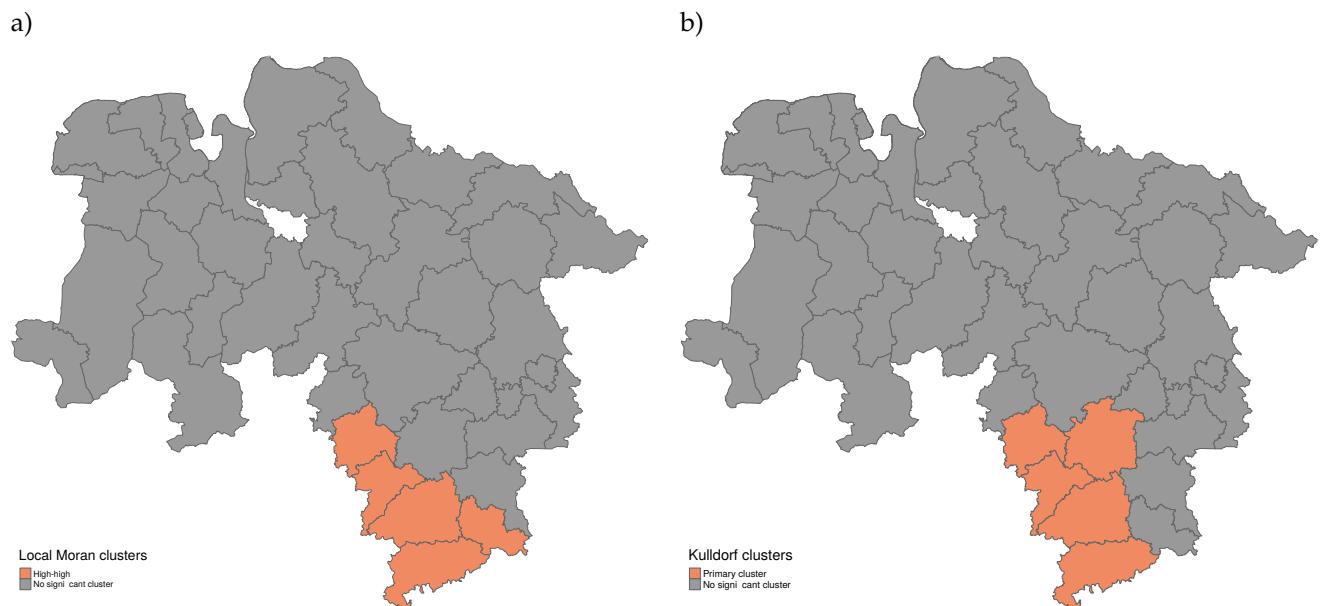
<sup>30</sup> slide 39

### *Local clustering*

While the global Moran I gives a single test statistic value and a p-value for all analysed districts, the local Moran method determines those for each area unit separately (Anselin, 1992). By Moran scatterplot,<sup>31</sup> the per district estimates can be visualised helping to identify the districts forming clusters.

<sup>31</sup> slide 40-42

Numerous so-called spatial scan statistics are used to localize geographic and/or temporal clusters. Perhaps the most commonly used scan statistic in spatial epidemiology is the Kulldorff's one (Kulldorff & Nagarwalla, 1995). We place a grid on the study area, from each grid point as a centre point, we form circles. The radius of the circles is continuously increased until the upper limit (what usually is the half of the whole study area). For each generated circle, we sum the number of cases inside and outside the circle and calculate the number of expected cases inside and outside the circle. The observed case numbers are compared with the expected case numbers based on theoretical distributions, and the circle in which the number of cases is the least likely compared to the out-of-circle cases is identified as the primary cluster.



As can be seen, by the Fig 29, both methods identified south-western districts as an aggregated high-risk area. It is also worth noting that although the results of the two methods overlap greatly, they are not completely identical.

### *Barrier detection*

While cluster analyzes seek to answer the question of whether spatial aggregation of risk can be identified, barrier detection approaches draw boundaries based on differences between neighbours (Solymosi et al., 2005). On slide 46 the barriers were constructed by the method of greatest differences (Monmonier, 1973).

Figure 29: Spatial clusters identified by local Moran (a) and Kulldorff' scan statistic (b).

## *Ecological analysis*

Based on disease mapping and pattern analysis, it seems clear that the risk of infection is much higher in the south-western districts of Lower Saxony than in most of the state. From both the animal and human health viewpoint, it would be important to know where this additional risk may come from. There is a stage in the development of *E. multilocularis* when the egg is directly exposed to external environmental factors. Thus, among many factors influencing the risk of infection, external environmental conditions may play a role too. Based on literature data, the development of the parasite can be influenced by the ambient temperature, the amount of precipitation, and the soil moisture.

To analyze the dependence of the prevalence per district in Lower Saxony on these environmental variables, we obtained those data. On the temperature map (Fig 30) most of the districts with the lowest mean temperature are identical to the districts with the highest prevalence of *E. multilocularis*. According to the precipitation map (Fig 31), the districts with the highest prevalence are not the districts that receive the highest amount of precipitation. But if we also compare it with the temperature map, we can see that in the districts that receive the most precipitation, the average temperature is also in the highest range. The soils of the areas with the highest prevalence belong to the driest type (Fig 32).

So far, we have formulated possible relationships between the risk of infection and some environmental factors on a subjective, visual basis. However, this is only an introduction to the mathematical-statistical description of associations. In this approach, the dependence of a dependent variable (e.g., prevalence) on independent variables (e.g., temperature, precipitation, soil moisture) is formulated using some mathematical model. On the one hand, these models allow us to quantify the size of the effect of independent variables on the dependent variable. On the other hand, the numerical values of these effects allow us to give predictions.

What does prediction mean in spatial epidemiology? On the one hand, it means that our model predicts what values of dependent variables are estimated for the places (e.g., districts) from which we have information about the dependent variable (e.g., prevalence). On the other hand, since we believe that there are rules in the physical world and these rules can be formulated mathematically, we also believe that these rules are not only valid in one region but can be extended to other areas.

Our goal is to describe a model from which the prediction shows the smallest deviation from the observed values. To do this, we create

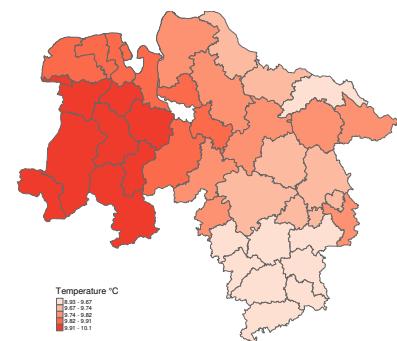


Figure 30: Temperature map.

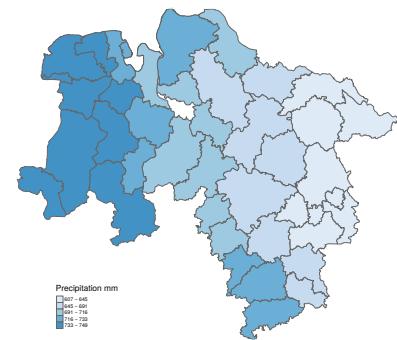


Figure 31: Precipitation map.



Figure 32: Hydrogeological map.

models from our three independent variables (temperature, precipitation, soil moisture) for all possible combinations and fit them. Based on each fitted model, we predict the expected prevalence in each district. Using these predicted prevalences, we calculate the difference between the observed and predicted values for each model. For this, we use the mean absolute error (MAE). The model that resulted in the predicted prevalence with the lowest MAE value is considered as the best.

In addition to the assessment of the global deviation (e.g. MAE) of the prediction results, it is also important to see prediction deviates from the observed values (here the prevalence) per districts as well. For the *E. multilocularis* example that map is in Fig 33.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

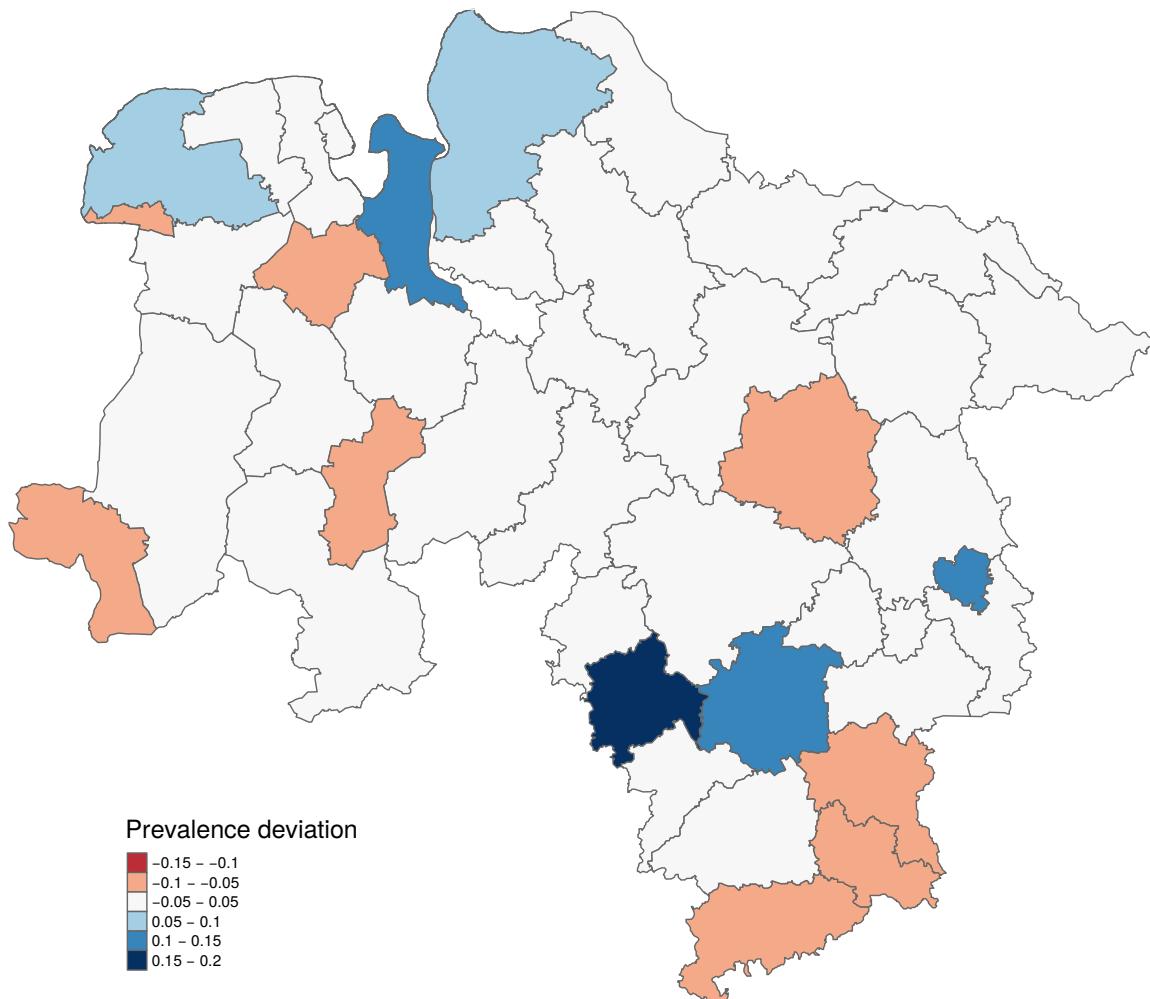


Figure 33: Map of the differences between the predicted and observed prevalences.



## Bibliography

- Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, 567(7748), 305–307.
- Anselin, L. (1992). Local indicators of spatial association – LISA. *Geogr. Anal.*, 27(2), 93–115.
- Bastos, M. L., Tavaziva, G., Abidi, S. K., Campbell, J. R., Haraoui, L.-P., Johnston, J. C., Lan, Z., Law, S., MacLean, E., Trajman, A., et al. (2020). Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis. *BMJ*, 370.
- Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1), 1–20.
- CDC (1990). *Guidelines for investigating clusters of health events*. Technical Report RR-11.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37–46.
- Cuzick, J. (2005). Forest plots and the interpretation of subgroups. *The Lancet*, 365(9467), 1308.
- Del Valle, D. M., Kim-Schulze, S., Huang, H.-H., Beckmann, N. D., Nirenberg, S., Wang, B., Lavin, Y., Swartz, T. H., Madduri, D., Stock, A., et al. (2020). An inflammatory cytokine signature predicts covid-19 severity and survival. *Nature medicine*, (pp. 1–8).
- Dohoo, I., Martin, W., & Stryhn, H. (2002). *Veterinary Epidemiologic Research*. Charlottetown, Prince Edward Island, Canada: VER Inc, 2nd edition.
- Farkas, R., Tánczos, B., Gyurkovszky, M., Földvári, G., Solymosi, N., Edelhofer, R., & Hornok, S. (2013). Serological and molecular detection of Theileria equi infection in horses in Hungary. *Veterinary parasitology*, 192(1-3), 143–148.
- Gardner, I. A. (2012). Analysis and use of diagnostic data. In *Diseases of Swine* (pp. 94–105).: John Wiley & Sons, Ltd.
- Glasziou, P., Burls, A., & Gilbert, R. (2008). Evidence based medicine and the medical curriculum. *BMJ (Clinical Research Ed.)*, 337(7672), 704–705.
- Greenhalgh, T. (2020). Will covid-19 be evidence-based medicine's nemesis? *PLOS Medicine*, 17(6), 1–4.
- Grimes, D. A. & Schulz, K. F. (2005). Refining clinical diagnosis with likelihood ratios. *The Lancet*, 365(9469), 1500–1505.
- Halasa, T., Boklund, A., Cox, S., & Enøe, C. (2011). Meta-analysis on the efficacy of foot-and-mouth disease emergency vaccination. *Preventive veterinary medicine*, 98(1), 1–9.
- Kastelic, J. (2006). Critical evaluation of scientific articles and other sources of information: an introduction to evidence-based veterinary medicine. *Theriogenology*, 66(3), 534–542.
- Kulldorff, M. & Nagarwalla, N. (1995). Spatial disease clusters: Detection and inference. *Statistics in Medicine*, 14, 799–810.

- Lakos, A., Igari, Z., & Solymosi, N. (2012). Recent lesson from a clinical and seroepidemiological survey: low positive predictive value of borrelia burgdorferi antibody testing in a high risk population. *Advances in medical sciences*, 57(2), 356–363.
- Lakos, A. & Solymosi, N. (2010). Maternal lyme borreliosis and pregnancy outcome. *International Journal of Infectious Diseases*, 14(6), e494–e498.
- Mehra, M. R., Desai, S. S., Kuy, S., Henry, T. D., & Patel, A. N. (2020). Cardiovascular disease, drug therapy, and mortality in covid-19. *New England Journal of Medicine*.
- Messam, L. L. M., Branscum, A. J., Collins, M. T., & Gardner, I. A. (2008). Frequentist and Bayesian approaches to prevalence estimation using examples from Johne's disease. *Animal Health Research Reviews*, 9(1), 1–23.
- Monmonier, M. S. (1973). Maximum-difference barriers: An alternative numerical regionalization method. *Geogr. Anal.*, 3, 245–261.
- Nakeshbandi, M., Maini, R., Daniel, P., Rosengarten, S., Parmar, P., Wilson, C., Kim, J. M., Oommen, A., Mecklenburg, M., Salvani, J., et al. (2020). The impact of obesity on COVID-19 complications: a retrospective cohort study. *International Journal of Obesity*, 44(9), 1832–1837.
- Noordhuizen, J. P. T. M., Frankena, K., Thrusfield, M., & Graat, E. A. M. (2001). *Application of Quantitative Methods in Veterinary Epidemiology*. Wageningen, The Netherland: Wageningen Pers.
- Patai, Á., Solymosi, N., Mohácsi, L., & Patai, Á. V. (2017). Indomethacin and diclofenac in the prevention of post-ERCP pancreatitis: a systematic review and meta-analysis of controlled trials. *Gastrointestinal Endoscopy*, 85(6), 1144–1156.
- Pfeiffer, D. (2002). Veterinary epidemiology: An introduction. [ww3.panaftosa.org.br/Comp/MAPA/431857.pdf](http://ww3.panaftosa.org.br/Comp/MAPA/431857.pdf).
- Pfeiffer, D. (2010). *Veterinary Epidemiology: An Introduction*. Oxford, UK: Wiley.
- Pollán, M., Pérez-Gómez, B., Pastor-Barriuso, R., Oteo, J., Hernán, M. A., Pérez-Olmeda, M., Sanmartín, J. L., Fernández-García, A., Cruz, I., de Larrea, N. F., et al. (2020). Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *The Lancet*, 396(10250), 535–544.
- Sackett, D. L., Strauss, Sharon E Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence-based Medicine: How to Practice and Teach EBM*. Churchill Livingstone.
- Sanhueza, J., Wilson, P., Benschop, J., Collins-Emerson, J., & Heuer, C. (2018). Meta-analysis of the efficacy of leptospira serovar hardjo vaccines to prevent urinary shedding in cattle. *Preventive veterinary medicine*, 153, 71–76.
- Smith, R. D. (2005). *Veterinary Clinical Epidemiology: From Patient to Population*. Boca Raton, Florida, USA: CRC Press, 3rd edition.
- Solymosi, N., Reiczigel, J., Harnos, A., Mészáros, J., Molnár, L. D., & Rubel, F. (2005). Finding spatial barriers by Monmonier's algorithm. In *International Society for Clinical Biostatistics Conference Szeged*, Hungary.
- Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation*, volume 13. John Wiley & Sons.
- Stevenson, M. (2012). An introduction to veterinary epidemiology. EpiCentre, IVABS, Massey University, Palmerston North, New Zealand.
- Thrusfield, M., Christley, R., Brown, H., Diggle, P. J., French, N., Howe, K., Kelly, L., O'Connor, A., Sargeant, J., & Wood, H. (2018). *Veterinary Epidemiology*. Oxford, UK: Wiley, 4th edition.
- Ward, M. P. & Carpenter, T. E. (2000a). Analysis of time-space clustering in veterinary epidemiology. *Prev. Vet. Med.*, 43(4), 225–237.

- Ward, M. P. & Carpenter, T. E. (2000b). Techniques for analysis of disease clustering in space and in time in veterinary epidemiology. *Prev. Vet. Med.*, 45, 257–284.
- Woodworth, G. G. (2004). *Biostatistics: A Bayesian Introduction*. Probability and Statistics Series. Hoboken, New Jersey, USA: Wiley & Sons. ISBN 0471468428.