



Trainer: Bondor Cosmina-Ioana, PhD

# Probabilities



ALWAYS



SEEK



KNOWLEDGE

# Objectives

The meaning of the term Probability

Methods of sampling

Probability distribution

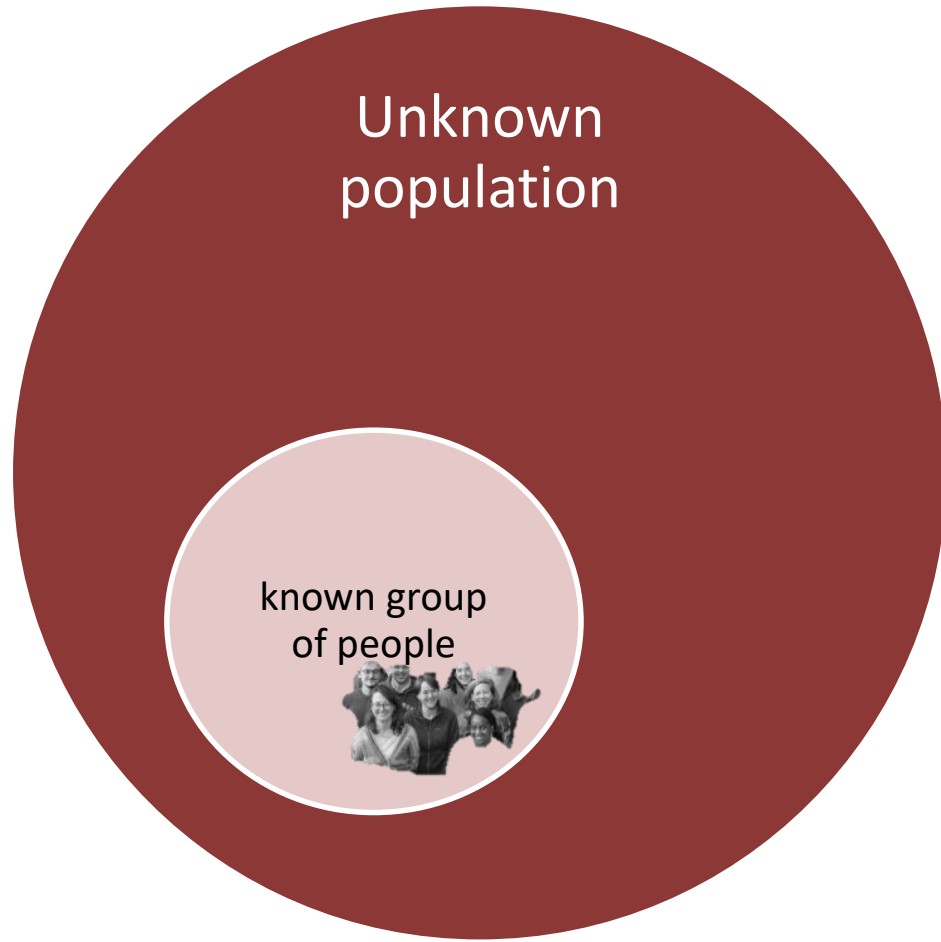
Examples

# Probability – we use this term

- "Today is probably raining"
- "At the party (probably) you will need a vegetarian menu"
- "I'll probably take 10 for the exam"

# Aim of the lecture

- To introduce the basic concepts of statistical inference



# Example 1

- a group of students
  - they are happy because they take all their exams
- However, we don't know if the same outcome will occur for all the students
- the **probability** of this being the case can be calculated.

trial – student give the exam  
possible events – fail/pass

# Example 2

- The police stopped the driver for a quick alcohol test.
- The result was !!! positive
- What is the probability of being wrong?
- the quick test should be confirmed by the blood test

trial – quick alcohol test and standard test from blood  
possible events for quick test – positive/negative  
possible events for blood test – positive/negative



# Example 3

- Alzheimer disease diagnostic “quick” test
  - Magnetic Resonance Imaging (MRI)
- The result was !!! positive
- What is the probability of being wrong?
- the quick test should be confirmed by the autopsy (the standard test)

trial – MRI and autopsy

possible events for MRI – positive to Alzheimer /negative to Alzheimer

possible events for autopsy test – positive to Alzheimer /negative to Alzheimer

# Probability theory

- unpredictable phenomena
- the laws of manifestations
- mass character
- in various areas of interest
  - nature,
  - society,
  - biology,
  - medicine,
  - etc.



the probability of rain



the probability of flu

# Example 3

- The researchers know
  - the distribution of blood types within the population,
  - the frequency of critical accidents.
- They aim to calculate the probability of needing a certain quantity of blood type ABIV in the upcoming month to ensure they have enough available.

# Life sciences

- If you want to know about living beings than you need to use statistics
- Why?
- Living beings are not the same
- Variable characteristics
- Populations

# Probability

- Basic definitions
  - experiment
  - trial
  - Outcome, event
- Rules about how to combine the probabilities of events: AND, OR, NON
- Events: mutually exclusive, impossible, complementary, certain, independents

# Probability

An **experiment** = an action / process repeated many times,  
**random experiment** = with unpredictable **outcome**

**trial** = a repetition of the experiment

Possible outcomes = possible **events** noted with A, B, C etc.

The probability (A) =  $\frac{\text{the number of times that A occurs}}{\text{the total number of trials}}$



# Example 1

# P (death by Ebola virus) = ?

The event = death by ebola virus in 2014 Outbreak

Ebola virus disease (EVD) – acute infectious disease with high mortality risk.

CDC. 2014 Ebola Outbreak in West Africa - Case Counts. Nov.2015. Available at: <http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/case-counts.html>

Country	Cases	Deaths	Survival	<b>Death probability</b>	Survival probability
Siera Leone	14,122	3,955		?	

# P (deaths Ebola virus) = ?

The event = death by ebola virus in 2014 Outbreak

Ebola virus disease (EVD) – acute infectious disease with high mortality

CDC. 2014 Ebola Outbreak in West Africa - Case Counts. Nov.2015. [counts.html](#)

[counts.html](#)

Country	Cases	Deaths			
Siera Leone	14122	3955			

Trial: an individual get infected with Ebola virus  
Possible events: death or survival (which are imprevisible outcomes)  
In Siera Leone there were 14,122 cases of infection, from which 3,955 died

$$P(\text{death Ebola illness}) = \frac{\text{the number deaths}}{\text{the total number of cases}}$$

- trial = a case with the Ebola illness
- event = death in Ebola illness

Country	Cases	Deaths	Survival	Death probability	Survival probability
Siera Leone	<b>14122</b>	<b>3955</b>		<b>=3955/14122</b>	

$$P(\text{death Ebola illness}) = \frac{\text{the number deaths}}{\text{the total number of cases}}$$

- $P(\text{death Ebola illness}) = \frac{3955}{14122} = 0.28$
- The probability that a random person infected with Ebola will die is 0.28 (28% fatality rate)

Country	Cases	Deaths	Survival	Death probability	Survival probability
Siera Leone	14122	3955		<b>0.28</b>	

# Certain event

- an event that is guaranteed to happen
- it occurs in every possible outcome of an experiment.

# Complementary events

- also called opposite events
- two events that together include all possible outcomes, but cannot happen at the same time

# Example 2

Death  $\leftrightarrow$  Survival

– opposite events

No. of deaths + No. of survival = Total

Deaths (%) + Survival (%) = 100 %

$P(\text{survival Ebola virus}) = P(\text{non A}) = ?$

$$P(\text{non } A) = P(\text{survival Ebola virus}) = 1 - P(A)$$

Country	Cases	Deaths	Survival	Death probability	Survival probability
Siera Leone	14122	3955	10167	<b>0.28</b>	<b>=1 - 0.28</b> or <b>=10167/14122</b>

$$P(\text{survival Ebola illness}) = \frac{10167}{14122} = 0.72$$

The probability that a random person infected with Ebola will survive is 0.72 (72% survival rate)

Country	Cases	Deaths	Survivals	Death probability	Survival probability
Siera Leone	14122	3955	10167	<b>0.28</b>	<b>0.72</b>

Deaths (%) + Survival (%) = 100 %

$$P(A) + P(\text{non } A) = 1$$

$$\Rightarrow P(\text{non } A) = 1 - P(A) = 1 - 0.28 = 0.72$$

Country	Cases	Deaths	Survivals	Death probability	Survival probability
Siera Leone	14122	3955	10167	<b>0.28</b>	<b>0.72</b>

# Home Exercises

Deaths (%) + Survival (%) = 100 %

$$P(A) + P(\text{non } A) = 1$$

Country	Cases	Deaths	Survivals	Death probability	Survival probability
Guinea	3805	2536	?	?	?
Siera Leone	14122	3955	10167	0.28	0.72
Liberia	10672	4808	?	?	?
Others	36	15	?	?	?
Total	28635	11314	?	?	?

- A,B mutually exclusive events

$$P(A \text{ and } B) = 0$$

- B the **complementary** event of A

$$P(\text{complementary event of } A) = P(\text{non } A) = 1 - P(A)$$

- A the **certain** event

$$P(\text{certain event}) = 1$$

- A the **impossible** event

$$P(\text{impossible event}) = 0$$



# Impossible event

- Is an event that cannot occur at any time

# Mutually exclusive events

- events that **cannot happen at the same time.**

# Example 3

$$P(\text{A and B}) = P(\text{Type O and Type A}) = ?$$

Blood type	Frequency	P(Blood type)
O	400	0.40
A	450	0.45
B	142	0.142
AB	8	0.008
<b>Total</b>	1000	1.0

$P(A \text{ and } B) = 0$  for mutually exclusive events

Blood type	Frequency	P(Blood type)
O	400	0.40
A	450	0.45
B	142	0.142
AB	8	0.008
<b>Total</b>	1000	1.0

$P(\text{Type 0 and Type A}) = ?$

Type 0 and Type A are mutually exclusive:

**$P(\text{Type 0 and Type A}) = 0$**

Two events are mutually exclusive = they can't happen at the same time

# Home exercises

Blood type	Frequency	P(Blood type)
O	400	0.40
A	450	0.45
B	142	0.142
AB	8	0.008
<b>Total</b>	1000	1.0

$P(\text{Type AB and Type A}) = ?$

$P(\text{Type AB and Type B}) = ?$

- **A,B mutually exclusive** events

$$P(A \text{ and } B) = 0$$

- B the **complementary** event of A

$$P(\text{complementary event of } A) = P(\text{non } A) = 1 - P(A)$$

- A the **certain** event

$$P(\text{certain event}) = 1$$

- A the **impossible** event

$$P(\text{impossible event}) = 0$$



## Events A and B:

- **reunion**  $C=A\cup B$  - an event that occurs if at least one of events A **or** B takes place
- **intersection**  $D=A\cap B$  - the event that occurs only when A **and** B occur simultaneously
- The opposite (complementary) of an event A (non A) - an event that occurs whenever there is not happening A
- Incompatible A and B events (mutually exclusive, disjunctive) - cannot occurred simultaneously ( $A\cap B = \emptyset$ )
- Compatible events A and B - can occur simultaneously.

# Example 4

$$P(\text{A or B}) = P(\text{Type O or Type A}) = ?$$

Blood type	Frequency	P(Blood type)
O	400	0.40
A	450	0.45
B	142	0.142
AB	8	0.008
<b>Total</b>	1000	1.0

$P(A \text{ or } B) = P(A) + P(B)$  for mutually exclusive event

Blood type	Frequency	P(Blood type)
O	400	0.40
A	450	0.45
B	142	0.142
AB	8	0.008
<b>Total</b>	<b>1000</b>	<b>1.0</b>

$$\begin{aligned} P(\text{A person have type O or A}) &= \\ &= P(\text{Type O or Type A}) = P(\text{Type O}) + P(\text{Type A}) = \\ &= \frac{400}{1000} + \frac{450}{1000} = \frac{950}{1000} = 0.95 \end{aligned}$$

Two events are mutually exclusive if they can't happen at the same time

A, B two events:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

A, B two independent events:

$$P(A \text{ and } B) = P(A) * P(B)$$

A, B two independent events:

$$P(A \text{ or } B) = P(A) + P(B) - P(A) * P(B)$$

A, B two mutually exclusive events:

$$P(A \text{ or } B) = P(A) + P(B)$$



# Independent events

- events where the occurrence (or non-occurrence) of one **does not affect** the probability of the other

- A **included** in B ( $A \subseteq B$ ) - the B set contains the A set
- A **dependent** on B ( $A | B$ ) - whether A's occurrences is influenced (depends on) B's occurrence.
- A and B are **independent** - whether the occurrence of one or the other does not depend on whether or not the other event occurred.

# Example 5

$$P(\text{A and B}) = P(\text{Boy and Boy}) = ?$$

Families with 2 children

Birth	Frequency		Family with two children to be boys	P("and")	Frequency	Probability
Girl	500	$P(\text{A=girl}) = 0.5$ $P(\text{B=boy}) = 0.5$	Two girls	$P(\text{A and A})$	250	$0.5 * 0.5 = 0.25$
Boy	500		Two boys	$P(\text{B and B})$	250	$0.5 * 0.5 = 0.25$
			One girl, one boy	$P(\text{A and B})$	250	$0.5 * 0.5 = 0.25$
			One boy, one girl	$P(\text{B and A})$	250	$0.5 * 0.5 = 0.25$
<b>Total</b>	1000				1000	

$$P(\text{A and B}) = P(\text{Boy and Boy}) = ?$$

Families with 2 children

Birth	Frequency		Family with two children to be boys		Frequency	Probability
Girl	500	$P(\text{A=girl}) = 0.5$	Two girls	$P(\text{A and A})$	250	$0.5 * 0.5 = 0.25$
Boy	500	$P(\text{B=boy}) = 0.5$	Two boys	$P(\text{B and B})$	250	$0.5 * 0.5 = 0.25$
			One girl, one boy	$P(\text{A and B})$	250	$0.5 * 0.5 = 0.25$
			One boy, one girl	$P(\text{B and A})$	250	$0.5 * 0.5 = 0.25$
<b>Total</b>	1000				1000	

A, B two events:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

A, B two independent events:

$$P(A \text{ and } B) = P(A) * P(B)$$

A, B two independent events:

$$P(A \text{ or } B) = P(A) + P(B) - P(A) * P(B)$$

A, B two mutually exclusive events:

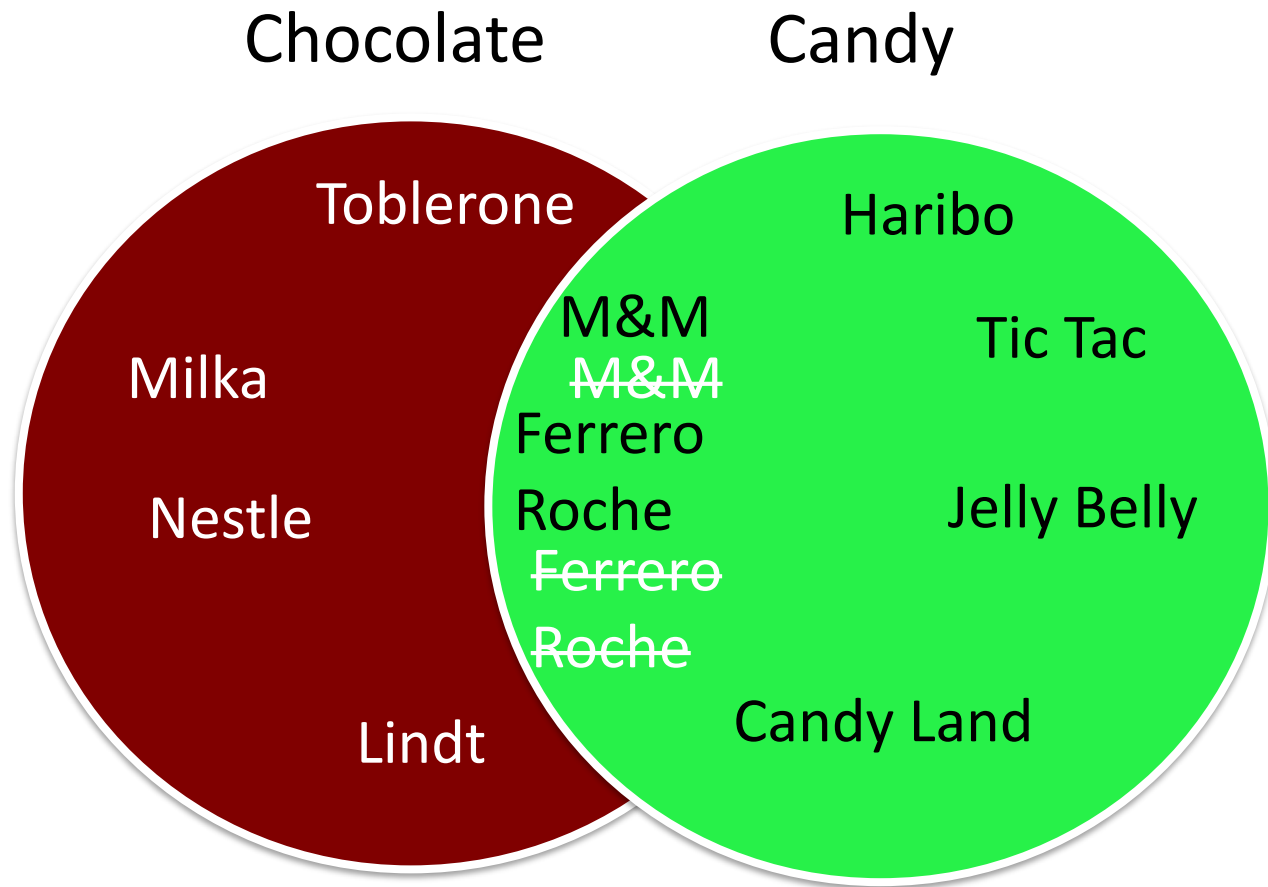
$$P(A \text{ or } B) = P(A) + P(B)$$



# Example 6

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

for any events A, B



$$\begin{aligned} P(A \text{ or } B) &= \\ &= P(\text{chocolate or candy}) = \\ &= P(A) + P(B) - P(A \text{ and } B) \end{aligned}$$

A, B two events:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

A, B two independent events:

$$P(A \text{ and } B) = P(A) * P(B)$$

A, B two independent events:

$$P(A \text{ or } B) = P(A) + P(B) - P(A) * P(B)$$

A, B two mutually exclusive events:

$$P(A \text{ or } B) = P(A) + P(B)$$



# Probability

- Theoretical
  - how it should be
- Empirical
  - how it is if we experiment
- Subjective
  - an individual's opinion on how it should be
- Example: gender
  - 0.5 is the theoretical probability of a male birth (it means 50%)
  - 0.494 is the empirical probability of a male birth in 2023 in Romania

$$P(\text{death Ebola illness}) = \frac{\text{the number deaths}}{\text{the total number of cases}}$$

Empirical probability of death by Ebola virus

Country	Cases	Deaths	Survival	Death probability	Survival probability
Siera Leone	14122	3955		<b>0.28</b>	

# Probability

- Theoretical
  - how should be
- Empirical
  - how it is if we experiment
- Subjective
  - the opinion of a person about how it should be
- Example: gender
  - 0.5 is the theoretical probability of a male birth (it means 50%)
  - 0.494 is the empirical probability of a male birth in 2023 in Romania

At birth there are two gender: male and female. We expect that male birth probability to be 0.5 (a half from all births) = the theoretical probability. In practice is near 0.5. Each year the number of male births from the total number of births is different.

# Summary

- A,B **mutually exclusive** events

$$P(A \text{ and } B)=0$$

- B the **complementary** event of A

$$P(\text{complementary event of } A) = P(\text{non } A) = 1 - P(A)$$

- A the **certain** event

$$P(\text{certain event}) = 1$$

- A the **impossible** event

$$P(\text{impossible event}) = 0$$



# Summary

A, B two events:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

A, B two independent events:

$$P(A \text{ and } B) = P(A) * P(B)$$

A, B two independent events:

$$P(A \text{ or } B) = P(A) + P(B) - P(A) * P(B)$$

A, B two mutually exclusive events:

$$P(A \text{ or } B) = P(A) + P(B)$$



# Dependent events

event B dependent on event A

– occurrence of B influenced by the occurrence of A

$$P(\text{B **dependent** on A}) = P(B|A)$$



# Dependent events

event B dependent on event A

$$P(A \text{ and } B) = P(B|A) * P(A)$$



# $P(A | B)$ = conditional probability

- The probability of occurrence of event A given that event B is already happen

$$P(A | B) = P(A \text{ and } B) / P(B)$$



# Diagnostic test

Link between the disease and the diagnostic test is a probability

# Quick test to detect alcohol from breath

	Blood test (positive)	Blood test (negative)	Total
Breath test (positive)	20	80	100
Breath test (negative)	5	895	900
Total	25	975	1000

Driving under the influence of alcohol – the “disease”

blood test – Golden standard diagnostic test

positive =under influence – 25

negative =no influence – 975

Quick diagnostic – from breath

positive – 100

negative – 900

20 – true positive (TP)

blood test positive

breath test positive

5 – false negative (FN)

blood test positive

breath test negative

80 – false positive (FP)

blood test negative

breath test positive

895 – true negative (TN)

blood test negative

breath test negative

# Quick test to detect alcohol from breath

	Blood test (positive)	Blood test (negative)	Total
Breath test (positive)	20	80	100
Breath test (negative)	5	895	900
Total	25	975	1000

Driving under the influence of alcohol – the “disease”

From 100 of positive quick test 20 are correct = 20%

From 900 of negative quick test 895 are correct = 99%

# Quick test to detect alcohol from breath

	Blood test (positive)	Blood test (negative)	Total
Breath test (positive)	20	80	100
Breath test (negative)	5	895	900
Total	25	975	1000

Driving under the influence of alcohol – the “disease”

From 100 of positive quick test 20 are correct = 20%

From 900 of negative quick test 895 are correct = 99%

From 25 of people who drank alcohol 20 were detected by the quick test = 80%

From 975 of people who drank alcohol 895 were detected by the quick test = 92%

## Either

- B - the event that a person has a certain disease B
  - for example diabetes, HIV, etc.
- T - the event of obtaining a positive test
  - in case of applying a diagnostic test T to detect the disease B
- complementary events to events B and T respectively
  - non (B) (person without condition B)
  - non (T) (person with negative test)

# Diagnostic test

- Golden standard – the diagnostic of the disease
  - positive – the person have the disease
  - negative – the person didn't have the disease
- Sometimes a **new test** for the disease it's desirable because
  - the golden standard diagnostic
    - takes too long
    - invasive
    - it's expensive
    - it cannot be perform

# Diagnostic test

- New test
  - positive
    - correct (the golden standard diagnostic is also positive)
    - incorrect (the golden standard diagnostic is negative)
  - negative
    - correct (the golden standard diagnostic is also negative)
    - incorrect (the golden standard diagnostic is positive)

- Test applied to n persons:

<b>Disease / Test</b>	<b>B With illness</b>	<b>non(B) Without illness</b>	<b>Total</b>
<b>T Positive test</b>	<b>a (TP)</b>	<b>b (FP)</b>	<b>a+b</b>
<b>non (T) Negative test</b>	<b>c (FN)</b>	<b>d (TN)</b>	<b>c+d</b>
<b>Total</b>	<b>a+c</b>	<b>b+d</b>	<b>n</b>

a – true positive

b – false positive

c – false negative

d – true negative

a+b – Positive test

c+d – Negative test

a+c – persons with illness

b+d – persons without illness

Disease / Test	B With illness	non(B) Without illness	Total
T Positive test	a (TP)	b (FP)	a+b
non (T) Negative test	c (FN)	d (TN)	c+d
Total	a+c	b+d	n

# Positive predictive value PPV

- Probability of a positive test to be true (indicate the disease):

$$PPV = Pr(B|T) = \frac{Pr(B \cap T)}{Pr(T)} = \frac{\frac{AP}{n}}{\frac{AP + FP}{n}} = \frac{AP}{AP + FP} = \frac{a}{a + b}$$

Disease / Test	B With illness	non(B) Without illness	Total
T Positive test	a (TP)	b (FP)	a+b
non (T) Negative test	c (FN)	d (TN)	c+d
Total	a+c	b+d	n

# Negative predictive value NPV

- Probability of a negative test to be true (indicate no disease):

$$NPV = \Pr(\text{non}B / \text{non}T) = \frac{\Pr(\text{non}B \cap \text{non}T)}{\Pr(\text{non}T)} = \frac{AN}{FN + AN} = \frac{d}{c + d}$$

Disease / Test	B With illness	non(B) Without illness	Total
T Positive test	a (TP)	b (FP)	a+b
non (T) Negative test	c (FN)	d (TN)	c+d
Total	a+c	b+d	n

# Test sensitivity $Se$

- Probability of a people with disease to get a positive test:

$$Se = \Pr(T / B) = \frac{AP}{AP + FN} = \frac{a}{a + c} = \frac{\frac{a}{n}}{\frac{a + c}{n}} = \frac{\Pr(T \cap B)}{\Pr(B)}$$

Disease / Test	B With illness	non(B) Without illness	Total
T Positive test	a (TP)	b (FP)	a+b
non (T) Negative test	c (FN)	d (TN)	c+d
Total	a+c	b+d	n

# Test specificity Sp

- Probability of a people without disease to get a negative test:

$$Sp = \Pr(\text{non}(T) / \text{non}(B)) = \frac{AN}{FP + AN} = \frac{d}{b + d} = \frac{\frac{d}{n}}{\frac{b + d}{n}} = \frac{\Pr(\text{non}T \cap \text{non}B)}{\Pr(\text{non}B)}$$

Disease / Test	B With illness	non(B) Without illness	Total
T Positive test	a (TP)	b (FP)	a+b
non (T) Negative test	c (FN)	d (TN)	c+d
Total	a+c	b+d	n

# Interpretation

- sensitivity and/or specificity
  - 0.90 – 1.00 is good to excellent
  - 0.80 – 0.89 is acceptable
  - 0.70 – 0.80 is poor to acceptable
  - 0.60 – 0.70 is poor
  - 0.50 – 0.60 near tossing a coin
  - 0.5 – tossing a coin
  - < 0.50 – the opposite will have a better Se or Sp!!!

# Quick test to detect alcohol from breath



	Blood test (positive)	Blood test (negative)	Total
Breath test (positive)	20	80	100
Breath test (negative)	5	895	900
Total	25	975	1000

Driving under the influence of alcohol – the “disease”

!!! Possible interpretation mistakes to confuse sensitivity (80%) for the positive predictive value (20%). PPV depends on the prevalence of the disease: when the prevalence is very low, even a highly accurate test will produce a relatively low post-test probability for a positive result.

Positive predictive value = From 100 of positive quick test 20 are correct = 20%

Negative predictive value = From 900 of negative quick test 895 are correct = 99%

Sensitivity = From 25 of people who drank alcohol 20 were detected by the quick test = 80%

Specificity = From 975 of people who drank alcohol 895 were detected by the quick test = 92%

# Magnetic resonance imaging (MRI) for Alzheimer

	With Alzheimer (B)	Without Alzheimer (non B)	Total
MRI positive (T)	80	120	200
MRI negative (non T)	20	99,880	99,900
Total	100	100,000	100,100

**Sensitivity** of the MRI test =  $P(T|B) = P(T \text{ and } B) / P(B) = (80/100) = 0.8$  (80%)

**Specificity** of the MRI test =  $P(\text{non } T|\text{non } B) = P(\text{non } T \text{ and non } B) / P(\text{non } B) = (99,880/100,000) = 0.9988$  (99.88%)

**Positive predictive value** of the MRI test =  $P(B|T) = P(T \text{ and } B) / P(T) = 80/200 = 0.4$  (40%)

**Negative predictive value** of the MRI test =  $P(\text{non } B|\text{non } T) = P(\text{non } T \text{ and non } B) / P(\text{non } T) = (99,880/99,900) = 0.9997$  (99.97%)

# HIV Test

	With HIV (B)	Without HIV (non B)	Total
Test positive (T)	100	100	200
Test negative (non T)	1	99.999	100.000
Total	101	100.099	100.200

**Sensitivity** =  $P(T|B) = (100/101) = 0.99$  (99%)

**Specificity** =  $P(\text{non T}|\text{non B}) = (99.999/100.099) = 0.999$  (99.9%)

**Positive predictive value** =  $P(B|T) = (100/200) = 0.5$  (50%) !!! Possible interpretation mistakes

**Negative predictive value** =  $P(\text{non B}|\text{non T}) = (99.999/100.000) = 0.9999$  (99.99%)



# Questions at theoretical exam - Example

\*In a dental office 10,000 people come for an intervention. The doctors mark the situation of each intervention in the patient file. The number of extracted left upper canine was 1500. Compute the probability of extracted left upper canine.

A. 85

B. 15

C. 0.15

D. 8500

E. 0.85

# Questions at theoretical exam - Example

\*In a dental office 10,000 people come for an intervention. The doctors mark the situation of each intervention in the patient file. The number of extracted left upper canine was 1500. The number of extracted right upper canine was 500. Compute the probability of extracted left upper canine or right upper canine if we suppose that this events are independent.

A. 1500

**B. 0.1925**

C. 0.1500

D. 19.25%

E. 0.05

# Questions at theoretical exam - Example

\*In a dental office 10,000 people come for an intervention. The doctors mark the situation of each intervention in the patient file. The number of extracted left upper canine was 1500. The number of extracted right upper canine was 500. Compute the probability of extracted left upper canine and right upper canine if we suppose that this events are independent.

A. 1500

**B. 0.0075**

C. 0.1500

D. 0.75%

E. 0.1925

# Questions at theoretical exam - Example

Mirela had frequent episodes of allergy when she was a child, when she presented rhinorrhea (nasal secretions). She is now a teenager. The probability of an individual to have at least one respiratory tract episode this winter season is 0.80. Symptoms of respiratory viruses include coughing, rhinorrhea and sore throat. If Mirela has a respiratory episode this season, which of the following statements do we know are true, given her history of allergy episodes?

- A. It is more likely to have “cough” or “nasal discharge” than “cough and nasal discharge”
- B. “Cough and nasal discharge” are more likely than just “coughing”
- C. “Cough and nasal discharge” are more likely than “cough, nasal discharge and sore throat”
- D. It is more likely to present only “coughs” than “coughs and nasal secretions”
- E. “Cough and nasal discharge” are more likely than “cough or nasal discharge”

Name of the  
revue

CLINICAL PRACTICE

# Example from scientific literature

## A Closer Look at Diagnosis in Clinical Dental Practice: Part 1. Reliability, Validity, Specificity and Sensitivity of Diagnostic Procedures

Title

• Iain A. Pretty, BDS(Hons), MSc, PhD •  
• Gerardo Maupomé, PhD •

Authors

### A b s t r a c t

*Dentists are involved in diagnosing disease in every aspect of their clinical practice. A range of tests, systems, guides and equipment — which can be generally referred to as diagnostic procedures — are available to aid in diagnostic decision making. In this era of evidence-based dentistry, and given the increasing demand for diagnostic accuracy and properly targeted health care, it is important to assess the value of such diagnostic procedures. Doing so allows dentists to weight appropriately the information these procedures supply, to purchase new equipment if it proves more reliable than existing equipment or even to discard a commonly used procedure if it is shown to be unreliable. This article, the first in a 6-part series, defines several concepts used to express the usefulness of diagnostic procedures, including reliability and validity, and describes some of their operating characteristics (statistical measures of performance), in particular, specificity and sensitivity. Subsequent articles in the series will discuss the value of diagnostic procedures used in daily dental practice and will compare today's most innovative procedures with established methods.*

**MeSH Key Words:** *decision support techniques; predictive value of tests; risk assessment/methods*

© J Can Dent Assoc 2004; 70(4):251-5  
This article has been peer reviewed.

**T**he need for cost-effective treatments is becoming increasingly important in resource-conscious health care systems. Yet appropriate treatment depends on

she can give appropriate weight to the result<sup>1</sup> in clinical decision making.<sup>2</sup> An objective assessment of a given diagnostic procedure would ascertain the reliability and validity

the 2 clinicians agree in their assessments. How can this level of agreement be further quantified?

A simple index would be the proportion of agreements between the 2 observers: 21/29 (i.e., there were 21 agreements out of 29 decisions) = 0.724, or 72.4% agreement. However, this measure ignores the agreement that would have occurred purely by chance. To correct for this chance agreement, Cohen's kappa statistic is used. While it is theoretically possible to achieve a negative value for kappa, the values normally fall between 0 (no agreement beyond chance alone) to 1 (perfect agreement). Landis and Koch<sup>13</sup>

*Journal of the Canadian Dental Association*

April 2004, Vol. 70, No. 4

253

Notably, a procedure can be accurate (i.e., no systematic error) without being valid, but it cannot be valid if it is inaccurate.

### Sensitivity and Specificity

Sensitivity and specificity are 2 of the operating characteristics that indicate the accuracy of a diagnostic procedure, i.e., its ability to correctly identify those individuals with and those without the disease or condition of interest.

A typical diagnostic situation allows for 2 outcomes: either the person has or does not have the disease.<sup>1</sup> When the results of a procedure are compared with those of a gold

# From the same article

About specificity and sensitivity

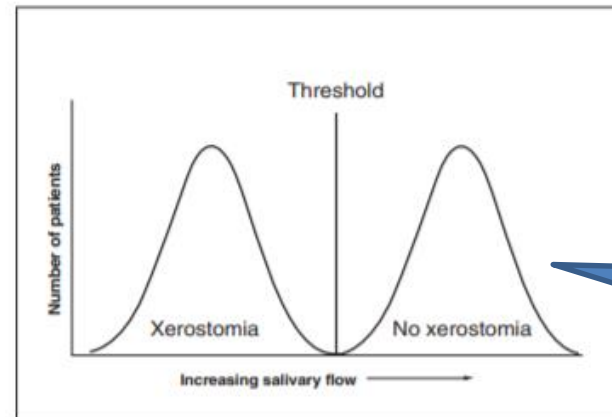
Year of publication, volume number from the start, volume number in the same year

Pretty, Maupomé

**Table 4** A 2 × 2 contingency table illustrating the outcomes of a comparison between a diagnostic or management procedure and a gold standard

	Gold standard result		
	Positive	Negative	Total
Procedure result			
Positive	True positive (TP)	False positive (FP)	TP + FP
Negative	False negative (FN)	True negative (TN)	FN + TN
Total	TP + FN	FP + TN	TP + FN + FP + TN

standard (either an established clinical procedure such as radiography for caries or a confirmatory test such as examination of histologic sections for caries). there are 4 possible



**Figure 1:** Probability distributions of results for a hypothetical perfect diagnostic procedure. This procedure would correctly identify all those with and without the disease or condition, and therefore its specificity and sensitivity are both 100%. In reality, such a situation occurs only when the disease is so obvious, gross or advanced that a diagnostic procedure is not required.

Probability distribution will be presented in the next course

# Article indexed in PubMed – medical article database

The image shows a screenshot of a PubMed article page. Several blue callout bubbles are overlaid on the page to highlight specific information:

- Top right:** "Sign in to NCBI" link.
- Search bar:** "PubMed" dropdown menu and "Advanced" search options.
- Format:** "Format: Abstract" dropdown menu.
- Metadata:** "J Endod. 2018 May;44(5):694-702. doi: 10.1016/j.joen.2018.01.021. Epub 2018 Mar 20." Callout: "Name of the journal, Year of publication, volume number from the start, volume number in the same year".
- Title:** "Diagnostic Accuracy of 5 Dental Pulp Tests: A Systematic Review and Meta-analysis." Callout: "Title, authors".
- Authors:** "Mainkar A<sup>1</sup>, Kim SG<sup>2</sup>".
- Abstract:** "INTRODUCTION: The aim of this systematic review was to investigate and compare the diagnostic accuracy including sensitivity, specificity, adjusted accuracy, adjusted positive predictive value (PPV), and adjusted negative predictive value (NPV) of cold pulp testing (CPT), heat pulp testing (HPT), electric pulp testing (EPT), laser Doppler flowmetry (LDF), and pulse oximetry (PO). METHODS: Three electronic databases were searched from January 1964 to December 2016. True-positive, false-positive, true-negative, and false-negative values were extracted from data in each study. Sensitivity, specificity, adjusted accuracy, adjusted PPV, and adjusted NPV were calculated from those values, if not presented. A random effects model was used to calculate pooled estimates of sensitivity, specificity, adjusted accuracy, adjusted PPV, and adjusted NPV. RESULTS: A total of 125 articles were identified, and 28 studies were included for the final review. The pooled estimates of sensitivity for CPT, EPT, HPT, LDF, and PO were 0.87, 0.72, 0.78, 0.98, and 0.97, respectively. Those of specificity were 0.84, 0.93, 0.67, 0.95, and 0.95, respectively. Those of adjusted accuracy were 0.84, 0.82, 0.72, 0.97, and 0.97, respectively. For adjusted PPV, they were 0.81, 0.89, 0.62, 0.94, and 0.94, respectively, and for adjusted NPV, they were 0.87, 0.80, 0.79, 1.00, and 0.99, respectively. CONCLUSIONS: LDF and PO were the most accurate diagnostic methods, and HPT was the least accurate diagnostic method. EPT showed high accuracy when testing vital teeth (specificity = 0.93) but low accuracy when assessing nonvital teeth (sensitivity = 0.72). CPT had moderate accuracy when evaluating vital (specificity = 0.84) and nonvital (sensitivity = 0.87) teeth." Callout: "About specificity and sensitivity".
- Full text links:** "Full text links" section with an "ELSEVIER FULL-TEXT ARTICLE" button. Callout: "Full text link".
- Similar articles:** "Similar articles" section listing related research.
- Footer:** "Copyright © 2018 American Association of Endodontists. Published by Elsevier Inc. All rights reserved." and "KEYWORDS: Cold test; electric pulp test; heat test; laser Doppler flowmetry; pulp test; pulse oximetry".

Thank you!