

Clinical and statistical interpretations for solving the proposed practical activities

Variable types	6
Qualitative (categorical).....	6
Quantitative	6
Survival.....	6
Descriptive statistics	6
How to choose the chart/table to describe one variable	6
Details	6
Qualitative (categorical) data	6
Frequency tables and Pie charts	6
Quantitative data	7
Mean, standard deviation (SD)	7
95% confidence interval for means	7
Histograms and frequency tables for quantitative data.....	7
Survival data.....	8
Median of survival time	8
Contingency table between grouping variable and censure status	8
Kaplan-Meier survival probability charts	8
Analysis of potential relations between two variables, using tables or charts	9
How to choose a table/chart for analyzing the relation between two variables	9
Details	9
Qualitative (categorical) data	9
Contingency tables and the associated proportional charts	9
Quantitative data	9
Scatter charts	9

Survival data.....	10
Kaplan-Meier survival probability charts	10
Statistical tests	10
How to choose a statistical test (the most common used ones).....	10
Choice between statistical test on independent samples and statistical test on dependent samples (pairs)	11
Interpretation	11
Confidence intervals	12
Statistical interpretation	12
The precision of the results.....	12
Clinical interpretation	12
Assessing if the result is statistically significant ($p < 0.05$) by looking at the confidence interval	12
Domains of study/Study types/ Data collection types	12
Domain of research.....	12
Study type	13
Based on study objectives.....	13
Based on the researchers role	13
Using the matching technique in studies comparing two groups.	13
Type of therapeutic study.....	13
Data collection method.....	14
Based on the studied population.....	14
Based on the duration of data collection	14
Based on the grouping method	15
Inferential Statistics (Data Analysis)	15
Case-control studies.....	15
Evaluation of a potentially significant link between a risk factor and a disease	15
Chi-square (χ^2) test	15
Quantifying the importance of a link between a risk factor and a disease	16

Quantifying the importance of the relationship - for studies of risk factors– how to choose based on the data collection method	16
Odds Ratio (OR).....	16
Cohort studies.....	18
Evaluation of a potentially significant link between a risk factor and a disease	18
Chi-square (χ^2) test	18
Quantifying the importance of a link between a risk factor and a disease	18
Risk Ratio (RR).....	18
Attributable risk (AR) – risk difference	19
Therapeutic studies.....	20
Contingency table	20
Evaluation of a potentially significant effect of a treatment on a disease	20
Quantifying the importance of the relationship/difference between treatments - for therapeutical studies – how to choose based on the data collection method	20
Student’s test (t-test).....	20
Quantifying the importance of the effect of a treatment on a disease	21
Ratio of the Experimental Event (REE).....	21
Ratio of the Control Event (RCE).....	21
Absolute Risk Reduction (ARR)	22
Relative Risk Reduction (RRR).....	22
Relative Risk (RR)	22
Number Needed to Treat (NNT)	22
Number Needed to Harm (NNH).....	22
Diagnostic studies	23
Contingency table	23
Testing if there is a significant link between the result of a diagnostic test and the presence or absence of a disease	23
Quantifying the importance of the relationship - for diagnostic studies– how to choose based on the data collection method.....	23
Mc Nemar test	23

Quantifying the characteristics of the evaluated diagnostic test	24
Indicators that may be computed only in representative samples	24
Phase II Indicators:	24
Phase III Indicators:	25
Receiver operating characteristic – ROC curve.....	26
Evaluation of a potentially significant link between a variable and the survival time	29
Log-rank test	29
Quantifying the importance of a link between a variable and the survival time	29
Hazard Ratio.....	29
Regression studies for proposing clinical models	31
Evaluation of a potentially significant regression model between different variables/characteristics	31
Single Linear Regression	31
Multiple Linear Regression	31
Quantifying the importance of a correlation between different variables/characteristics	32
Pearson’s correlation coefficient (r) and the r-squared value (r ²).....	32
Evidence hierarchy.....	33
Therapeutic Studies / adverse reactions /etiology/prevention.....	33
Prognostic Studies (ex. Case-control, exposed not-exposed).....	34
Diagnostic Studies.....	34
Assessing the validity of scientific articles.	35
Therapy studies.....	35
Prognostic studies (ex. Case-control, exposed not-exposed).....	36
Diagnostic studies	36
The relationship between the confidence interval and the P value.....	37
Searching for scientific articles using the PICO technique.....	37
Searching for therapy information:	37
Searching for diagnosis information:	37
Searching for prognosis (risk factors) information:	37

Citing references using the Vancouver system.....	38
References for papers presenting original research in printed journals:	38
References for papers presenting original research in electronic journals:.....	38
References for books:	39
Critical appraisal guide for scientific medical writing (articles, thesis).....	39
Oral presentation principles (article, thesis).....	42

Variable types

Qualitative (categorical)

Nominal (e.g. hair color)

Nominal dichotomous/binary (e.g. gender)

Ordinal (e.g. disease severity)

Quantitative

Continuous (e.g. weight)

Discrete (e.g. number of children in a family)

Survival

(e.g. survival time before death)

Descriptive statistics

How to choose the chart/table to describe one variable

- a. for **qualitative** variables:
 - Frequency tables
 - Pie charts, column charts, bar charts
- b. for **quantitative** variables:
 - Mean and standard deviation – for normal distributed data
 - Mean and interquartile range – for non-normal distributed data
 - Frequency tables
 - Histograms
 - Synthetic graphics
 - normally distributed variables: error plot, means plot, columns with error bars
 - non normally distributed variables: box and whiskers chart (box plot)
- c. for **survival** variables:
 - Median of survival time
 - Survival probability chart Kaplan Meier

Details

Qualitative (categorical) data

Frequency tables and Pie charts

Interpretation: evaluate and discuss which category of the described variable has a higher presence in the studied sample.

Quantitative data

Mean, standard deviation (SD)

Format to be used: mean \pm standard deviation (e.g. “the cholesterol level in the studied sample has been 190 ± 15 mg/dl”).

Interpretation:

- the interval mean \pm 1SD contains approx. 68% of all observations in the studied sample
- the interval mean \pm 2SD contains approx. 95% of all observations in the studied sample
- the interval mean \pm 3SD contains approx. 99% of all observations in the studied sample

95% confidence interval for means

Format to be used: mean (95% CI lower limit-upper limit) (e.g. “a mean body length at birth of 45.7 cm (95% CI 43.7-47.7) has been found”). Values are usually written using one decimal.

Interpretation:

- **Simplified interpretation:** the 95% confidence interval for a parameter (e.g. the mean body length) is the interval that contains the respective parameter for the target population with a 0.95 probability. In other words, we can be 95% sure that the interval contains the investigated parameter (e.g. the mean body length) of the target population.
- **Correct interpretation:** the 95% confidence interval for a parameter (e.g. the mean body length) is the interval that includes 95% of the computed parameters for any number of samples that were randomly extracted from the target population.
- The 95% confidence interval of a parameter approximately equals the parameter \pm 2 standard errors (SE). **Attention!** Do not confound standard error (SE) with standard deviation (SD). SD is a measure of data spread in the studied sample, SE measures how far the parameter estimated on the sample lies from the true population parameter.

Histograms and frequency tables for quantitative data

Interpretation:

A histogram represents a frequency distribution by means of rectangles whose **widths** represent class intervals and whose **areas** are proportional to the frequencies in the corresponding frequency table.

The frequency distribution represented by a histogram is interpreted to be a **normal** or Gaussian distribution if it resembles the shape of a bell, i.e. the majority of data is distributed around the mean value and there are progressively fewer values towards the upper and lower extremes.

The frequency distribution is a **non-normal** one if it is significantly skewed, displaying an asymmetrical shape, not resembling the shape of a bell.

Survival data

Median of survival time

Format to be used: median value followed by time unit (e.g. *“the median survival time in the studied sample has been 14 months”*).

Interpretation: the median of survival time in a group represents the time for which half of the subjects in that group survive (i.e. they do not complete the predefined event)

Contingency table between grouping variable and censor status

Interpretation: if one or several categories of the grouping variable display a small number of completed events (fewer than 10), the estimations of our study will be imprecise, while tests and quantifying indicators (e.g. the Hazard Ratio) will probably not reach statistical significance.

Kaplan-Meier survival probability charts

Interpretation: the survival probability chart shows the time dependency of the probability to survive (i.e. not to complete the predefined event) for the studied subjects. The shape of the curve may show, for instance, a high rate of completing the predefined event rather in earlier than in later stages of the follow-up (e.g. there will be a relatively high number of deaths immediately after an operation, due to operative risks, while deaths due to the investigated disease will start appearing at a later time), or there may be numerous other possible shapes for different survival curves.

Analysis of potential relations between two variables, using tables or charts

How to choose a table/chart for analyzing the relation between two variables

- a. for the **relation between two qualitative** variables:
 - Contingency tables
 - Column charts
- b. for the **relation between a quantitative variable and a qualitative** variable (a quantitative variable compared by groups – eg. Age in function of the treatment – a drug vs. placebo):
 - for one quantitative variable depending on a qualitative variable (For grouped quantitative data)
 - Mean and standard deviation – for normal distributed data
 - Mean and interquartile range – for non normal distributed data
 - Frequency tables
 - Histograms
 - Synthetic graphics
 - normally distributed variables: error plot, means plot, columns with error bars
 - non normally distributed variables: box and whiskers chart (box plot)
 - for the **relation between two qualitative variables** - one quantitative variable depending on another quantitative variable - Scatter charts
- c. for the **relation between a survival variable and a qualitative** variable (a survival variable compared by groups – eg. Time to death in subjects with pancreatic cancer in function of the treatment – a drug vs. placebo):
 - a. Survival probability charts – Kaplan Meier

Details

Qualitative (categorical) data

Contingency tables and the associated proportional charts

Interpretation: evaluate and discuss whether there is a proportional or a non-proportional distribution between the investigated variables.

A non-proportional distribution usually points towards a potential link between the investigated variables. However, the existence of such a link needs to be tested using inferential statistics (i.e. statistical tests that compare frequency distributions).

Quantitative data

Scatter charts

Interpretation:

If the cloud of points in the scatter chart is disposed more or less along a sloping line, there probably is a **linear relation** between the two variables. If the sloping line is ascending, the relation between the two variables is directly proportional. If the sloping line is descending, the relation between the two variables

is inversely proportional. To further investigate and analyze the relation between two such variables, correlation and linear regression can be used.

If the cloud of points in the scatter chart is disposed more or less along a curve, such as a polynomial, an exponential or a logarithmical curve, then there probably exists **another relation** (i.e. polynomial, exponential or logarithmical) between the two variables. Linear regression and correlation can not be used to further investigate and analyze such a relation.

If the cloud of points in the scatter chart is disposed in an approximate circle or along a vertical or horizontal line, then there is probably **no relation** between the two quantitative variables.

Survival data

Kaplan-Meier survival probability charts

Interpretation:

If two or several survival curves partially overlap, by crossing each other several times, then there is most likely no significant difference in survival between categories corresponding to each curve.

If the survival curve corresponding to one of the investigated categories is constantly above the curve of another category, then the survival of the first category is longer than the survival of the second category. The larger the distance between curves, the higher the difference in survival probability between the respective categories.

Statistical tests

How to choose a statistical test (the most common used ones)

- Evaluation of the existence of a relationship between a disease and a prognostic factor:
 - Comparison of **qualitative** data
 - 2 / >2 independent samples (expected frequencies in the contingency table > 5 for > 80% of cells)
 - Chi square Test
 - 2 / >2 independent samples (expected frequencies in the contingency table < 5 for > 20% of cells)
 - Fisher's exact test
 - Two dependent / paired samples
 - Mc Nemar Test
 - Comparison of **quantitative** data that **follow a normal distribution**
 - Two independent samples
 - Student test for independent samples
 - Assuming equal / unequal variances
 - Two dependent / paired samples

- Student test for paired samples
 - Excel: t test paired two samples for means
- Comparison of quantitative data that do not follow a normal distribution
 - Two independent samples
 - Mann Whitney U test
 - Two dependent / paired samples
 - Wilcoxon test for paired samples
- Comparison of survival data
 - 2 / >2 independent samples
 - Log-rank test

Choice between statistical test on independent samples and statistical test on dependent samples (pairs)

Dependent samples (paired): where each subject in the sample has a correspondent with similar characteristics in the other sample

- Comparison between twins
- Comparison on the same subjects repeatedly observed over time (eg Before and After Treatment)
- Comparison of matched samples - by researchers (for each subject in the case group there is a control subject with the same characteristics - age, gender and other confounding factors)
- Comparison on the same subject (comparing the left and the right side of the skin treated with dermatological products)
- Comparing two measurements methods or diagnostic methods used on the same subjects

Independent samples: each subject in the sample does not have a correspondent with similar characteristics in the other sample

Interpretation

The p-value is the probability to end up by chance with a result like the one found (i.e. the frequency distribution of risk factor and disease) in a situation when there is in fact no link between the investigated parameters (the risk factor and the disease).

If $p < 0.05$, we reject the **statistical null hypothesis** (there is no statistically significant relation between a risk factor and a disease, or there is no statistically significant difference between groups in respect of a characteristic) and we can accept the **statistical alternative hypothesis** to be true (there is a statistically significant relation between a risk factor and a disease, or there is a statistically significant difference between groups in respect of a characteristic)-.

If $p > 0.05$ we cannot reject the **statistical** null hypothesis.

Confidence intervals

Statistical interpretation

- **Simplified interpretation:** the 95% confidence interval for a parameter (e.g. AR) is the interval that contains the respective parameter for the target population with a 0.95 probability. In other words, we can be 95% sure that the interval contains the investigated parameter (e.g. the RR) of the target population. Ex. The true value of AR in the target population is between 10,3% and 30,2% with a probability of 95%.
- **Correct interpretation:** the 95% confidence interval for a parameter (e.g. AR) is the interval that includes 95% of the computed parameters for any number of samples that were randomly extracted from the target population.

The precision of the results

look at the width of the confidence interval

- very precise / relatively precise / imprecise (*wide range - imprecise results, narrow interval - precise results*)

Clinical interpretation

- strong clinical significance (*Both ends with highly significant clinical values*)
- weak clinical significance (*Both ends with less important clinical values*)
- unclear clinical importance (*One end with a high clinical importance the other with a less important value*)

Assessing if the result is statistically significant ($p < 0.05$) by looking at the confidence interval

- If the absence of a link/ difference is within the confidence interval then the result is NOT statistically significant ($p > 0.05$)
- If the absence of a link/ difference is outside the confidence interval then the result IS statistically significant ($p < 0.05$).
- For the ratio type indicators eg. RR, OR, HR, the absence of the effect is indicated by value 1
- For the difference type indicators eg. RA, ARR, means difference, or percentage difference, the absence of the effect is indicated by value 0.)

Domains of study/Study types/ Data collection types

Domain of research

Description of a health phenomenon (the study must be descriptive, not analytical!)

Evaluation of a diagnostic test

Evaluation of a therapeutic approach

Evaluation of prognostic (risk or protective) factors

Study type

Based on study objectives

- a. **Descriptive** (no comparisons are performed, no links, relationships, correlations are pursued)
- b. **Analytical** (comparisons are performed, links, associations, relationships, correlations are pursued)

Based on the researchers role

- a. **Observational** (the researcher does not intervene either on subjects or on the course of the disease)
- b. **Experimental** (the researcher does intervene on both subjects and on the course of the disease – e.g. by administering treatments, performing operations, etc.)

Using the matching technique in studies comparing two groups.

This means that for each subject in the case/exposure group, the researcher identifies a subject with similar characteristics in the comparison group (control/non exposed group). I.e. matching made on: same sex, same age (identical or +/- 2 years), same risk factors (i.e. diabetics). The matching can be made 1:1 or 1:2, 1:3, 1:4 (so one case to 1/2/3/4 controls)

Type of therapeutic study

- Depending on the **design**
 - parallel group (each subject is assigned to one group. Finally we obtain several groups, each patient receiving an intervention only (except where intervention is the absence of any intervention))
 - Cross over (each subject receives as an intervention and then another intervention after a period of time (washout). The intervention order that is given here is random - the subjects are their own witnesses)
- Depending on the **objective**
 - efficacy / explanatory (trial conducted under restrictive conditions tending to the ideal situation in which patients are followed to make sure they take the correct dosage, with the correct schedule) :
 - effectiveness / pragmatic (trial conducted under conditions similar to those in current clinical practice, with patients selected less strictly - provides information that may be recommendations for use in practice) :
- Depending on the **hypothesis**
 - superiority (trial hypothesis is that treatment A is superior to treatment B)
 - non-inferiority (trial hypothesis is that intervention is at least as good compared to other reference intervention) :
 - equivalence (trial hypothesis is that the results of two interventions / treatments are similar) :

- Depending on the **phase** of development of a drug
 - Phase I (safety assessment (observation of serious side effects or less serious ones, identifying safe dosage of medication, assessment of absorption, metabolism, distribution and excretion of the substance) of the drug in humans. Carried out on a small number of subjects (tens).)
 - Phase II (initial evaluation of the effectiveness of the drug. Carried on hundreds of subjects. Moreover it aims to identify other adverse effects) :
 - Phase III (confirmatory phase. Assess whether the drug really is efficient, comparing it with other existing treatments. Performed on a large number of subjects (several thousand). Moreover it continues tracking adverse reactions. If this phase ends well, the drug may be approved by the national agencies for the use of the product on the market) :
 - Phase IV (pharmacovigilance phase / post marketing - after the release (approval) on the market. Evaluates other risks (long-term), evaluates the drug in different populations, or even to see if others may be benefiting of the drug. It can change its use for optimal results (i.e. change in dosage)) :
- Subjects were randomly assigned treatments (**randomized**) ?
- state whether allocation was concealed (**allocation concealed** ?) (Yes / unclear / no) (one who introduced the study subjects must not know what intervention the patient will receive.)
- All patients were analyzed in the groups to which they were randomized (analysis was like "**intention to treat analysis** ") (ie subjects who received the new treatment, even if take it or take it discontinuously or take it wrong, are analyzed as if they have taken all the treatment correctly, and the same for the control group)
- the masking method that was used (mask - **blind**) for
 - open (**open label**) (both patient and doctor which assesses the outcome know what the patient received)
 - **single-blind** (the patients are not aware of which interventions they are receiving)
 - **double-blind** (neither the patient nor the doctor knows which evaluates the results of intervention know what intervention the patient is receiving)
- The trial was **controlled** ? (the treatment of interest was compared with one control (placebo, reference treatment, other treatment))

Data collection method

Based on the studied population

- a. **Exhaustive** (the whole target population is studied)
- b. **Sampling** (a sample of the target population is studied)

Based on the duration of data collection

Cross-sectional (data is collected at a given moment, like a „photographic image“ of the medical situation)

Longitudinal (data is collected from the past or from the future, when referred to the start of a study)

Retrospective (data is collected from the past (e.g. exposure to risk factors), from questionnaires or patient records)

Prospective (data is collected in the future, by following patients over time (e.g. following the appearance or remission of a disease))

Based on the grouping method

a. one group (which can be a **representative sample** (a group that faithfully copied the characteristics of the population from which it was extracted - for this a probabilistic sampling is used - simple, random stratified, systematic, cluster), **a cohort** of subjects group of subjects with a common feature, or the same inclusion and exclusion criteria), or a **consecutive series of cases**, or a **series of cases**, without specifying whether the patients were taken successively in the study)

b. **Exposed – nonexposed** (two groups are followed: one exposed, the other non-exposed to a prognostic factor)

c. **Case - control** (two groups are followed: the cases, who have the predetermined disease, and the controls, who are free of the predetermined disease)

Inferential Statistics (Data Analysis)

Case-control studies

Contingency table

	Disease present	Disease absent	Total
Factor present	a	b	a+b
Factor absent	c	d	c+d
Total	a+c	b+d	a+b+c+d

Evaluation of a potentially significant link between a risk factor and a disease

Chi-square (χ^2) test

Motivation of choice: the chi square test compares frequency distributions of qualitative variables.

Format to be used: p =two-tailed p -value (using up to 3 decimals) – name of the test (e.g. “ $p=0.01$ – chi square test” or “ $p<0.001$ – chi square test”)

Statistical Interpretation: The p -value is the probability to end up by chance with a result like the one found (i.e. the frequency distribution of risk factor and disease) in a situation when there is in fact no link between the investigated parameters (the risk factor and the disease).

If $p < 0.05$, we reject the **statistical** null hypothesis (stating the absence of a significant difference between the observed distribution of risk factor and disease and the frequency distribution expected by chance - Ex. the absence of a link between the risk factor and the disease) and we can accept the **statistical** alternative hypothesis to be true (stating the existence of a significant difference between the observed distribution of risk factor and disease and the frequency distribution expected by chance - Ex. the existence of a link between the risk factor and the disease).

If $p > 0.05$ we cannot reject the **statistical** null hypothesis.

Quantifying the importance of a link between a risk factor and a disease

Quantifying the importance of the relationship - for studies of risk factors- how to choose based on the data collection method

- **The point estimator** (it provides information to what was observed in the studied sample) **and its 95% confidence interval** (it provides information regarding what is happening in the target population / in reality - if the samples are correctly chosen)
 - one group (**representative sample, or a cohort, or a consecutive serie of cases, or a serie of cases**)
 - Relative Risk / Risk Ratio - RR
 - Attributable Risk / Risk Difference – RD, , OR is possible too
 - **exposed - not exposed sampling**
 - Relative Risk / Risk Ratio - RR
 - Attributable Risk / Risk Difference - RD, OR is possible too
 - **case – control sampling**
Odds Ratio - OR

Odds Ratio (OR)

Motivation of choice: a case-control study does not allow computing the risk ratio (RR), since cases and controls have different selection criteria, thus, diseased and non-diseased subjects have different selection criteria inside the exposed, respectively non-exposed groups.

Therefore, in case-control studies, RR **must** be approximated by OR.

Format to be used: OR=point estimate value (95% CI lower limit-upper limit) (e.g. “OR=3.2 (95% CI 2.5-3.9)”). Values are usually written using one decimal.

Interpretation: odds ratio is a good approximation of risk ratio if the investigated disease is rare or has a long period of incubation, resulting in the use of a case-control study. Ex. The odds of lip cancer is 3.2 times higher in those who drink strong alcoholic beverages compared to those who don't drink.

OR – point estimate

- An **Odds Ratio of 1** indicates that the exposure condition under study is equally likely in both the case and the control group.
- An **Odds Ratio greater than 1** indicates that that the exposure condition under study is more likely in the first group (i.e. the case group).
- An **Odds Ratio lower than 1** indicates that the exposure condition under study is less likely in the first group (i.e. the case group).

OR – 95% CI

- **Simplified interpretation:** the 95% confidence interval for a parameter (e.g. OR) is the interval that contains the respective parameter for the target population with a 0.95 probability. In other words, we can be 95% sure that the interval contains the investigated parameter (e.g. the OR) of the target population. Ex. The true value of OR in the target population is between 2.5 and 3.9 with a probability of 95%.
- **Correct interpretation:** the 95% confidence interval for a parameter (e.g. OR) is the interval that includes 95% of the computed parameters for any number of samples that were randomly extracted from the target population.

Cohort studies

Evaluation of a potentially significant link between a risk factor and a disease

Chi-square (χ^2) test

Motivation of choice: the chi square test compares frequency distributions of qualitative variables.

Format to be used: p =two-tailed p -value (using up to 3 decimals) – name of the test (e.g. “ $p=0.01$ – chi square test” or “ $p<0.001$ – chi square test”).

If $p<0.05$, we reject the **statistical** null hypothesis (stating the absence of a significant difference between the observed distribution of risk factor and disease and the frequency distribution expected by chance - Ex. the absence of a link between the risk factor and the disease) and we can accept the **statistical** alternative hypothesis to be true (stating the existence of a significant difference between the observed distribution of risk factor and disease and the frequency distribution expected by chance - Ex. the existence of a link between the risk factor and the disease).

If $p>0.05$ we cannot reject the **statistical** null hypothesis.

Quantifying the importance of a link between a risk factor and a disease

Risk Ratio (RR)

Motivation of choice: a cohort study allows computing the risk ratio (RR), since diseased and non-diseased subjects have similar selection criteria inside the exposed, respectively non-exposed groups.

Format to be used: RR=point estimate value (95% CI lower limit-upper limit) (e.g. “RR=3.2 (95% CI 2.5-3.9)”). Values are usually written using one decimal.

Interpretation: Risk ratio (or relative risk) is the ratio of the probability of the disease occurring in the exposed group versus the non-exposed group. Ex. The risk of lip cancer is 3.2 times higher in those who drink strong alcoholic beverages compared to those who don't drink.

RR – point estimate

- A **Risk Ratio of 1** means there is no difference in risk between the exposed group and the non-exposed group.
- A **Risk Ratio greater than 1** means the disease is more likely to occur in the exposed group than in the non-exposed group.
- A **Risk Ratio lower than 1** means the disease is less likely to occur in the exposed group than in the non-exposed group.

RR – 95% CI

- **Simplified interpretation:** the 95% confidence interval for a parameter (e.g. RR) is the interval that contains the respective parameter for the target population with a 0.95 probability. In other words, we can be 95% sure that the interval contains the investigated parameter (e.g. the RR) of the target population. Ex. The true value of RR in the target population is between 2.5 and 3.9 with a probability of 95%.
- **Correct interpretation:** the 95% confidence interval for a parameter (e.g. RR) is the interval that includes 95% of the computed parameters for any number of samples that were randomly extracted from the target population.

Attributable risk (AR) – risk difference

Motivation of choice: a cohort study allows computing the attributable risk (AR), since diseased and non-diseased subjects have similar selection criteria inside the exposed, respectively non-exposed groups.

Format to be used: AR=point estimate value (95% CI lower limit-upper limit) (e.g. “RR=15% (95% CI 10,3%-30,2%)”). Values are usually written using one decimal.

Interpretation: Attributable risk is the probability difference of the disease occurring in the exposed group versus the non-exposed group. Ex. The risk of lip cancer is 15% higher in those who drink strong alcoholic beverages compared to those who don’t drink.

AR – point estimate

- An **Attributable Risk of 0** means there is no difference in risk between the exposed group and the non-exposed group.
- An **Attributable Risk greater than 0** means the disease is more likely to occur in the exposed group than in the non-exposed group.
- An **Attributable Risk lower than 0** means the disease is less likely to occur in the exposed group than in the non-exposed group.

AR – 95% CI – confidence interval

- **Simplified interpretation:** the 95% confidence interval for a parameter (e.g. AR) is the interval that contains the respective parameter for the target population with a 0.95 probability. In other words, we can be 95% sure that the interval contains the investigated parameter (e.g. the RR) of the target population. Ex. The true value of AR in the target population is between 10,3% and 30,2% with a probability of 95%.
- **Correct interpretation:** the 95% confidence interval for a parameter (e.g. AR) is the interval that includes 95% of the computed parameters for any number of samples that were randomly extracted from the target population.

Therapeutic studies

Contingency table

	Event present	Event absent	Total
New treatment	a	b	a+b
Control treatment	c	d	c+d
Total	a+c	b+d	a+b+c+d

Evaluation of a potentially significant effect of a treatment on a disease

Quantifying the importance of the relationship/difference between treatments - for therapeutic studies – how to choose based on the data collection method

- **The point estimator** (it provides information to what was observed in the studied sample) **and its 95% confidence interval** (it provides information regarding what is happening in the target population / in reality - if the samples are correctly chosen)
- one group (**representative sample, or consecutive series of cases, or a series of cases**)
 - **ARR, RR, OR, NNT** (REE, REC can be calculated, but they do not quantify the importance of the difference)
- exposed - not exposed sampling
 - **ARR, RR, OR, NNT** (REE, REC can be calculated, but they do not quantify the importance of the difference)
- case – control sampling
 - **Odds Ratio - OR**

Student's test (t-test)

Motivation of choice: Student's t-test is applied to compare means of quantitative variables, for small samples (<50 for independent samples or <100 for paired samples) with normal distribution (i.e. bell-shaped histogram), or large samples exhibiting a non-normal histogram.

Attention!

When comparing variables of the same group in different situations (e.g. blood pressure measured for the same group of patients, before and after a hypotensive treatment) the compared samples are paired and the correct choice for a t-test performed in *Excel* is **t-Test: Paired Two Sample for Means**.

When comparing variables of different groups (e.g. blood pressure after a hypotensive treatment in diabetic versus non-diabetic patients) the compared samples are independent and the correct choice for a t-test performed in *Excel* depends on the p-value rendered by the **F-Test Two-Sample for Variances**:

- If the p-value ≥ 0.05 , a **t:Test: Two-Sample Assuming Equal Variances** should be chosen
- If the p-value < 0.05 , a **t:Test: Two-Sample Assuming Unequal Variances** should be chosen

Format to be used: p =two-tailed p -value (using up to 3 decimals) – name of the test (e.g. “ $p=0.01$ – Student’s t -test for paired samples” or “ $p<0.001$ – Student’s t -test for independent samples, assuming equal variances”).

Statistical Interpretation: The p -value is the probability to end up by chance with a result like the one found (i.e. the observed mean values of the investigated variable in the two samples) in a situation when there is in fact no difference between the mean values of the investigated variable in the two samples.

If $p<0.05$, we reject the **statistical** null hypothesis (stating the absence of a significant difference between the compared means) and we can accept the **statistical** alternative hypothesis to be true (stating the existence of a significant difference between the compared means).

If $p>0.05$ we cannot reject the **statistical** null hypothesis.

Clinical Interpretation: If $p>0.05$, the absence of a statistically significant difference between the compared means always translates in a similar absence of clinical significance.

However, the same does not hold true in the presence of statistical significance. A statistically significant difference between the compared means ($p<0.05$) does not necessarily mean that this difference is also of clinical significance. In order to establish clinical significance in such a case, the actual **difference between the compared means** must be computed by subtracting the two means rendered in the test-result worksheet and the obtained difference must be compared with a predefined threshold value established beforehand in the research protocol as a minimum level for clinical significance (e.g. we may impose by protocol that a cholesterol-lowering treatment results in a mean cholesterol value at least 20 mg/dl lower compared to placebo, in order to consider its effect clinically significant.)

Quantifying the importance of the effect of a treatment on a disease

Ratio of the Experimental Event (REE)

REE = the proportion of patients treated with the treatment to be evaluated (i.e. patients in the experimental group) who responded favorable to the treatment = $a/(a+b)$. Eg. REE=3% - There is a 3% of infarction for those who took aspirin

Ratio of the Control Event (RCE)

RCE = the proportion of patients treated with the control treatment (i.e. patients in the control group) who responded favorable to the treatment = $c/(c+d)$. Eg. REC=6% - There is a 6% of infarction for those who took placebo

Absolute Risk Reduction (ARR)

ARR = the absolute difference between the ratio of the experimental event and the ratio of the control event = $|REE - RCE|$. Eg. ARR=3% - There is a 3% less risk of infarction for those who took aspirin compared to those who took placebo

Relative Risk Reduction (RRR)

RRR = the relative difference (i.e. ratio) between the ratio of the experimental event and the ratio of the control event = $|REE - RCE| / RCE$.

Relative Risk (RR)

RR = the relative risk between the ratio of the experimental event and the ratio of the control event = REE / REC . Eg. RR=7 – there is 7 times more risk of being cured of Helicobacter pylori gastric infection in those treated with antibiotic compared to those treated with placebo. Eg. 2: RR=0.4 – there is 0.4 times more risk of infarction in those treated with aspirin compared to those treated with placebo – this means the risk is lower in the aspirin group.

Number Needed to Treat (NNT)

NNT = the number of patients that needs to be treated with the experimental treatment in order for one patient to benefit from this treatment (i.e. to treat the disease) = $1/ARR$. (e.g. NNT=3 for a new cholesterol-lowering treatment in which the threshold for clinical efficacy was set at 20 mg/dl, should be interpreted as: “3 persons with hypercholesterolemia need to be treated using the new cholesterol-lowering treatment in order for one person to have her cholesterol blood level lowered by more than 20 mg/dl.”)

Number Needed to Harm (NNH)

NNH = the number of patients that needs to be treated with the experimental treatment in order for one patient to have a negative reaction from this treatment (eg. An adverse reaction – an allergic reaction) = $1/ARR$. (e.g. NNH = 22, to have an gastric hemorrhage when using 30 mg aspirin for prevention of myocardial infarction. It is necessary to treat 22 patients with 30 mg aspirin used for the prevention of infarction, so that one will have a gastric bleeding.)

Diagnostic studies

Contingency table

	Disease present	Disease absent	Total
Test positive	TP a	FP b	a+b
Test negative	FN c	TN d	c+d
Total	a+c	b+d	a+b+c+d

Testing if there is a significant link between the result of a diagnostic test and the presence or absence of a disease

Quantifying the importance of the relationship - for diagnostic studies- how to choose based on the data collection method

- **The point estimator** (it provides information to what was observed in the studied sample) **and its 95% confidence interval** (it provides information regarding what is happening in the target population / in reality - if the samples are correctly chosen)
 - one group (**representative sample, or consecutive serie of cases, or a serie of cases**)
 - Se, SP, LR+, LR-, PPV, NPV, prevalence
 - **exposed - not exposed sampling**
 - PPV, NPV
 - **case – control sampling**
 - Se, SP, LR+, LR-

Mc Nemar test

Motivation of choice: the test compares frequency distributions of dependent (paired) qualitative variables.

Format to be used: p =two-tailed p -value (using up to 3 decimals) – name of the test (e.g. “ $p=0.01$ – chi square test” or “ $p<0.001$ – chi square test”).

Statistical Interpretation: The p -value is the probability to end up by chance with a result like the one found (i.e. the frequency distribution of the test result and the disease) in a situation when there is in fact no link between the investigated parameters (the test result and the presence or absence of the disease).

If $p<0.05$, we reject the **statistical** null hypothesis (there is no statistically significant difference between the two diagnostic tests - stating the absence of a significant difference between the observed

distribution of diagnostic test results and disease and their frequency distribution expected by chance) and we can accept the **statistical** alternative hypothesis to be true (there is a statistically significant difference between the two diagnostic tests - stating the existence of a significant difference between the observed distribution of diagnostic test results and disease and their frequency distribution expected by chance).

If $p > 0.05$ we cannot reject the **statistical** null hypothesis.

Clinical Interpretation:

If $p < 0.05$, we reject the **clinical** null hypothesis (stating the absence of a link between the result of a diagnostic test and the presence or absence of a disease) and we can accept the **clinical** alternative hypothesis to be true (stating the existence of a link between the result of a diagnostic test and the presence or absence of a disease).

If $p > 0.05$ we cannot reject the **clinical** null hypothesis.

Quantifying the characteristics of the evaluated diagnostic test

Format to be used: all indicators should be written as: Indicator=percentage value, without decimals (e.g. Se=89%). If an associated confidence interval has been computed, it should be written as (95% CI lower limit – upper limit).

Indicators that may be computed only in representative samples

Prevalence = $(TP+FN)/(TP+FP+FN+TN)$ = the total number of cases of the disease in the population at a given time = the total number of cases in the population divided by the number of individuals in the population.

Phase II Indicators:

Sensitivity = $Se = TP/(TP+FN) = p(T^+ | B^+)$ = the probability of a diseased subject to test positive using the new diagnostic test (e.g. Se=95% means that 95% of the diseased subjects tested positive using the new diagnostic test. In other words: there is a probability of 95% for a diseased subject to test positive using the new diagnostic test, or the probability of a subject with disease to have the new test positive is 95%, or the proportion of subjects with a positive result for the new diagnostic test within the group with the disease is 95%, or 95% of the subjects with disease have a positive result for the new diagnostic test.)

Specificity = $Sp = TN/(TN+FP) = p(T^- | B^-)$ = the probability of a non-diseased subject to test negative using the new diagnostic test (e.g. Sp=98% means that 98% of the non-diseased subjects tested negative using the new diagnostic test. In other words: there is a probability of 98% for a non-diseased subject to test negative using the new diagnostic test, or the probability of a subject without disease to have the new test negative test is 98%, or the proportion of subjects with a negative result for the new diagnostic test within the group without the disease is 98%, or 98% of the subjects without disease have a negative result for the new diagnostic test.)

Clinical Interpretation of Sensitivity and Specificity:

A diagnostic indicator as Se, Sp, and accuracy approaching 100% means a high diagnostic quality. (Approximately <80% - reduced >=80-90% - moderate, >= 90% high)- (ex. Se = 85% means that the new diagnostic test has a moderate sensitivity) But for some situations one needs values very close to 100% to be content with a test- ex. Test for HIV). A diagnostic indicator close to 50% (but less than 50%) as a diagnostic means almost absent.

Accuracy = $(TP+TN)/(TP+FP+FN+TN)$ = the proportion of correct (true) results among all test results.

Positive Likelihood Ratio = (+LR) = $Se/(1-Sp)$. It shows how many times a **positive** test result is more likely to appear in a diseased subject than in a non-diseased one.

Negative Likelihood Ratio = (-LR) = $(1-Se)/Sp$. It shows how many times a **negative** test result is more likely to appear in a diseased subject than in a non-diseased one.

Clinical Interpretation of Likelihood Ratio: the LR of a diagnostic test is a measure for its capacity to alter the pre-test probability of a subject to have the tested disease. The values of LR can induce more or less significant changes in the probability of a certain diagnosis:

- If $LR > 10$ or $LR < 0.1$, this leads to significant and often conclusive changes in the pre-test probability of a certain diagnosis.
- If LR is between 5 and 10 or LR is between 0.1 and 0.2, this leads to moderate and useful changes of the pre-test probability of a certain diagnosis.
- If LR is between 2 and 5 or LR is between 0.2 and 0.5, this leads to little but sometimes important changes of the pre-test probability of a certain diagnosis.
- If LR is between 1 and 2 or LR is between 0.5 and 1, this leads to very modest and rarely important changes of the pre-test probability of a certain diagnosis.

Phase III Indicators:

Positive Predictive Value = $PPV = TP/(TP+FP) = p(B^+ | T^+)$ = the probability of a subject who tested positive using the new diagnostic test to actually have the tested disease (e.g. $PPV=95\%$ means that 95% of subjects who tested positive using the new diagnostic test actually had the tested disease. In other words: there is a probability of 0.95 for a subject who tests positive using the new diagnostic test to actually have the tested disease.)

Negative Predictive Value = $NPV = TN/(TN+FN) = p(B^- | T^-)$ = the probability of a subject who tested negative using the new diagnostic test to actually not have the tested disease (e.g. $NPV=98\%$ means that 98% of subjects who tested negative using the new diagnostic test actually did not have the tested disease. In other words: there is a probability of 0.98 for a subject who tests negative using the new diagnostic test to actually not have the tested disease.)

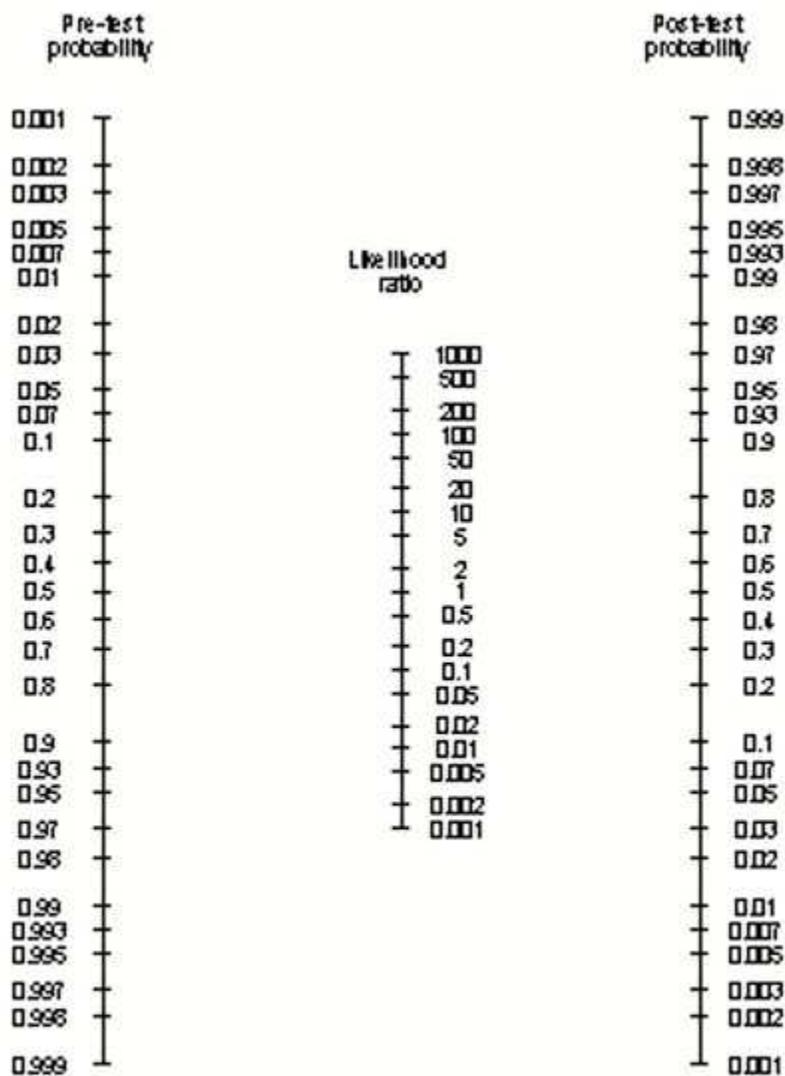


Fig. 1 Fagan nomogram for computing the post-test probability to have the tested disease, when knowing the pre-test probability, the test result and the likelihood ratio of the applied test.

On the first axis: the probability before using the diagnostic test

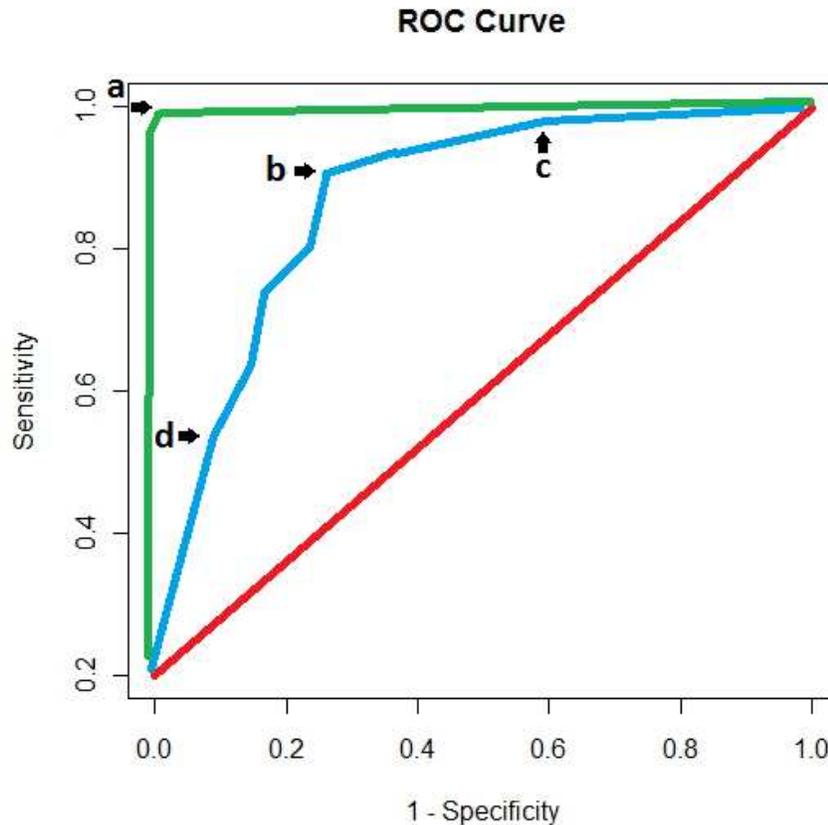
On the second axis (LR): if the result of the test is positive – use LR+, if the result of the test is negative use LR-

Receiver operating characteristic – ROC curve

The ROC chart

- horizontal axis = 1 - Specificity
- vertical axis = Sensitivity

- the curve represents all the values of Se, 1-Sp for all the possible values of the cut-off for the quantitative result of the test
- the closer the curve is
 - at the top left corner - high diagnostic value - discriminant test (see green curve) - - AUROC (area under the ROC) / AUC close to 1 – green curve
 - diagonally - diagnostic value absent - non-discriminating test (see red curve) - - AUROC (area under the ROC) / AUC close to 0.5 – red curve



The choice of the cut-off - to have the best combination of Se and Sp high

- Taking into account the purpose of the study (see the example for the blue curve)
 - Diagnostic screening - High sensitivity see (c) - Se = 0.95 = 95%, Sp = 1 - 0.6 = 0.4 = 40%
 - Precision diagnostics - High specificity - see (d) - Se = 0.52 = 52%, Sp = 1 - 0.1 = 0.9 = 90%
 - Best Se / Sp ratio - the closest point to the upper left corner - see (b) - Se = 0.9 = 90%, Sp = 1 - 0.25 = 0.75 = 75%

For AUROC one can do a statistical test for AUROC to see if it is statistically significant.

- The null hypothesis of the statistical test for AUROC is that AUROC is equal to 0.5

- The alternative hypothesis of the statistical test for AUROC is that AUROC is equal to 0.5
- If the p value of the test is <0.05 - the result is statistically significant
- If the p value of the test is > 0.05 – one cannot say the result is not statistically significant

There is a relationship between the 95% confidence interval of AUROC and the p value of the test for AUROC

- if the 95% confidence interval of AUROC contains the value 0.5, - the null hypothesis could not be rejected, the p value of the test is > 0.05 , thus one cannot say the result of the statistical test for the AUROC is statistically significant
- If the 95% confidence interval of AUROC does not contain the value 0.5, - we can reject the null hypothesis, accept the alternative hypothesis, the p value of the test is <0.05 , the result of the statistical test for AUROC is statistically significant

Survival studies

Evaluation of a potentially significant link between a variable and the survival time

Log-rank test

Motivation of choice: the Log-rank test is a non-parametric test to compare the survival distributions of two samples.

Format to be used: p =two-tailed p -value (using up to 3 decimals) – name of the test (e.g. “ $p=0.01$ – Log-rank test” or “ $p<0.001$ – Log-rank test”).

Statistical Interpretation: The p -value is the probability to end up by chance with a result like the one found (i.e. the observed survival distributions in the two samples) in a situation when there is in fact no difference between survival times in the two samples.

If $p<0.05$, we reject the **statistical** null hypothesis (stating the absence of a significant difference between the compared survival distributions) and we can accept the **statistical** alternative hypothesis to be true (stating the existence of a significant difference between the compared survival distributions).

If $p>0.05$ we cannot reject the **statistical** null hypothesis.

Clinical Interpretation: If $p>0.05$, the absence of a statistically significant difference between the compared survival distributions always translates in a similar absence of clinical significance.

However, the same does not hold true in the presence of statistical significance. A statistically significant difference between the compared survival distributions ($p<0.05$) does not necessarily mean that this difference is also of clinical significance. In order to establish clinical significance in such a case, the actual **difference between the median survival times** in the compared groups must be computed, by subtracting the two median survival times. The obtained difference must be compared with a predefined threshold value established beforehand in the research protocol as a minimum level for clinical significance (e.g. we may impose by protocol that only a difference in median survival time higher than 3 month will be considered to be clinically significant.)

Quantifying the importance of a link between a variable and the survival time

Hazard Ratio

Motivation of choice: a survival study allows computing the hazard ratio (HR) as an estimate of relative risk. The hazard ratio is the effect of an explanatory variable on the hazard or risk of an event.

Format to be used: HR=point estimate value (95% CI lower limit-upper limit) (e.g. “ $HR=0.6$ (95% CI 0.4-0.8)”).

Interpretation: For two individuals who differ only in a relevant membership to a group (e.g. female vs. male, or treatment vs. control), HR is the ratio between the predicted hazard for a member of one group and that for a member of the other group, holding everything else constant in a Cox regression.

HR – point estimate

For a **qualitative** explanatory variable:

- A **Hazard Ratio of 1** means that the survival time is the same (the hazard is the same) for the compared groups (e.g. for an urban group coded with 1 and a rural group coded with 0, if the HR=1 for the survival time of the two groups, then, while holding everything else constant in a Cox regression, the survival time is the same (the hazard is the same) for the urban and the rural group. In other words, the hazard of a person living in an urban area is the same as the hazard of a person living in a rural area.)
- A **Hazard Ratio greater than 1** means that the survival time is shorter (the hazard is higher) in the first group than in the second group. (e.g. for an urban group coded with 1 and a rural group coded with 0, if the HR=1.23 for the survival time of the two groups, then, while holding everything else constant in a Cox regression, belonging to the urban group raises the hazard by $1.23-1=0.23$, or 23%. In other words, the hazard of a person living in an urban area is with 23% higher than the hazard of a person living in a rural area.)
- A **Hazard Ratio lower than 1** means that the survival time is longer (the hazard is lower) in the first group than in the second group. (e.g. for an urban group coded with 1 and a rural group coded with 0, if the HR=0.67 for the survival time of the two groups, then, while holding everything else constant in a Cox regression, belonging to the urban group lowers the hazard by $1-0.67=0.33$, or 33%. In other words, the hazard of a person living in an urban area is with 33% lower than the hazard of a person living in a rural area.)

For a **quantitative** continuous explanatory variable, the same interpretations apply to a unit difference. (e.g. if a HR=1.02 regarding the survival time of patients with lung cancer grouped by the level of LDH, then, while holding everything else constant in a Cox regression, the raise of the LDH level for a patient by one unit will raise the hazard that the patient dies from lung cancer by $1.2-1=0.2$, or 2%.)

HR – 95% CI

- **Simplified interpretation:** the 95% confidence interval for a parameter (e.g. HR) is the interval that contains the respective parameter for the target population with a 0.95 probability. In other words, we can be 95% sure that the interval contains the investigated parameter (e.g. the HR) of the target population.
- **Correct interpretation:** the 95% confidence interval for a parameter (e.g. HR) is the interval that includes 95% of the computed parameters for any number of samples that were randomly extracted from the target population.

Regression studies for proposing clinical models

Evaluation of a potentially significant regression model between different variables/characteristics

Single Linear Regression

Motivation of choice: the need to predict one dependent variable (y) when knowing the value of one independent variable (x)

Format to be used: $y=ax+b$ (e.g. “age of pregnancy (weeks) = 4.7015 * length of femur (cm) + 3.2012”);
p=value – F test

- instead of x and y , the name of each variable will be used, followed between parentheses by its units of measurement .
- to determine the values for coefficients a and b (intercept), look in the third table that results after using *Regression* from *Data Analysis* in *Microsoft Excel*.
- to determine the significance of the single linear regression model, use the p-value returned by the **F test** found in the second table that results after using *Regression* from *Data Analysis* in *Microsoft Excel*.

Model Evaluation: if the F test $p < 0.05$, the single linear regression model is acceptable, i.e. the independent variable predicts the dependent variable well. Implicitly, there exists a significant link between the dependent variable and the independent variable.

Statistical Interpretation: an increase of one unit of the value for the independent variable will determine the increase of the value for the dependent variable with the value of the coefficient associated to the independent variable (e.g. for the above mentioned example regarding the prediction of the age of pregnancy using the length of the femur, “for each centimeter of femoral growth in the fetus, the age of the pregnancy increases with 4.7015 weeks”)

Clinical Interpretation: the size of the coefficient associated to the independent variable should be evaluated along with its 95% confidence interval. If the 95% CI is narrow, the clinical prediction for the target population is good. Vice-versa, if the 95% CI is wide, the clinical prediction for the target population is low.

Multiple Linear Regression

Motivation of choice: the need to predict one dependent variable (y) when knowing the values of several independent variables (x_1, x_2, \dots, x_n)

Format to be used: $y=a_1x_1+a_2x_2+ \dots +a_nx_n+b$; p=value – F test

- instead of x_1, x_2, \dots, x_n and y , the name of each variable will be used, followed between parentheses by its units of measurement .

- to determine the values for coefficients a_1, a_2, \dots, a_n and b (intercept), look in the third table that results after using *Regression* from *Data Analysis* in *Microsoft Excel*.
- to determine the significance of the multiple linear regression model, use the p-value returned by the **F test** found in the second table that results after using *Regression* from *Data Analysis* in *Microsoft Excel*.

Model Evaluation: if the F test $p < 0.05$, the multiple linear regression model is acceptable, i.e. the independent variables predict the dependent variable well. Implicitly, there exists a significant link between the dependent variable and the independent variables.

Still, if some of the independent variables and their coefficients exhibit non-significant p-values ($p > 0.05$), they should be excluded from the regression model and a new regression model, that doesn't include those independent variables, should be computed.

Statistical Interpretation: for each independent variable, the increase of its value with one unit will determine the increase of the value for the dependent variable with the value of the coefficient associated to the respective independent variable.

Clinical Interpretation: the size of each coefficient that is significant for the regression model (its p-value is lower than 0.05) should be evaluated along with its 95% confidence interval. If the 95% CI is narrow, the clinical prediction for the target population is good. Vice-versa, if the 95% CI is wide, the clinical prediction for the target population is low.

Quantifying the importance of a correlation between different variables/characteristics

Pearson's correlation coefficient (r) and the r-squared value (r²)

Motivation of choice: Pearson's correlation coefficient (r) and the r-squared value (r²) are quantifiers for the importance of a link between two quantitative, continuous variables.

Format to be used: r=value, r²=value (values will be written using 2 decimals).

Interpretation:

Pearson's correlation coefficient (r) shows the degree of linear correlation between the variations of two normally distributed quantitative continuous variables:

A positive r shows a **direct proportional** relation between the investigated variables (the increase of one variable associates with an increase of the other variable).

A negative r shows an **inverse proportional** relation between the investigated variables (the increase of one variable associates with a decrease of the other variable).

The **absolute value of r** , $|r|$ between:

- 0 – 0.25 indicates a weak or absent linear correlation between the investigated variables
- 0.25 – 0.5 indicates a moderate linear correlation between the investigated variables
- 0.5 – 0.75 indicates a good linear correlation between the investigated variables
- 0.75 – 1 indicates a very strong linear correlation between the investigated variables

For non-normally distributed quantitative variables as well as for qualitative, ordered nominal variables, a largely similar interpretation can be given to *Spearman's rank correlation coefficient*, which may be computed in these cases.

(e.g. if $r=0.78$ then “*there is a very strong linear correlation between the investigated variables*”)

The R-squared value (r^2), also known as *determination coefficient*, shows what percentage of the variation of one variable can be explained by the variation of the other variable. (e.g. if $r^2=0.89$ then “*89% of the variation of y can be explained by the variation of x* ”)

Attention! Although related, Regression and Correlation are two different things!

Regression is a method used to predict the value of a dependent variable from one or several independent variables.

Correlation evaluates the strength of a linear association between two variables.

Evidence hierarchy

Below there are presented three hierarchies for certain types of clinical questions: (from Centre for Evidence-Based Medicine, adapted, simplified for didactic purposes):

Therapeutic Studies / adverse reactions /etiology/prevention

1a: Systematic reviews for randomized controlled trials

1b: Randomized controlled trial with narrow confidence intervals

2a: Systematic reviews for cohort studies

2b: Cohort study (including low quality randomized controlled trials (eg. Followed-up patients below 80%))

3a: Systematic reviews for case-control studies

3b: Case-control study

4: Series of cases (or a cohort or case-control studies of low quality)

5: Experts' opinion without explained critical evaluation or, based on physiology or, based on fundamental research, or based on certain principles.

Prognostic Studies (ex. Case-control, exposed not-exposed)

1a: Systematic reviews for cohort studies (ex. exposed not-exposed)

1b: Cohort study (ex. exposed not-exposed) with followed-up patients below 80%

2a: Systematic reviews for retrospective cohorts (case-control studies) or untreated control groups out of randomize controlled trials

2b: For retrospective cohorts (case-control studies) or untreated control groups out of randomize controlled trials

3.-

4: Series of cases (or a cohort or case-control studies of low quality)

5: Experts' opinion without explained critical evaluation or, based on physiology or, based on fundamental research, or based on certain principles.

Diagnostic Studies

1a: Systematic reviews for diagnostic studies of 1b level

1b: Independent comparison, using the blind method, patients taken in a corresponding consecutive manner, on which the reference test and the new test were applied.

2a: Systematic reviews for diagnostic studies of 2 level

2b: Any of the following:

3a: Comparison using the blind method or the objective one

A study with a group of nonconsecutive patients, or a narrow spectrum of subjects; but on whom both tests were applied, the reference one and the new test.

3b: Independent comparison, using the blind method, on a spectrum of corresponding subjects, but for which the reference test wasn't applied to all patients.

4: Any of the following situations:

Comparison using a reference test which was not objective, or the blind method was not used, or the comparison was not independently done.

5: Experts' opinion without explained critical evaluation or, based on physiology or, based on fundamental research, or based on certain principles.

Assessing the validity of scientific articles.

Do you consider the study valid? *We will consider a study valid if the essential validity criteria are met. Each validity criteria that is not met lowers the quality of the stud.*

Therapy studies

1. Essential criteria in validity evaluation

- a. The treatments were randomly assigned to the subjects (YES/NO)
- b. The treatment allocation was concealed („allocation concealed“)? (YES/Vague/NO) (the person who registers the subjects into the study doesn't have to know which patient received a certain treatment)
- c. Were the subjects followed enough time (so that the treatment effect can be visible)? (YES/NO)
- d. Were there enough patients left in the study in order to do a non-biased analysis? (YES/NO) (if **withdrawal/ quitting** rate < 20% - then the number of subjects KEPT in the study or their percentage is $\geq 80\%$)
- e. Were all the patients analyzed in the samples they were randomized into (the **analysis** was of the „**intention to treat**“ type - „intention to treat analysis“)? (are the subjects analyzed as if they all took the treatment correctly- natalizumab or placebo- even if they did not do so or they may have taken it discontinuously) (YES/NO)
- f. Was the trial **controlled**? (YES/NO) (The intended treatment was compared to a control one – nothing, placebo, reference treatment, another treatment)

2. Secondary criteria in validity evaluation

- a. **Are the followed group and the control one comparable?** (YES/NO) (The comparability of samples shows how efficient the randomization was. Randomization's purpose is to make the compared samples as similar as possible when it comes to the factors that may influence the desired result- known and unknown factors. The two samples are comparable if the known factors which may influence the results are evenly distributed among the 2 groups. In case they are not evenly distributed among the compared groups, their effect may be controlled using different techniques (stratification, regression). Through control their effect on the results is eliminated.)
- b. Is the **blind method** used in case of
 - i. **Patients?** (YES/Vague/NO) (the patient doesn't know which treatment he/she receives)
 - ii. **Doctors who treat?** (YES/Vague/NO) (the doctor who treats doesn't know which treatment he/she gives)
 - iii. **Investigators who observe the result?** (YES/Vague/NO) (the investigator who observes the result doesn't know who received a certain treatment)
- c. Is the study a negative one? (There aren't any significant statistical differences between the compared ($p > 0,05$))? (YES/NO).
Only if the study is negative answer to the following question: The **power** of the study is big enough? (YES/NO) (The power of a study consists of the minimal difference that can be observed, if the difference is real, with a certain probability (usually 80%), Eg. A study may have the power of observing a ARR of 10% (a difference of 10% between the treatments) with a probability of 80% - thus, if there really is a difference between the

treatments greater or equal to 10% (ARR) the study may observe this difference in 80 cases out of 100 (if we were to have 100 identical studies, 80 of them would see the difference))

Prognostic studies (ex. Case-control, exposed not-exposed)

1. Essential criteria in validity evaluation
 - a. Is the studied **sample representative**? (YES/NO)
 - b. Is the studied sample **clearly defined**? (YES/NO) (*using inclusion, exclusion criteria*)
 - c. Is the illness stage well defined for the subjects of the study? (YES/NO/Can't be applied)
 - d. In a **cohort** study, the patients were identified **before the development of the expected results**? (YES/NO) (*i.e. before the disease/death*)
 - e. Were the subjects **followed** for a period of **time** long enough (*so that the effects may occur*) (YES/NO)
 - f. Were there enough subjects left in the study for an unbiased analysis?? (YES/NO) (*if **withdrawal/ quitting** rate < 20% - then the number of subjects KEPT in the study or their percentage is >= 80%*)
2. Secondary criteria in validity evaluation
 - a. **The results** were evaluated using the **blind** method? (YES/NO) (*the investigator who observes the result doesn't know who received a certain treatment*)
 - b. The criteria for the **results** are **objective**? (YES/NO)
 - c. Was there an adjustment **made** for the **important prognostic factors**? (YES/NO) (*in case there are known factors which may influence the prognostic, especially if they aren't equally distribute among the samples- their effect may be controlled using different techniques (stratification, regression). Through control their effect on the results is eliminated.*)

Diagnostic studies

1. Essential criteria in validity evaluation
 - a. Is the **reference** test **acceptable**? (YES/NO) (*can we be sure that the test shows the truth- is it the best diagnostic test or among the best ones used for the desired aspect*)
 - b. Were the new and standard tests **independently evaluated**? (YES/Vague/NO) (*different investigators evaluated the 2 tests*)
 - c. Was the blind **method** used? (YES/Vague/NO) (*neither the new test's investigator, nor the standard test's one knew about the other's results*)
 - d. Was the test evaluated on a **sufficient number** of patients? (YES/Vague/NO) (*similar to what happens in real practical cases: including subjects in initial, medium, severe stages of disease; subjects with similar pathologies; treated and untreated subjects – a group of such subjects may be created using randomized sampling or consecutive sampling*)
2. Secondary criteria in validity evaluation
 - a. Is the **standard** test **applied no matter** of the new test's outcome? (YES/Vague/NO)
 - b. The way of applying the tests- were they detailed enough so that they may be **replicated** ? (YES/NO) (*it describes the way of executing the test, the patient's preparations for the test, contraindications / precautions for the test, precautions after the test, risks or uncomfortable issues for the patient who is subjected to the test*)

The relationship between the confidence interval and the P value

If the absence of a link/ difference is within the confidence interval then the result is NOT statistically significant ($p > 0,05$), but If the absence of a link/ difference is outside the confidence interval then the result IS statistically significant ($p < 0,05$). For the ratio type indicators eg. RR, OR, HR, the absence of the effect is indicated by value 1 and for the difference type indicators eg. RA, ARR, means difference, or percentage difference, the absence of the effect is indicated by value 0.)

Searching for scientific articles using the PICO technique

Searching for therapy information:

P –the Patient/the problem to deal with (usually the name of the disease),

I – the intervention that we are interested in using to deal with the problem (e.g. a drug),

C – the Comparison intervention, the intervention with which we want to test against, it can be optional (e.g. another drug),

O – the Outcome (what we want to achieve with the interventions).

Searching for diagnosis information:

P –the Patient/the problem to deal with (usually the name of the disease),

I – the intervention that we are interested in using to deal with the problem (e.g. a diagnosis technique),

C – the Comparison intervention, the intervention with which we want to test against, it can be optional (e.g. another diagnosis technique),

O – the Outcome (what we want to achieve with the interventions).

Searching for prognosis (risk factors) information:

P –the Patient/the problem to deal with (usually the name of the disease),

I – the intervention that we are interested in using to deal with the problem (e.g. a risk factor),

C – the Comparison intervention, the intervention with which we want to test against, here it is optional,

O – the Outcome (what we expect to happen – e.g. a complication due to the risk factor).

Citing references using the Vancouver system

References for papers presenting original research in printed journals:

Example of a reference for a paper presenting original research in a printed journal:

Pompili M, Riccardi LR, Covino M, Barbaro B, Di Stasi C, Orefice R, et al. Contrast enhanced gray-scale harmonic ultrasound in the efficacy assessment of ablation treatments for hepatocellular carcinoma. *Liver Int* 2005;25(5):954-61.

Layout of references for papers presenting original research in printed journals:

Author's last name separated by 1 space from the initial(s) of his/her first name (*notes: between several initials of the same author no space or other separator should be used; several authors should be separated using a comma and 1 space; if the paper is signed by more than 6 authors, only the first six authors should be mentioned, followed by the expression et al*). **Title of the paper. Abbreviated name of the journal that published the paper**

(*note: find the official abbreviation of medical journals by accessing: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=journals>*) **year of publication;number of volume(number of issue):pages.** (*note: write the first and the last page of the paper, separated by a dash. When writing the last page do not write digits that repeat from the first page*)

References for papers presenting original research in electronic journals:

Examples of references for papers presenting original research in electronic journals:

Migeon BR, Axelman J, Jeppesen P. Differential X Reactivation in Human Placental Cells: Implications for Reversal of X Inactivation. *Am J Hum Genet* [Internet] 2005 [cited 2005 September];77:355-64. Available from: URL: <http://www.journals.uchicago.edu/AJHG/journal/issues/v77n3/42333/brief/42333.abstract.html>

Morse SS. Factors in the emergence of infectious diseases. *Emerg Infect Dis* [Internet] 1995 [cited 1996 Jun 5];1(1):[24 screens]. Available from: URL: <http://www.cdc.gov/ncidod/EID/eid.htm>

Layout of references for papers presenting original research in electronic journals:

Author(s) (*see previous layout for details*). **Title of the paper.** **Abbreviated name of the journal that published the paper** [serial online] **year of publication** [cited date (*date of today*)]; **number of volume(number of issue):pages or number of screens between square brackets.** **Available from: URL:** web address

References for books:

Example of a reference for a book:

Eisen HN. Immunology: an introduction to molecular and cellular principles of the immune response. 5th ed. New York: Harper and Row; 1974.

Layout of references for books:

Author(s). **Title of the book.** **edition, if other than first edition.** **Place of publication** (i.e. city or town where the publisher is located): **Publisher; year of publication.**

(notes: the above pieces of information may be retrieved from the outside and inside covers as well as from the first pages of the book; if the number of the edition is not stated, it is usually safe to assume the book to be at its first edition)

Example of reference for a chapter of a book:

Ptasznik R. Sonography of the shoulder. In: van Holsbeek MT, Introcaso JH (eds) Musculoskeletal Ultrasound. 2nd Edition. St. Louis : Mosby Inc; 2001:463-516.

Layout of references for chapter of books:

Author(s). **Title of the chapter.** In: **Editor(s).** **Title of the book.** **edition, if other than first edition.** **Place of publication** (i.e. city or town where the publisher is located): **Publisher; year of publication.**

Critical appraisal guide for scientific medical writing (articles, thesis)

Title:

- Too concise
- Too precise
- Length (no. of words is less than 15? – relative limit):

- Strong (forte) position (the first words – here the most important concepts should be stated)

Introduction:

- Presents the general aspect of the subject
- Presents the particular aspect of the subject
- The aim of the study is clearly stated at the end of the chapter
- No statements unjustified by references

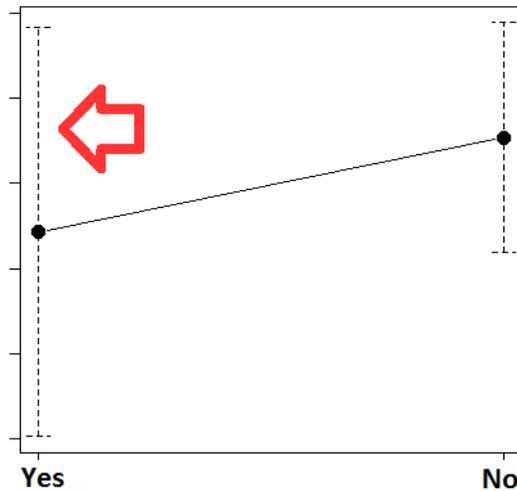
Material and methods:

- The authors explain on whom or on what they have performed their study
- Inclusion criteria are clearly stated
- Exclusion criteria are clearly stated
- The study design is presented (e.g. *series of consecutive cases, retrospective, prospective, randomized clinical trial*)
- The data collection type is presented (e.g. *case-control, representative sample, exposed-nonexposed*)
- Details regarding the sampling method are given
- Description of the sample (e.g. *age, gender, socio-professional characteristics, symptoms, morphology, etc.*)
- The authors explain what they tried to evaluate?
- The authors explain what they tested/compared/computed?
- The authors explain the evaluation criteria of their results and the validation methods of their results?
- The authors explain which statistical methods they have used?
- The tense of the verbs is past tense
- **ERRORS:** any comments, explanations, comparisons

Results:

- the authors present both positive and negative results
 - Nothing else than results
 - The tense of the verbs is past tense
 - **ERRORS:**
 - any comments, explanations, comparisons, allusions to the studied population or the method
 - marginal results, that are not linked to the aim of the study
 - references placed in this chapter
 - results of other studies
 - **Figures and tables:**
 - Understandable regardless of their context
 - Referred to in the text
 - Numbered in order of insertion
 - **ERROR:** Same information in both figures and tables (redundant)
 - **Figures**
 - have proper legends and properly defined axes, including units of measurement if appropriate
 - their titles are inserted below
 - numbered using indo-arabic numerals
- Specific criteria for specific types of graphs:**

- **Pie charts:**
 - categories should be indicated in the legend or the chart
- **Histograms:**
 - columns should be joined
 - the legend is useless
- **Graphs comparing averages:**
 - say what the chart bars represent (e.g. 1 or 2 standard deviations, 1 or 2 standard errors or the 95% confidence interval)
 - groups must be specified on the chart



- **Tables**
 - have their titles inserted above them
 - contain correct units of measurement in headers, if appropriate
 - overall and partial totals are coherent and add-up to 100%
- **Numbers** not too many decimals

Discussions:

- The authors explain whether the aim of their research has been reached or not
- main results are summarized
- the authors write a critical appraisal of the quality and validity of their results
- they discuss whether the number of subjects has been sufficient in order to draw conclusions
- they discuss whether the chosen method has been optimal in solving the research question
- they discuss why they have preferred it over other methods
- the authors discuss how results should be interpreted based on employed methods and the power of the employed statistical tests
- they compare the obtained results with those of other authors
- they express a critical appraisal of other similar works
- **ERRORS:**
 - repeating all results already presented in chapter *Results*
 - Repeating what has been said in chapter *Introduction*
 - Citing an author without mentioning him as a proper reference

Abstract: it must provide answers to the following questions:

- Why was the research performed

- How was the research performed
- What has been found
- What conclusions may be drawn
- The length of an abstract varies depending on the publishing journal (generally between 150-300 words)
- **ERRORS:**
 - the abstract contains references, figures, tables, abbreviations
 - it presents results that don't appear in chapter *Results*
 - it reports on the results of other authors

Regarding the whole article:

- **References** (the system depends on the publishing journal; generally, medical journals observe the Vancouver system):
 - References are inserted in the text, immediately after the corresponding statement
 - References should not appear in chapters: title, abstract, results.
- **Overall style:**
 - correct use of tenses (past tense for past actions, present tense for well-established notions)
 - objective, neutral tone (avoiding emotional expressions like “fantastic”)
- **Precision**
- **Clarity**

Oral presentation principles (article, thesis)

- **Structure of a scientific presentation** for an oral presentation of 8-10 minutes (the approx. time for your thesis defense):
 - title, authors (1 slide)
 - introduction – depending on the public (0-2 slides)
 - material and methods (1-2 slides)
 - results (2-3 slides)
 - only results
 - only the most important results
 - apply the KISS principle (Keep It Short and Simple)
 - discussions (1 slide)
 - conclusions (1 slide)
 - closing slide (1 slide)
- **Content of the presentation:**
 - **Introduction:**
 - General interest of the subject
 - Particular interest of the subject
 - Aim of the research (*clearly stated, using past tense*)
 - **Material and methods** (*presented using past tense*):
 - Note on whom the study has been performed
 - Inclusion criteria
 - Exclusion criteria

- Type of study: consecutive cases, cross-sectional, prospective, retrospective, randomized, ...
 - Data collection type: *case-control, exposed-nonexposed, representative sample*); Indicate the sampling method.
 - Describe the sample (*e.g. age, gender, socio-professional, ethnic, symptoms, morphology, etc.*)
 - What has been evaluated
 - What has been tested/compared/computed
 - Judgment criteria for the evaluation of results; methods used for validation of results
 - Statistical methods that have been used (*e.g. Chi square test, t-test, LogRank*)
- **Results** (*presented using past tense*):
 - Note both positive and negative results (not statistically significant)
 - Nothing else, but results
 - Do not insert: comments, explanations, comparisons to other results, marginal results that are not related to the stated aim of your research
 - **Figures and tables:**
 - Do not repeat information in both a figure and a table
 - Self-explaining, regardless of their context (correct titles, legends, axes, labels, etc.)
 - Note units of measurement and explain abbreviations
 - Ensure coherence of presented data (parts add up to totals, etc.)
- **Discussion:**
 - Note whether the aim of the study has been reached or not
 - Critically appraise the validity of your results
 - Appraise whether the number of subjects included in your study has been sufficient in order to apply the results to the target population of your study
 - Appraise whether the chosen study method has been optimal for solving the research problem or explain why it has been preferred over other methods.
 - How do you interpret the results based on the employed methods and the power of the employed statistical tests?
 - Compare your results with results obtained by other authors.
 - Do not repeat statements from the introduction.
- **Form of the presentation:**
 - **Slide text:**
 - Avoid a phrase if a sentence or idea will do!
 - Avoid a sentence if 1 or 2 keywords will do!
 - Not more than 7 lines (relative criteria) of text per slide
 - Ideally 1 idea per line.
 - **Illustration material:**
 - Content:
 - figures

- tables
- Use
 - What cannot be said otherwise
 - Materials that are difficult to listen to
 - Helps you stay in time (1 picture equals 1000 words)
- Choosing between figures and tables
 - Depending on your objectives
 - Preferably figures, curves, histograms
- **Practical creation of slides:**
 - The title
 - single line, if possible
 - underlined, bolded, different color
 - not numbered
 - The text
 - must be readable from the last row of chairs in the room
 - b/w or colored but with high contrast towards background
 - Figures and tables
 - must not be numbered
 - must have a complete title above them
 - must be self-explaining
 - verbal explanations should only help the public to understand
 - tables should not have more than 15 cells (max. 3-4 columns and 5-7 rows)
 - Images
 - patient pictures can only be used with individual consent from patients and while masking their eyes and mouth with dark strips or blurring, to ensure confidentiality.
 - Movies
 - very short and only if they are absolutely necessary
 - Avoid:
 - slide transitions (preferably none or very discrete)
 - text animation (especially if excessive, long and serves no purpose, it will get in the way of your oral presentation and it becomes annoying for the public)
 - abbreviations
 - unclear references to earlier introduced coding (e.g. group A, B, ...)
 - “elegant” / complicated expressions