A Primer in Research Methodology and Biostatistics

Bacârea Vladimir
Ghiga Dana
Măruşteri Marius
Oláh Péter
Petrişor Marius

UNIVERSITY PRESS
19 99
TÂRGU-MUREŞ

2014

A Primer in Research Methodology and Biostatistics

Authors:

Bacârea Vladimir

Ghiga Dana

Măruşteri Marius

Oláh Péter

Petrişor Marius

ISBN: 978-973-169-285-2

UNIVERSITY PRESS
19 99
TÂRGU-MUREŞ

2014

Table of Contents

Introduction

Medical research is a dynamic field in a continuous expansion. Knowledge of the necessary steps to carry out a high quality research is essential. Without prior knowledge of the basic concepts of research methodology and biostatistics is difficult to elaborate a research strategy.

There is no argument that a research starts from an innovative idea or a question coming from the researcher's curiosity, but from this point until the final stage of research, publication of results, many steps must be followed.

Basic notions of research methodology regarding the methods for bibliographic research or scientific hypothesis and protocol development must be known and understood by a researcher before starting a research study. Furthermore, it is essential to know how to define the study population and how to choose the proper means to select an adequate study sample. Also a very important step in conducting a medical research is data collection. Data must represent the reality and no bias should occur in this stage. Considering the need to choose the proper methods of data collection, the researcher should be familiar with the circumstances in which a certain method of data collection should be applied. Another important aspect in conducting a research is choosing the adequate type of study depending on the purpose and objectives of the study. Also data analysis and results interpretation are fundamental in order to be able to extrapolate the research findings to the entire population. It is also important that a researcher knows the notions regarding how to display the research results and how to write a scientific paper. These notions will help the researcher to complete the research project and to publish the research findings in order to be evaluated by other scientist or to help other researchers in their work.

A major aspect of the medical research that must be taken into consideration when conducting a research project is represented by ethics. In all research studies human or animal rights should be considered.

Biostatics is an essential tool in scientific research, especially in the medical field. If the statistical analysis is not adequate, biases can occur and might lead to major consequences in medical practice.

For those working in medical field, a science derived from mathematics could seem something very difficult. Since we all use statistical analysis software, this books tries to get beyond the mathematical details of each test and focuses on instructing the researcher what test to choose and when to choose it. This is a very important decision in a research study, because applying an inappropriate test most certain it will lead to errors in study results.

This book aims to help students, PhD students and novice researchers that are involved in conducting a medical research. Comprehensive and yet succinct this book offers basic notions regarding both research methodology and biostatics and hopefully will help the reader to have a broader scientific perspective.

The authors have equal contributions to the contents of this book.

Chapter 1 – Evidence-based medicine

Even from ancient Greece there are traces regarding evidence-based medicine, but only in 20th century this impact began to have impact in almost all fields of health care. A research group led by David Sackett and Gordon Guyatt were the first to establish the methodologies used to determine the best evidence in medical research and in 1992, Gordon Guyatt*et al.* were the first who wrote about the "evidence-based medicine" medical term.

David Sackett, in 1996, defined evidence-based medicine as the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. In other words it can be stated that evidence-based medicine represents the medicine that appeals to the scientific documented methods.

Several reasons led to the introduction of evidence-based medicine, which include the necessity of knowledge and to apply new methods in medical practice, the extremely high volume of medical journals and the doctors' limited time to browse all these medical sources.

The evidence-based medicine practice is defined by Sackett as integrating individual clinical expertise with the best available external clinical evidence from the systematic research (Figure 1).

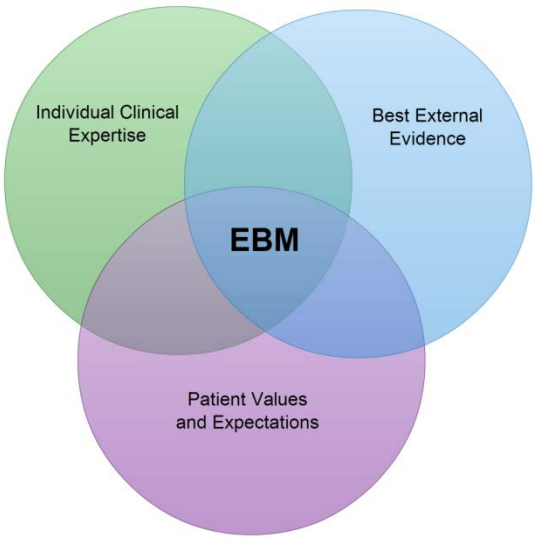


Figure 1: The Evidence-based Medicine Triad

It can be stated that evidence-based medicine represents the medicine based on methods that are scientifically documented.

In order to apply evidence-based medicine, besides time, expertise and experience, several steps are required:

- Translation of the medical probleminto an answerable question;
- Systematic search of relevant records to answer the question;
- Critical evaluation of the internal validity of the founded evidence:
    - Systematic errors;
    - Quantitative aspects of diagnosis and treatment;
    - The effect size and aspects regarding its precision;
    - Clinical importance of results;
    - External validity or generalizability.
- Application of relevant medical records in practice;
- Evaluation of their efficiency and effectiveness.

There are two sources of evidence:

- People: attendance and participation to national and international conferences, accessing online networks and participating in professional and interprofessional conversations;
- Literature:
    - Primary literature: original sources of information (research fundings);
    - Secondary literature: any information that is a reconsideration of primary data (systematic reviews, guidelines, editorials).

Regarding research fundings, not all research is valued to the same degree, and so hierarchies of evidence were developed. One of the most well-known hierarchy of evidence is the one published by Guyatt *et al.*(Figure 2).

One other system to stratify evidence by quality has been developed by the United States Preventive Services Task Force for ranking evidence about the effectiveness of treatments or screening:

- Level I: Evidence obtained from at least one properly designed randomized clinical controlled trial.
- Level II-1: Evidence obtained from well-designed controlled trials with no randomization.
- Level II-2: Evidence obtained from well-designed cohort or case-control analytic studies, preferably from more than one center or research group.
- Level II-3: Evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled trials might also be regarded as this type of evidence.
- Level III: Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees.

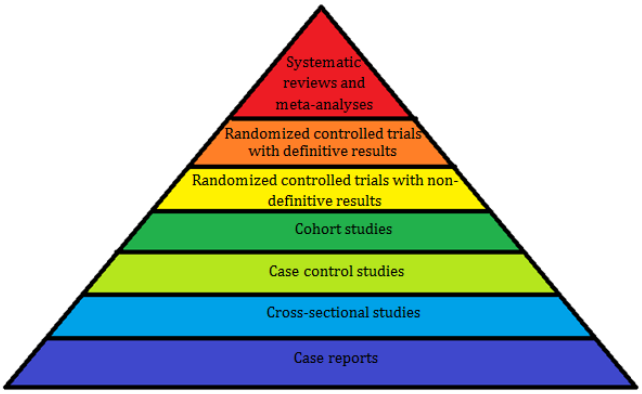


Figure 2:Hierarchy of evidence

The majority of the evidence ranking schemes focuses on evidences regarding therapy and prevention. The Oxford CEBM Levels of Evidence addresses this problem and provides "levels" of evidence for demands about prognosis, diagnosis, treatment benefits, treatment harms, and screening.

Another, newer system, that takes into account more dimensions than only the quality of medical research, was developed by the GRADE (The Grading of Recommendations Assessment, Development and Evaluation) working group.

Logistics evidence-based medicine is structured on three levels:

- Organizational level:
    - Cochrane Collaboration – the most well-known organization that conducts systematic reviews regarding the latest news from various medical fields, which are published in Cochrane Library. In 1979 Archie Cochrane criticized the medical society of that time for lack of ways to provide to doctors the latest medical research results. As response to Cochrane's call, Cochrane Collaboration was founded in 1993 and now consists of a group of more than 31,000 volunteers in more than 120 countries.
    - AGREE (The Appraisal of Guidelines for Research and Evaluation) Collaboration – is an international organization dealing with the implementation of clinical practice guidelines. The original AGREE Instrument has been updated and methodologically refined. The AGREE II is now the new international tool for the assessment of practice guidelines. The AGREE II is both valid and reliable and comprises 23 items organized into the original 6 quality domains.
- Technological and financial level: is provided from research grants, the expenses are necessary for the creation and updating of primary and

secondary evidence sources, implementation and database administration, database access and staff training.

- Human resources:
    - Specialists– who elaborate the medical evidences.
    - IT personnel – deals with IT support for the creation and management of evidences and computer-aided medical decision software.
    - People involved in the development of health policies.
    - Medical practitioners – integrate the available valid evidences in current practice along with individual experience.

Chapter 2 – Philosophy of science

Anessentialassignment of the biomedical science is to find the causes of phenomena, such as disease.

A major breakthrough was accomplished in 1880s when Robert Koch (1843–1910) defined his guidelines (Koch's postulates) to determine if a microorganism is the cause of a disease. According to Robert Koch if it can be demonstrated that the presence of a parasite is not due to a random accident, then that parasite can be considered as the cause of a disease.

Initially, The Koch's postulates that he presented at the Tenth International Congress of Medicine in Berlin in 1890 were:

- The parasite occurs in every case of the disease in question and under circumstances which can account for the pathological changes and clinical course of the disease.
- The parasite occurs in no other disease as a fortuitous and nonpathogenic parasite.
- The parasite, after being fully isolated from the body and repeatedly grown in pure culture, can induce the disease anew.

Today the Koch'spostulates are known to be:

- The organism must be found in all animals suffering from the disease, but not in healthy animals.
- The organism must be isolated from a diseased animal and grown in pure culture.
- The cultured organism should cause disease when introduced into a healthy animal.
- The organism must be reisolated from the experimentally infected animal.

At the time when Robert Koch formulated the postulates, they were essential for the progress of infectious diseases knowledge, but these postulates are elusive, and cannot be applied in certain instances.

Sir Austin Bradford Hill (1897–1991) a remarkable pioneer in medical statistics and epidemiology published in 1965 an article which had a major impact on epidemiologists and medical researchers. In this article Hill presented nine considerations for a causal association. Very often these considerations were applied by other researchers as a checklist of criteria, althoughHill himself did not consider in this way.The nine considerations, or viewpoints as Bradford Hill called them,are:

- Strength of association: as stronger the association is, the more less is likely that the association relationship between exposure and disease is due to chance or a confounding variable.
- Consistency of the observed association: has the association been observed by different people, in different places, circumstances and times?
- Specificity: the association relation between a risk factor and a disease supports causality if the association is limited to certain individuals, places and types of disease, and if there is no association between the risk factor and other modes of dying.
- Temporality: the risk factor in question must precede the outcome by a period of time.
- Biological gradient: there is a gradient of risk associated with the degree of exposure (dose–response relationship).
- Biological plausibility: refers to a plausible explanation regarding the association between exposure and disease based on known facts that describes hoe the exposure may alter the risk of disease.
- Coherence: the observed association should not be in contradiction with already known facts about the disease natural history and biology.
- Experiment: the strongest support for causation may be obtained through controlled experiments (clinical trials, intervention studies, animal experiments).
- Analogy: in some cases, it is fair to judge cause–effect relationships by analogy. For example if a drug from a certain class of drugs causes a specific disease it easy to presume that another drug from the same class of drugs will cause a similar effect.

If it is to determine the causal link between a specific factor (such as cigarette smoking) and a disease (such as emphysema or lung cancer), the nine criteria would be:

- Strength of association. The lung cancer rate for smokers is quite a bit higher than for nonsmokers.
- Consistency. Different methods (e.g., prospective and retrospective studies) produced the same result.
- Specificity. In smokers, death rate due to lung cancer ishigher than death rates due to other causes.
- Temporality. Smoking in majority of cases precedes the lung cancer debut.
- Dose – response relationship. Data shows that death rate from lung cancer rises linearly with the number of cigarette smoked daily.
- Theoretical plausibility. The biological theory that statesthat smoking is the cause of tissue damage which over time will lead to cancer in the cells is a very plausible explanation.

- Coherence. The conclusion, that smoking will cause lung cancer,is understandableknowing the lung cancer biology and history.
- Experimental evidence. In other experiments is shown that tobacco tarproduce cancer, so carcinogens are present in tobacco tar.
- Analogy. Experimental studies on animals show that between induced smoking and lung cancer is causal relationship, then it is easy to assume the same conclusion on humans also.

Even if these criteria are more complex than Koch's postulates, there are, still, cases where it might be a causal relationship, but where criteria do not apply. The Bradford–Hill criteria show that causation in the biomedical sciences is not deterministic, but is more of a mixture of more general criteria.

Conditions for causation:

- *necessary* but *not sufficient condition* for an event
    - multiple factors are prerequisites for something happening, but no one of them alone is sufficient

For example, having a leg is a necessary condition for having an inflammation of it, but having a leg is not said to be the cause of the inflammation. In this case, multiple factors are prerequisites for something happening, but no one of them alone is sufficient.

- *sufficient* but *not necessary* condition
    - if a sufficient condition is present, we know that the event (the effect) will occur

For example, when a person develops leukemia after being exposed to ionizing radiation known to be of the sort and strength that results in leukemia, we tend to say that the radiation caused the leukemia. The leukemia could, of course, have been caused by other factors (sufficient conditions) in the absence of the radiation. Nevertheless, if a sufficient condition is present, we know that the event (the effect) will occur.

- Each of the factors is an *insufficient* but *necessary* part of a *sufficient* and *necessary* condition for the event

For example, just being stung by a wasp or just being hypersensitive to wasp venom will not cause an anaphylactic reaction. However, acting together, both may be sufficient and necessary for an anaphylactic reaction to develop, so both are said to cause the event. Each of the factors is an insufficient but necessary part of a sufficient and necessary condition for the event to occur (Figure 1).

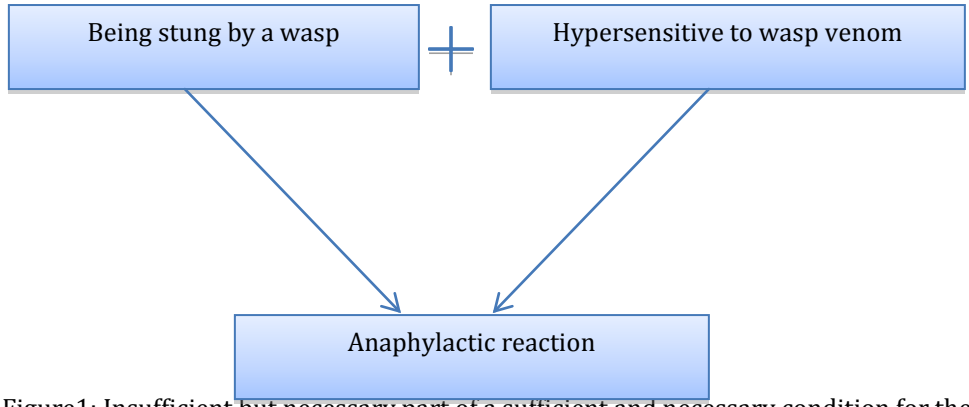


Figure1: Insufficient but necessary part of a sufficient and necessary condition for the event to occur

- The factor is an *insufficient* and *non-redundant* part of an *unnecessary* but *sufficient* condition for death

For example, when a person drinks a lethal dose of poison, if no antidote is administered andno gastric lavage is performed,the person dies. Only drinking poison is not a sufficient condition to cause death.Many people drink poison without dying, because gastric lavage was performed or antidote is administered. However, drinking poison is part of a series of conditions that acting together are sufficient to cause death (Figure 2). Drinking poison, in this situation, is non-redundant, because if poison is not ingested, in the same circumstances, deaths do not occur.

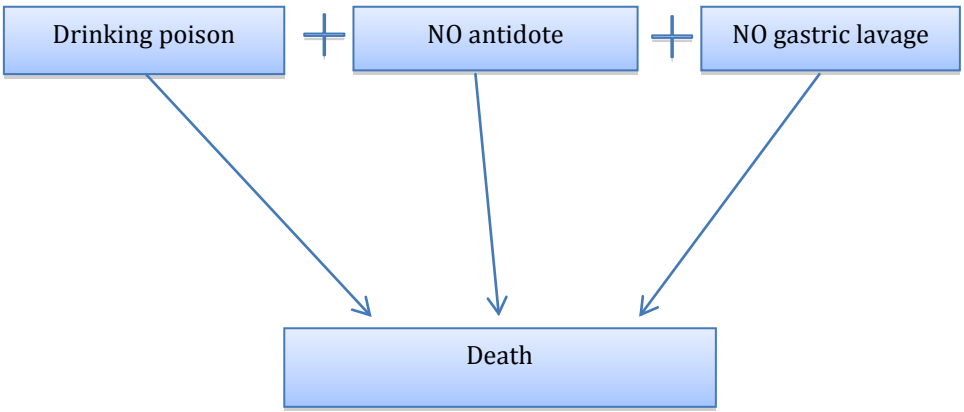


Figure 2: INUS condition

In this case, drinking poison is an insufficient and non-redundant part of an unnecessary but sufficient condition for death. This is called an INUS condition (Mackie 1974).

Another situation is represented by the relation between smoking and lung cancer. In this case the causal link is represented by the increased probability that lung

cancer will develop among smokers. This situation represents a probabilistic approach to causation.

The probabilistic theories regarding causationfocus on the idea that causes increase the probability of their effects to occur.

From the time of Aristotle, philosophers have realized that a distinction could be made between two kinds of scientific knowledge:

- knowledge *that* – descriptive
- knowledge *why* – explanatory

For example it is one thing to know that myocardial infarction is associated with certain kinds of pain (angina pectoris) and it is a different thing to know why this is so.

According to Salmon (1990), the explanatory knowledge is the one which provides scientific understanding of the world.

In the biomedical science something is explained by showing the way by which it is expected to happen according to the laws of nature (nomicexpectability) (Hempel 1965).

There are three models of explanation according to Salmon (Table I):

Deductive–nomological model (DNM) – a singular event is explained if the initial conditions and the universal laws of science are followed. For example the increase in the number of erythrocytes in high altitudeis explained by the laws that govern the increase in the number of erythrocytes and the initial conditions regarding medium (decrease of oxygen concentration in the air).

Deductive–statistical model (DSM) – is similarthe deductive–nomological model. Something is explained by the initial conditions and the statistical laws. For example: if the initial condition is represented by "having tonsillitis" and the statistical laws state that "antibiotics administration will lead to recovery", then the conclusion is "People treated with antibiotics will recover". This model cannot explain singular events only general regularities.

Inductive–statistical model (ISM) –explains a singular eventfrom statistical laws.For example: if a person having tonsillitis and is taking antibiotics and the statistical laws states that "probability of recovery in tonsillitis when antibiotics are administered is approximately 1", then the conclusion is "that person who has tonsillitis and is taking antibiotics will recover".

Table I: Models of explanations according to Salmon (1990)

Laws	Singular events	General regularities
Universal laws	DNM	DNM
Statistical laws	ISM	DSM

Modes of inference

Charles Sanders Peirce (1839–1914), an American scientist was described asan innovator in philosophy and research methodology, was the first to examine the modes of inference in the biomedical science. The modes of inference are: deduction, induction and abduction.

Deduction involves relatingfrom general rules to specific cases.For example if all people with glaucoma present atrophy of the optic nerve and Miss. Peterson has glaucoma, then Miss. Peterson presents atrophy of the optic nerve.

Induction involvesrelatingfrom multiple cases to a general rule. For example if all people observed with atrophy of the optic nerve presents symptoms of glaucoma and all people observed are from the general population, then all people with atrophy of the optic nerve presents symptoms of glaucoma.

*Abduction*involves an observation which leads to a possible hypothesis to prove. For example, if Miss. Peterson presents atrophy of the optic nerve and all people with atrophy of the optic nerve presents symptoms of glaucoma, then Miss. Peterson has glaucoma.

The differences between these three modes of inference are illustrated below (Figure 3, 4, 5).

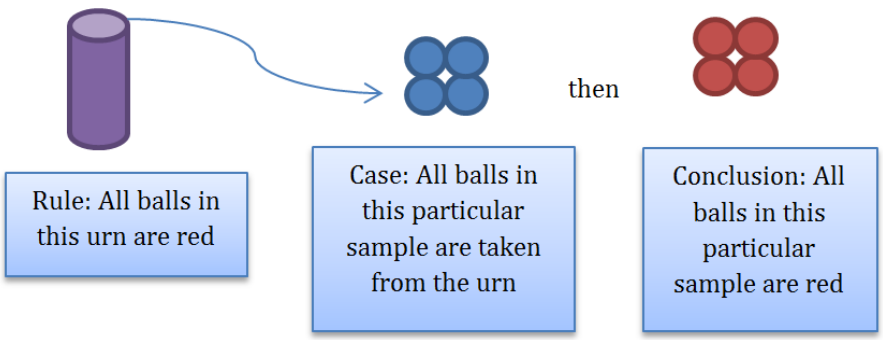


Figure 3: Deduction

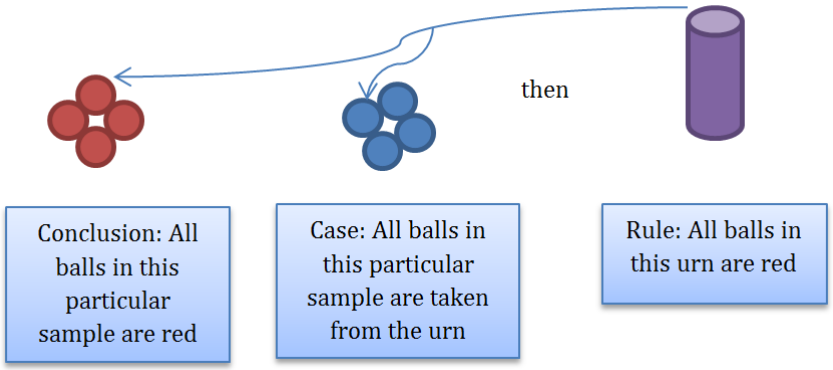


Figure 4: Induction

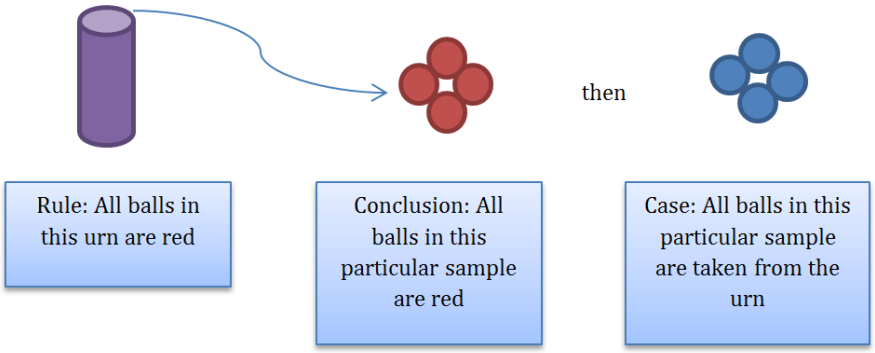


Figure 5: Abduction

Theory testing

There are three challenges in medical treatment and research:

- The obtained effects may be due to luck or accident (and not intervention);
- Theobtained effect occurs even if there is no intervention;
- The effect may notbe obtained despite intervention.

In the hypothetical–deductive model the scientific hypotheses are tested by predicting particular events that can be observed, and then observe if the events will occur as predicted. In the situation when the events occur as predicted, the scientific hypothesis is confirmed, and if not the hypothesis is rejected. The steps of the hypothetical–deductive model are:

- State a clear testable hypothesis;
- Deduce the empirical consequences of this hypothesis.

- Perform empirical experiments (in order to compare their results with thededuced empirical consequences);
- Comparison of experimental findings with derived empirically consequences.If the results coincide the hypothesis is confirmed, if not the hypothesis is rejected.

Chapter 3 – Scientific Medical Research

Attempts to detect correlations between health or disease and risk factors, to assess the mechanisms by which these factors act, to find cures or preventive drugs, arises a series of questions and the answers to these questions are brought by the scientific research.

3.1 Historicalreferences

"We see further because we stand on giants shoulders"

Isaac Newton

In ancient Greece the diseases treatment was based more on philosophy than on actual understanding of human anatomy. Surgical procedures were rare and human dissection was not an accepted practice as result physicians had little first-hand information about the inside of human body. It wasn't until the Renaissance that the science of human body was born. The Belgian physician Andreas Vesalius (Figure 1A) shocked many by deciding to investigate anatomy by dissecting human bodies. He decided to publish his discoveries and the result was the book entitled "De HumaniCorporisFabrica" (Figure 1B) describing the structure of human body. The book was first published in 1538 and considered one of the greatest books in medical literature. It is also regarded one of the greatest discovery in medicine because it contains the first precise description of the interior of the human body. Because of Vesalius the study of human anatomy through dissection became an essential component in medical training.

Figure 1: A. Andreas Vesalius; B. De HumaniCorporisFabrica

Regarding blood circulation, in the early part of 17 century, how blood works in the body was misunderstood. The dominant theory was that blood is pushed and flows through the heart by the existing pores in the soft tissue in the body. English physician William Harvey (Figure 2) he was fascinating with the working of the heart, studying it by dissecting animals. In his dissection, Harvey observed that the heart had one way valves that kept blood flowing in one direction. Some valves let the blood in, while others let it out. His great discovery was that the heart was pumping blood into the arteries when thencirculated through the veins coming full circle back into the heart. Harvey discovery led to major progress in anatomical research and surgery.

Figure 2: William Harvey

Another great discovery regarding blood occurred in Vienna in 1900. An Austrian physician Karl Landsteiner mixed samples of blood from different individuals and studied the effects. In some cases samples mixed safely, but in other combinations the blood clumped and became sticky. Landsteiner discovered that clumping occurred when certain proteins called “antibodies” in the recipient’s blood bounded with other proteins called “antigens” on the donor’s red blood cells. After this discovery he determined that human blood can be divided in four distinct groups (A, B, AB, 0). He realized that blood transfusion can be carried out safely only when people receive blood from someone who shares the same blood group. Landsteiner’s discovery made possible the organ transplant in the 50’s.

Also the need for blood replacement has been researched and over the years many potential substitutes were investigated: milk, salt or saline solutions, hemoglobin and animal plasma and Ringer’s solution. Although these solutions were considered a breakthrough at their time, none of them is a true blood substitute. The latest bearer of hope for a potential breakthrough in developing artificial blood comes from a Romanian researcher, RaduSilaghi-Dumitrescu, a professor at Babes-Bolyai University from Cluj-Napoca. In 2013, the researcher stated that he and his team successfully transfused a blood substitute into mice without any record of any ill effects.

Who discovered anesthesia is still a matter of debate. In 1840, two dentists from Boston, Horace Wells and William Morton, and from Georgia, Crawford Long, experimented with two chemicals, of which was believed to have the ability to reduce pain, nitrogen dioxide or nitrous oxide (laughing gas) and ether, a liquid mix of alcohol and sulfuric acid. All three doctors claimed to discover anesthesia.

A turning point in the history of medicine was the discovery of X-rays. Wilhelm Conrad Röntgen (Figure 3A), a German physicist, in 1895, conducting experiments with a cathode ray tube, had accidental discover a radiation unknown to science, which he called X ray (Figure 3B). Roentgen discovery led to a revolution in medical diagnostic.

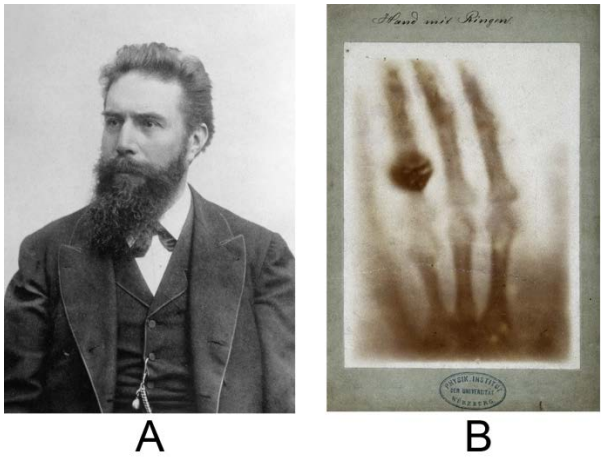


Figure 3: A. Wilhelm Conrad Röntgen; B. Röntgen's first X-ray

Some discoveries, like the X-rays, were made by accident, while others were developed through time. In Vienna, in 1846, many women were dying due to childbed fever or puerperal fever. Doctor IgnazSemmelweis, noticed that in the wards where mothers were assisted by physicians, the percentage of death due to childbed fever was higher than in the wards where mothers were assisted by midwives. He also noted that the physicians also conducted the autopsy of the mothers, and after they would go back and deliver babies or examine other pregnant women, without washing their hands. To find out if the physicians were passing from an infected woman, an invisible matter, through their hands, to the other patients, he conducted an experiment. He asked his students to wash their hands in chlorine solution before assisting to a baby delivery. The result was the percentage of deaths due to childbed fever dropped. After this experiment Semmelweis realized that infectious disease has a single cause, and if the source of infection is eliminated the disease does not occur. But at that time the connection between bacteria and infections, so Semmelweis idea was ignored. Only after 10 years, Louis Pasteur, a French chemist and microbiologist, was determined to find out the cause of infectious diseases. It all started when Pasteur was trying to find out what was spoiling the wine production of the country. He discovered that the spoiled wine was contaminated by microorganisms, germs, and these germs were

causing the wine to sour. But through a simple heat treatment, he showed that the germs can be killed and the wine saved, and so the pasteurization process was born. Starting from what he discovered in wine industry, Pasteur conducted a series of experiments and demonstrations to prove that germs are the cause of specific diseases and this led to his great discovery, the germ theory. The central idea of this theory is that one microorganism causes one disease in every body. This theory marks the beginning of modern medicine.

Louis Pasteur is regarded as one of the three main founders of bacteriology, together with Ferdinand Cohn and Robert Koch, and is popularly known as the "father of microbiology".

Robert Koch discovered the Tuberculosis bacillus for which he received the Nobel Prize inPhysiology or Medicine in 1905. His research led to the creation of Koch's postulates, a series of four generalized principles linking specific microorganisms to particular diseases which remain today the "gold standard" in medical microbiology.

In 18 century, smallpox killed an estimated 40 million people all around the world. A country doctor, Edward Jenner, discovered than in an English village, some locals who were involved with the dairy business were immune to smallpox, because they already had been infected by cowpox. In 1796, during an outbreak of smallpox, Edward Jenner decided to conduct an experiment. He extracted pus from the wounds of a woman who worked at a dairy factory and then he inoculated an 8 years old healthy boy with the cowpox virus. In the days that followed, the boy developed a slight fever and some cowpox blisters, and then recovered. Six weeks later, Jenner inoculated the boy with the smallpox virus. After a few days he noticed that the boy was healthy, resistant to smallpox. Vaccination to smallpox was revolutionary and it was the first time a man-made product had been used actively to prevent a disease before it occurred.

Louis Pasteur, 50 years after Jenner's discovery, pushed forward the concept of vaccination, developing vaccines for rabies in human and anthrax in sheep, and in the 20th century, Jona Salk and Albert Sabin, independently, developed vaccines againstPoliomyelitis.

In World War I, more than 10 million people died and many from the infection of their wounds. After the war the research intensified to find safe methods for repellingthe bacterial invasion. Scottish physician, Alexander Fleming, while studying staphylococcus bacteria, he observed on the culture dish, the growth of a mold, *penicilliumnotatum*. He noticed that the bacteria around the mold had died which led him to speculate that the mould was producing a lethal substance for the bacteria. He named the substance penicillin. Fleming, for the next several years, tried to extract penicillin and applying it to treat infection, but he was unsuccessful and he eventually gave up. In 1935, scientists, Howard Florey and Ernst BorisChain, from Oxford University, found Fleming's writings and decided to carry on his work. They successful

extracted and purified the penicillin and in 1940 they tested it by injecting 8 mice with lethal doses of streptococcus. After that, 4 mice were injected with penicillin and within hours, the 4 mice which weren't injected with penicillin, died, but 3 from those who had been given penicillin were alive.

In 1945, Sir Alexander Fleming (Figure 4), Howard Florey and Ernst Boris Chain shared the Nobel Prize in Physiology or Medicine for the discovery of the first antibiotic.

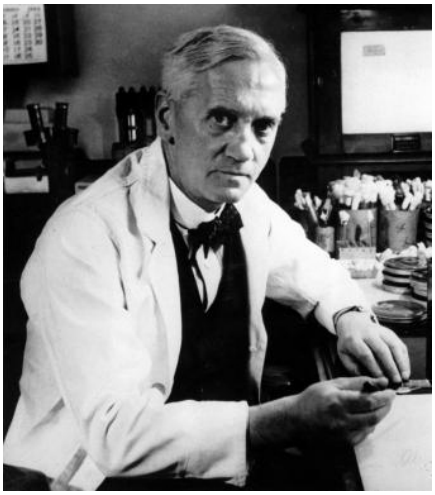


Figure 4: Sir Alexander Fleming

Pathologist Gerhard Domagk, in 1932, was studying the potential medical applications of some new medical dyes. He worked with a new synthetic chemical dye, called Prontosil, which he injected in lab mice that were infected with streptococcus bacteria. The dye attached to the bacteria and inhibited its growth and the spread of the infection stopped. Because Prontosil contained sulfanilamide molecular structure, it was called a "sulfa drug", first of its kind, a synthetic chemical substance that can cure and prevent bacterial infections.

In 1939, Domagk received the Nobel Prize in Medicine for this discovery, the first drug effective against bacterial infections. He was forced by the Nazi regime to refuse the prize and was arrested by the Gestapo for a week, but after the war, in 1947, Domagk was finally able to receive his Nobel Prize.

AIDS is one of the worst epidemics in modern history. The first clues about the disease emerged in the early 80's. The physicians reported a large number of deaths due to rare infections and different types of cancer. Blood sample of the patients showed an extremely low level of CD4T lymphocytes, the white blood cells vital for the immune system of the body. In 1982, the Center of Disease Control and Prevention, gave the disease a name: AIDS (Acquired Immunodeficiency Syndrome). Luc Montagnier, from the Pasteur Institute in Paris, and Robert Gallo, from the National

Cancer Institute from Washington D.C., shared the credit for the discovery of HIV (Human Immunodeficiency Virus), the cause of AIDS.

Claude Bernard, a French physiologist, was one of the first to suggest the use of blind experiments to ensure the objectivity of scientific observations, and he was the first to define the term“milieu intérieur”, now known as homeostasis.Among many other accomplishments, Claude Bernard performed cardiac catheterization in an experimental study, on a horse. In this study, from the jugular vein and carotid artery, performing a retrograde approach Claude Bernard entered in both the left and the right ventricles. After this study an era of investigation in the cardiovascular field started.

The first cardiac catheterization performed on a living person is attributed to Werner Forssmann, who at the age 25, inserted a 65 cm catheter into one of the left antecubital veins of his own forearm and guided it fluoroscopically until it entered his right atrium and afterwards took an X-ray picture of it (Figure 5). For this achievement, Forssmann shared the Nobel Prize in Physiology or Medicine in 1956 with André Frédéric Cournand and Dickinson W. Richards.

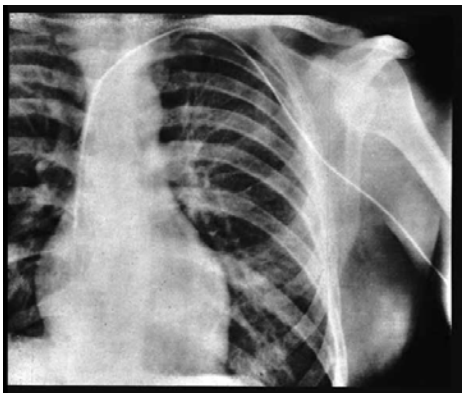


Figure 5: Werner Forssmann’s X-ray showing cardiac catheterization

3.2 Purpose, justification and qualities of a medical study

The late, undeniable technical progresseshad favorable repercussions on the general evolution of human society, but, also have contributed to the emergence of many risk factors that can influence health.

It can be stated that a research project based on which a clinical-epidemiological study will be carried out has its origins in:

- Researcher’s curiosity for a certain medical problem;
- The attempt to describe some inconsistencies between what is observed and what should happen in medical practice;

- Researcher's qualities regarding his abilities to observe and his professional experience;
- Researcher's ability to find the most effective methods for solving the arising problems.

3.2.1 Purpose of the study

The purpose of a research study is to give answers to questions regarding:

- Causation – Documenting a close relationship between one or more risk factors and disease;
- Description of disease natural history or dynamics –Events that occur in the natural history of a communicable diseaseare grouped into four stages: exposure, infection, infectious disease, and outcome (Figure 6).

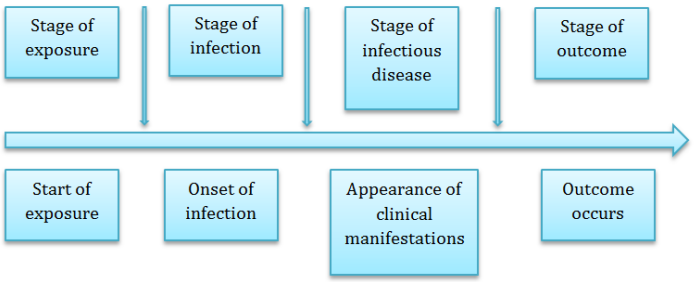


Figure 6: Stages in the natural history of communicable diseases.

The outcome can be represented by recovery, immunity, chronic state or death.If in infectious diseases these phases, especially the first two are better defined, in the non-communicable diseases the onset is imprecise, and the latency period is usually asymptomatic, so it is harder to detect the disease in these stages.

- Disease prognosis – The assessment of disease evolution close to certainty. Prognostic factors are appreciated through epidemiological cohort studies that can evaluate the indicators such as: a patient survival 5 years after a key moment in the natural history of the disease, fatality rate, the rate of response to treatment, relapse and remission rates of disease.
- Disease treatment – The efficiency and effectiveness of treatment are common research topics, for theserandomized controlled clinical trials are conducted. It is also possible to analyze new scheme of treatment, the effectiveness of some drugs or application schemes for prophylactic drugs (vaccine, disinfectants, etc.).

- Disease prevention – Includes all measures designed to influence modifiable causes of disease and carried out in four stages: foremost, primary, secondary and tertiaryprevention.

3.2.2 Study justification

The study justification begins with the report of available information related to the theme that the future research intends to debate. These information can be found in the literature. Bibliographical research can identify shortcomings regarding the subject under study, the lack of data on a disease in a geographical area (country, county, community), explain and express solutions to rectify or improve the issue. All these may be elements that can justify the initiation and the management of a study.

3.2.3 The qualities of a research study

A research study must be:

- **Pertinent**. This quality is fulfilled if the answer of the question or of the theme that is under study has repercussions in the future (a better knowledge of the disease natural history, prognosis, an improvement in the diagnosis and therapy)
- **New**. The conducted study must bring new information in the medical research.
- **Feasible**.For a study to be feasible, before initiating the study the researcher must ensure that:
    - The study is accessible and acceptable to the population;
    - There is sufficient number of subjects, laboratory determinations, in order to be able to draw appropriate conclusions;
    - There are no financial or technological constraints;
    - The study must not have too ambitious purposes or multiple purposes. In most of the cases, a research study should have one purpose, to find answer to only one question;
    - The study can be conducted in an adequate period of time and with an acceptable financial cost.
- To respect **medical ethics**. The four ethical principles must be followed:
    - The interest and benefit of the medical research;
    - Harmless of the research;
    - To respect theindividuals included in the study (respecting individuals' privacy, confidentiality);
    - To preserve fairness and respect for moral values.

Chapter 4 – Strategies and Planning a Research Study

In conducting a research study it is required to fulfill the following steps:

- Planning theresearch study;
- Protocol development;
- Data (information) collection;
- Data analysis;
- Analysis and interpretation of the results;
- Displaying results;
- Publishing the study.

4.1 Planning the research study

In order to prepare a research study a multiple steps are required to be followed:

- Bibliographical research
- Formulating a research problem
- Scientific or research hypothesis
- Study motivation
- Formulating the research topic

4.1.1 Bibliographical research

The body of scientific literature is extensive and overwhelming.Bibliographical research mission is to achieve a balance between the excess of medical documents and the loss ofrelevant information and implies:

- Research documentation on the study theme;
- Critical reading of the selected publications;
- Combine the acquired data from different scientific articles, on the same subject, into coherent information.

Bibliographical research requires a selection and evaluation of the literature on a certain problem and is achieved by following three steps:

I. The first step is **proper bibliographic research**, namely identifying publications related to the research problem.This step provides foundational knowledge about the problem area. The review of literature also helps a researcher to

find, on a specific research area,what studies were carried out, how the studies were conducted and what conclusions were drawn.

Selection of information, choosing those scientific articles that match the theme of the research is done by reviewing first of allthe title of the article. If the title corresponds with the search terms and is useful, than the relevance of the study must be checkedbyreviewing the article's abstract in order to determine whether the results can be applied in practice, and whether the methods used are similar or identical to those that will be used in the future research study. If all this criteria are fulfilled, the reader will evaluate the article in detail, if not he will pass to another article.

Bibliographical research stages are:

- Defining and precise delimitation of the research topic
- Establishment of a "keywords"list
- Keywords translation in international languages
- Establishment of a list with documents that will be evaluated
- Reference booksreview
- Primary publicationsreview
- Editingbibliographic records and arranging them in alphabetical order
- Inserting references in text

Primary bibliographical sources

Primary bibliographical sources are represented by all written materials containing original studies:

- Scientific articles: their main quality is that the published information are the latest one on a specific subject.
- PhD theses: useful because in such documents is described extensively a certain theme.
- Patents and trademarks: useful in experimental research.

The medical journals cover the most up to date information, so is better to review as many journals as possible, even though the number of these publications depends upon the field of the research; certain specialties have more journals than others.

Secondary bibliographical sources

The secondary bibliographical sources are represented by:

- Specialist academic or professional booksand medical guidelines textbooks: useful in the early stages of a study. These publications contain definitions, basic notions regarding a certain medical specialty.
- Medical reference books: represented by monographs, writings on a single subject or an aspect of a subject.

Books represent an important part of any bibliography and as other forms of documentation, have advantages and disadvantages. The most significant advantage is that the information published in books is important and of good quality and represents multiple findings from different researches integrated in a body of knowledge. The biggest disadvantage is that the material is not completely up to date due to fact that it requires a couple of years from completion of a book until its publication.

Tertiary bibliographical sources

These bibliographical sources are represented by indexes, medical journals containing only the articles abstracts.

Online bibliographical research

This type of research allows access to countless sources of information using the search engines. In scientific research, online research besides the use of browsers, search engines and knowledgeregarding the use of predefined keywords and of the logical operators, involves also the access of online medical sources and the use of MeSH terms (see Chapter10).

The most well-known online medical sources are:

- Medline – http://www.nlm.nih.gov/pubs/factsheets/medline.html
- Pubmed– http://www.ncbi.nlm.nih.gov/pubmed
- Elsevier– http://www.elsevier.com/
- ScienceDirect – http://www.sciencedirect.com/
- SpringerLink – www.springer.com
- DOAJ (Directory of Open Access Journals) – http://www.doaj.org/

The main disadvantage with online bibliographic research is represented by the validity of the information, because is not very complicated for someone to produce a Web page. The following criterion helps to differentiate a good information from a bad one:

- *Authority* exists if on the Web page is listed information about the author and his competence in writing on that specific topic, and if the page address has a preferred domain such .edu, .org or .gov.

  Information about the publisher of the Web document can be found by accessing the “About us”, “Mission”, or “Who we are” sections.

  The URL address shows the page affiliation. A URL ending with:

  - .edu is from an institution of higher education
  - .gov is from some branch of the federal government
  - .org is from a nonprofit organization

    - .com is from a commercial vendor
    - .net from anyone who can afford to pay for space on a server
- *Accuracy* exists if on the page is listed the author and institution affiliation and a way of contacting him or her. Also it must be identified the purpose of the information (news, advertising, public announcement), and if the information is the report of a research study.
- *Objectivity* is obtained when on the Web page there is no or little advertising and the information is accurate and objective. In this case it must be identified if there are any biases in the presented information. The following can suggest biases:
    - If the information cannot be traced to the actual information presented in bibliographic or Internet reference
    - If the authors express their opinion
- *Currency* exists if the page and any links that the page might provide are updated regularly. In this case it must be determined when the page was created and when it was last updated.
- *Coverage* exists when the information on the Web page can be seen without paying any fees or being necessary to install additional software.

The gathered information can be organized in 3 ways:

- Bibliographical notebook: records the information in chronological order;
- Bibliographical sheets: records the information source, journal
- Electronic bibliographical sheets: databases management systems, which use special programs to organize the references.

II. The second step of the bibliographical research is represented by the **critical reading** of the selected publications in order to judge their value regarding either the quality of the research or the relevance of the published results.

When evaluating a publication, the reader must assess:

- The internal validity – if the results correspond to reality.
- The external validity – if the published information can be applied in medical practice, **can** be generalized to the entire population and if it can be applied in the future research.

Critical reading of the medical literature is based on using a reading grid containing 8 evaluation criteria.

1. The *aim* can regard:
    - Disease natural history (occurrence and distribution), evolution (progression over time) and prognosis
    - The performance of a screening and a diagnostic test.
    - The effects of treatments or preventive efforts.
    - The identification of the cause of disease (aetiology).
2. The *type of the study* must be correspondent with the aim of the study.
3. The *studied factors*. The reader must recognize the studied factors which can be exposure or intervention factors. Also to assess the methods used to measure the factors and if it was used the same method for the studied groups, the methods used for comparison and if there are any biases.
4. The *outcome*refers to the event, state or condition caused by the studied factor and is represented by: disease, death, complications, relapses, survival. The methods used to measure the outcome must be described, so the reader can assess the validity of the information.
5. The *studied population (the sample)*. In this case it must be assessed if the studied population was properly defined, how the selection of the subjects was carried out, if there was or no randomization and the results external validity.
6. *Data analysis*. In this case the reader must evaluate if all the biases that may occur, were tacked into account by the author, especially the confusion bias.
7. *Results*must be consistent with the aim and type of the study. The reader will assess if there was performed the statistical analysis, if the proper measurements were calculated and if was also been interpreted the clinical or biological significance.
8. *Conclusions*must be drawn according to the aim of the study and must state if the results can be generalized to the studied population and if they can be applied inmedical practice.

III. The third step of the bibliographical research consists in combiningthe acquired data from different scientific articles, on the same subject and is accomplished through meta-analysis.Meta-analysis is a method of data analysis applied to summarizing research findings of individual studies. Meta-analysis has been used to give helpful insight regardingthe overall effectiveness of interventions,the relative impact of independent variables (e.g., the effect of different types of therapy), andthe strength of relationship between variables. Meta-analysis produces a stronger conclusion than can be provided by any individual study.

4.1.2 Formulating a research problem

After a rigorous review of the literature the current stage of knowledge regarding the research theme is revealed. Sometimes beginner researchers after selecting a research topic skip this stage and pass over the data collection stage. This will lead to the risk of not obtaining information on the problem of interest.

Kerlinger (1973) stated that a good research problem must fulfill three criteria:

- The variables in the problem should express a relationship
- The research problem should be stated in question form
- The statement of the problem should be made in such way to imply possibilities of empirical testing.

Is very important to formulate a specific problem, this will help to ensure that the investigators understand the problem and will help them in tacking decisions regarding: participants, instruments, devices and measures.

After the research problem has been stated, a scientific hypothesis must be formulated.

4.1.3 Scientific or research hypothesis

The scientific hypothesis is an affirmation regarding a presumed relation between the studied factors and the outcome and so the scientific hypothesis can be:

- Null hypothesis (H0) states that there is no relationship among the variables being investigated.

Example: 1. Vitamin C does not inhibit calcium absorption.
2. Laparoscopic cholecystectomy is as effective as the classical one.

- Alternative hypothesis (H1) states that there is relationship among the variables being investigated.

Example: 1. Vitamin C does inhibitcalcium absorption.
2. Laparoscopic cholecystectomy is more effective thanthe classical one.

The hypothesis will have to be confirmed or rejected accordingly with the research findings, the statistical test will calculate the probability that the observed association is real or it occurs due to chance.

4.1.4 Study motivation

A research study must be conducted to bring new information in the medical research, so its motivation is related to the lack of previous work on a certain problem.

4.1.5 Formulating the research topic

In this stage the aim, the objectives and the type of the study must be mentioned.

Epidemiological studies may be regarded as having four aims, as listed below:

- Describing disease occurrence and distribution of disease in population groups and development over time (trend).
- Identifying cause of disease (aetiology).
- Undertaking experiments to assess the effects of treatments, therapeutic approaches, or preventive efforts.
- Evaluating diagnostic procedures.

The study objectives are divided in:

- *General objectives*: a generic statement which describes in broad terms what the study wishes to accomplish
- *Specific objectives*: dividing the overall objective into smaller parts to systematically address the various aspects of the problem. In other words, these objectives must specify ***what*** the researcher will do in each phase of the study, ***how***, ***where***, ***when*** and***why***.

When expressing the aim and the objectives of the study, phrasing should beclear, precise, unambiguoususing action verbs.

In the dictionary the words “aim” and “objective” are synonymous, but in research terms the word “aim” is used to state what the researcher intends to achieve by undertaking his research, and the word “objective” is used to state what is intended to be achieved by the end of the research.

Example:

Aim: To demonstrate the role of chromium in the etiology of chronic obstructive pulmonary disease.

General objective: To calculate the role of chromium in the etiology of pulmonary disease.

Specific objectives:

- Define the study population;
- Select the sample;
- Ensuring the compatibility between the study groups.

The choice of study type depends on the problem to be studied and is closely related to the objective. If the objective of the study is to describe the frequency or the characteristics of a disease in a population according to time, place and individual, than

the conducted study will be a descriptive one. Depending on the objective, the type of the study is chosen.

4.2 Protocol development

The research protocol is a document that states the objective and the scientific hypothesis of the study, and describes how the future study will be carried out and what achievements it will bring.

Protocol development is important and necessary for the ethics committees and eventual sponsors of the research study.

Protocol includes several forms, which must state the following information regarding the research project:

- Title
- Summary
- Description
- The ethical standards of research
- Ways of communicate and disseminate the results
- Perspectives for further research
- Annexes

The research project must present information regarding:

- Human resources (team members, their institution affiliation);
- Theme relevance for knowledge development in the corresponding field
- The current state of knowledge and references
- Study justification
- Study objectives and research hypothesis
- Study methodology:
    - The studied factor/factors (the independent variable) and possible methods for their measurement;
    - The outcome (the dependent variable);
    - Type and design of the study;
    - The study sample/samples and the sampling methods;
    - The planning of activities;
    - The budget must be structured tacking into account the necessary expenses for:
        - Capital (equipment, software);
        - Staff (training, health insurance);
        - Bibliographic research (documentation, subscriptions);
        - Travel (visits to other research units, conferences attending);
        - Materials (kits, reagents, stationery);

        - Results dissemination (scientific papers publishing);
        - Indirect costs (electricity bills, heating bills, water bills);
    - Ways to harness the results;
    - Perspectives for other projects.

The protocol must also contains annexes, curriculum vitae of the principal investigator and co-investigators, calendar of activities, the ethics committee approval, special documents and the informed consent signed by the subjects included in the study. The protocol must be phrased clearly, properly prepared, and with favorable repercussions for the future.

Before implementation, the protocol will be evaluated by experts for eligibility (if the conditions for acceptance and financing are fulfilled), quality and relevance.

Chapter 5 – Study population and sampling

Data collection is performed according to objectives, study type, time, human and financial resources available.

There are two main possibilitiesof data collection:

- Regarding studied elements:
  - Exhaustive collection. All population subjects that we desire to study. Hard to accomplish due to high costs or study population alteration.
  - By sampling. The method used in medical studies.
- Regarding the length of collection:
  - Transverse. A group is studied in a precise moment in time.
  - Longitudinal (extended in time):
    - Retrospective – based on medical registers
    - Prospective – data is collected on pre-established time intervals.

5.1 The study population

For epidemiological purposes, the study population is any community of people, animals, plants or objectsall with measurable or quantifiable characteristics. The study population is the entire group we define, we wish to describe and draw conclusions about it.

The population may be seen as an immensity as well as a dynamic, and a study is always based on a sample of it and the sample must be representative for the population interested in study. For each study populationit can be selectedseveraltypes of sample. Based on these samples an estimator is calculated and based on the calculated estimator, which can be represented by the arithmetic mean, percentage, it can be drawn conclusions about the entire population. It is very important that the investigator defines very precise and carefully the study population (the inclusion and exclusion criteria) before selecting the sample units.

5.2 The concept of sampling

Sampling is the process of selecting a group of elements(sample) from a larger group (study population). Based on the sample, the prevalence of the outcome, any event or situations and any information regarding the study population, can be estimated or predicted. The sample represents a subgroup of the study population (Figure 1).

Example: It is desired to estimate the average age of the students in order to investigate a possible relation between age and the moment one starts smoking.There are two methods to estimate the average age.A method would be that all students be contacted and asked their age. To calculate the average age, all students' ages are added and then divided by the number of students.Another method is that from all students to be selected only a group of students. Only these students will be contacted and their age registered. The average age will be calculated by adding the ages and dividing it by the number of students selected. This calculated average age will estimate the average age of all students.

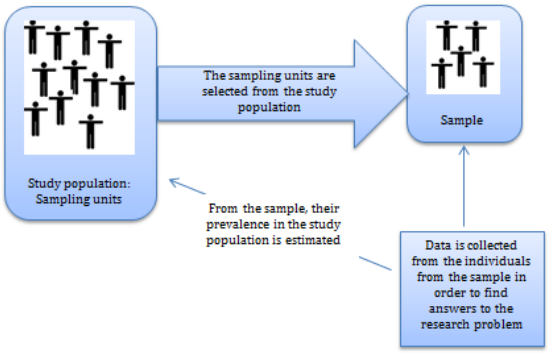


Figure 1: The concept of sampling

The advantages of the process of selecting a sample are that it saves time, human and financial resources, also samples can be studied more rapidly and more details can be found.

The disadvantage is that there are not ***found*** the information about the population characteristics but only ***estimation*** or ***prediction*** of them.

The goal of sampling is to obtain a representative sample, similar to the population on all characteristics. A perfect representative sample is a "mirror image" of the population from which it was selected.A sample is representative if it fits from both quantitative and qualitative point of view:

- Quantitative: in the sample must be included a sufficient large number of elements.
- Qualitative: each element from the study population must be randomly selected.

Sampling terminology

In the process of sampling there are few aspects that must be considered:

- **Study population**, usually denoted by the letter "N", represents the larger group from which the sample is selected.

- **Sample** is represented by the smaller group or the subgroup selected from the population that is interested in study.
- **Sample size**, usually denoted by the letter "n", is the number of elements included in the sample.
- **Sampling design (sampling strategy)**is represented by the sampling selection methods (types of sampling).
- **Sampling unit (sampling element)**represents each individual that becomes the basis for selecting the sample.
- **Sampling frame** is the list in which each individual from the study population is identified.
- **Sample statistics** represents the finding obtained following the information gathered from the individuals included in the sample and is a numerical characteristic of the sample data. Estimators are denoted by Latin letters.
- **Population mean (population parameter)** is the estimates drawn from sample statistics and is a numerical characteristic of a population.Each sample drawn from the study population will show an average value which can only estimate the parameter. Usually the parameters are denoted by Greek letters.
- **Sampling error** is used to refer to the difference between the value of a sample statistic and the value of the population parameter.
- **Census** is when data are collected from everyone in the population.
- **Response rate** is represented by the percentage of individuals in the selected sample who actually participate. This rate should be as high as possible.

Factors affecting the inferences drawn from a sample

There are two factors which could influence the degree of certainty regarding the inferences drawn from a sample:

- **The sample size** – the larger the sample size is, the more accurate the findings. Findings on a larger sample have more certainty than the ones on a smaller sample.
- **The extent of variation in the study population**–the higher the variation regarding the characteristics under study in the study population, the greater the uncertainty for a given sample size. If the study population is homogeneous (similar) considering the characteristics under study, is sufficient to select a small sample to be able to provide a good estimate, but if the study population is heterogeneous (diversified) a larger sample is needed to be selected to be able to obtain the same level of accuracy.

Biases in selecting a sample

Bias can occur if:

- In the sampling process is used a non-random method. In this situation subject selection can be influenced consciously or unconsciously by researcher choice;
- The sampling frame is not accurate and complete (does not cover all individuals from the study population);
- A fraction of the study population is impossible to find or refuses to cooperate.

Types of sampling

The sampling strategies can be categorized as follows (Figure 2):

- Random or probability sampling designs
- Non-random or non-probability sampling designs

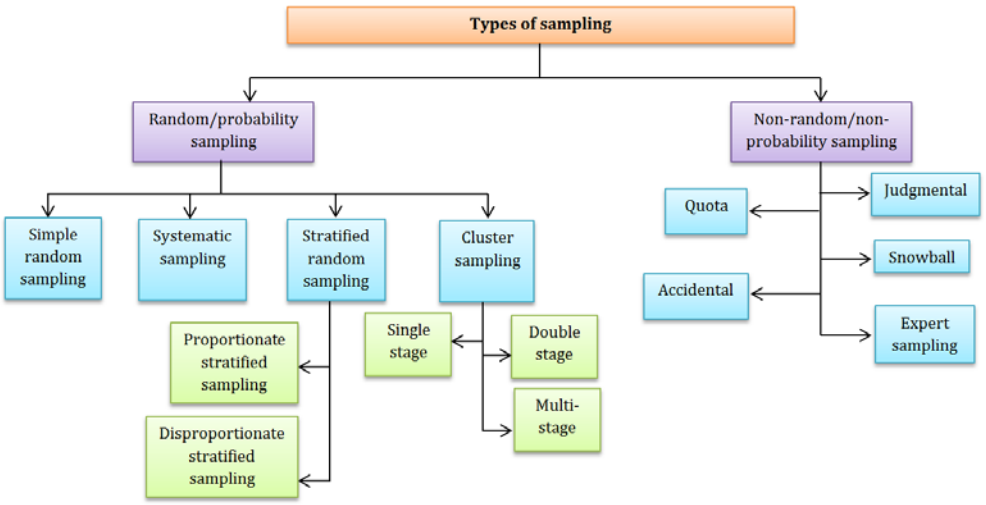


Figure 2: Types of sampling

5.2.1 Random or probability sampling designs

Regarding the methods of drawing a random sample, there are three most used methods:

- The "fishbowl" draws or "the hat" model can be applied if the study population is small. The procedure requires writing everyone's name or numbering on an equal-sized piece of paper, than put the pieces of paper into a box and picking them one by one without looking, until the

number of pieces of paper pulled from the box equals the desired sample size.

- Computer program which can help in selecting a random sample.
- A table of randomly generated numbers available in most statistics or research methodology books.The procedure for using a table of random numbers requires following a few steps:
    1. First there must be identified the total number of elements from the study population (79, 135, 793 or 1279) and those must be numbered. The total number of elements in a study population may run up to four or more digits:
        - If the total number of elements in a population is 9 or less than there will be numbers with 1 digit;
        - If the total number of elements in a population is 99 or less than there will be numbers with 2 digits.
    2. Each element must be numbered starting from 1;
    3. After choosing the table with random numbers, if it has more than one page, the starting page will be chosen randomly, also the starting column or row must be chosen randomly and the direction for selecting the numbers will be predetermined;
    4. Corresponding to the number of digits to which the study population runs, will be selected the same number, randomly, the digits of the column or row from the table. For example if the study population consists of 456 individuals (3 digits), than 3 columns or rows must be chosen randomly;
    5. Sample size must be decided;
    6. The last step is selecting the required number of elements for the sample from the table. The numbers will be selected from the columns or rows already chosen. Taking the example from above, 3 columns were randomly selected (7, 5, 2). If the sample size is 55, than 55 numbers will be selected from these columns. The selection will start from the first chosen column. Due to the fact that the study population consists of 456 individuals (3 digits), the last 3 digits of the numbers from the column will be underlined (Figure 3). The first number is 700, is more than 456 (total study population), so the 700^{th} element cannot be accepted, the second number is 423 which is below the total elements in the population, so the 423^{th} element becomes a part of the sample. This process will go on until all 55 elements of the sample will be selected.

	1	2	3	4	5	6	7	8	9	10
1	24697	87654	53367	64294	16436	81682	32700	39613	24127	69368
2	10917	19667	64436	28133	95333	35941	80423	19908	43043	42521
3	46768	11186	30081	63026	54008	11433	10983	47335	13544	66830
4	88611	47447	72546	14502	18353	78972	45857	79801	61375	45715
5	84444	66482	91751	78369	21489	58290	32364	22607	54812	59017
6	80672	19090	37005	39590	38536	58064	44617	24548	19584	58403
7	97765	80082	53100	66714	44238	78566	55492	64464	52660	24065
8	23621	98363	66452	30792	20937	24367	72160	16059	10958	94134
9	93950	83314	64793	98547	48587	55625	19982	65910	91782	83309
10	43494	58259	90223	84217	21321	30269	78896	14028	81523	18271
11	61515	14627	17852	23116	35385	57176	68909	80928	70078	37458
12	67314	79060	43832	47021	93334	44246	96148	21835	14487	69057
13	27915	81217	48813	12272	36312	39350	87865	49299	10038	76574
14	90968	72701	79635	74680	19939	14113	59206	90121	44616	59561
15	16178	23770	50119	60365	16934	49420	12947	39820	48216	49719
16	47984	57020	32432	98650	47711	19604	90484	40029	95604	40221
17	56264	73898	56717	61215	73932	41975	12748	48796	79478	34544
18	98332	36277	56620	41418	42288	38329	90533	85773	52675	73639
19	85982	55340	80088	20328	49495	24122	17768	77097	99203	59908
20	45289	45416	64290	76219	82002	63205	50917	14932	63838	35623
21	63179	32772	10218	64449	31965	25803	57113	94538	27847	45925
22	86822	64570	10625	99040	74198	97643	40099	43232	40412	26878
23	31834	75485	62296	56876	80246	82197	10032	56373	97404	53407
24	30741	25303	19744	20456	15594	46903	15268	26561	22854	16827
25	60102	26660	19436	69888	45020	66192	89879	77630	63883	61559
26	91350	38331	41914	76904	89313	90993	56201	36686	49850	81036
27	63528	94401	53224	64821	67249	66844	74728	21144	38892	96303
28	96836	91969	66737	45576	15328	58225	53605	10501	94754	30259
29	92060	23715	47844	66360	38705	50222	77395	88416	50869	42485
30	97875	44557	15541	17920	17750	70540	70812	86319	69918	61834
31	14697	13698	99520	40670	67542	47619	53029	52163	73524	85348
32	74752	95574	45295	18114	87641	41921	90311	45907	51632	37179
33	81586	76389	17125	42327	39893	60259	15808	26389	86488	27495
34	35055	92946	15804	65410	17642	50001	65323	12330	44167	36065
35	34357	87799	12546	17901	28309	64753	79891	20582	88366	99475
36	84021	93976	30991	39472	78736	22934	51432	49261	22905	43987
37	98167	64384	70231	99645	59256	95967	56368	67326	17663	41653
38	15229	32790	46175	81446	21260	80828	87315	60734	80916	49935
39	65714	69139	32238	89230	63547	40621	87780	52401	93220	36565
40	58994	78446	34720	41508	35820	59696	18637	86539	77003	65204
41	13899	11958	91740	18809	99582	99489	69042	52162	63632	31901
42	26127	20646	56795	59688	61707	34428	96433	24045	59315	54983
43	95196	98799	81279	76620	51291	91602	89704	44687	20825	42969
44	46273	69471	90915	61306	89085	94013	55915	19091	12069	11785
45	12361	69198	80960	47338	60014	76833	73583	48991	82058	68201

Figure 3: Selecting a sample using a table of random numbers

Random sampling can be selected by using two different systems:

- Sampling without replacement – doesn't allow the selection of a person more than once.
- Sampling with replacement – implies that the selected person to be replaced in the study population increasing in this way the possibility to be drawn again.

Types of random sampling designs

I. *Simple random sampling*means that each individual from the study population has an **equal** and **independent** chance of selection to be included in the sample. Equal chance implies that each individual in the study population presents the same probability of being selected in the sample and its selection is not influenced by other factors such as personal preferences of the researcher. Independent chance implies that the selection in a sample of an individual is not conditioned by the selection of an other person. The sample will be considered as a random/probability sample only if both these conditions are met, otherwise bias can be introduced into the study. The procedure for selecting a simple random sample requires following three steps (Figure 4).

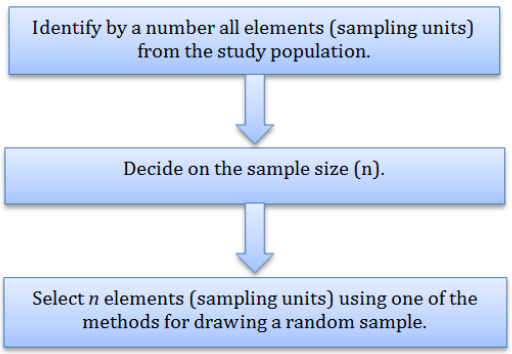


Figure 4: The procedure for selecting a simple random sample

II. *Systematic sampling* implies following a few steps (Figure 5). After the list with the elements in the study population is set and the sample size is decided, the sample interval must be determined. In order to determine this interval the **sample factor** is calculated by dividing the population size to the sample size. This factor is denoted by the letter "k". The **sample interval** is represented by the numbers between 1 and k. The next step is to randomly select a number between 1 and *k*(sample interval), and that person will be included in the sample. Then starting from the randomly selected number, each k^{th} element will be included in the sample.

For example, if the total number of elements in the study population is 200, and the desired sample size is 50, then *k* equals 4 which represent the sample interval.Next,

a number between 1 and 4 must be randomly selected. If this number is 3, then the first person to be included in the sample will be the third person from the total study population, the second person the 7th person (3+4), the third person will be person 11 (7+4), and so the process will continue until all 50 elements of the sample will be selected.

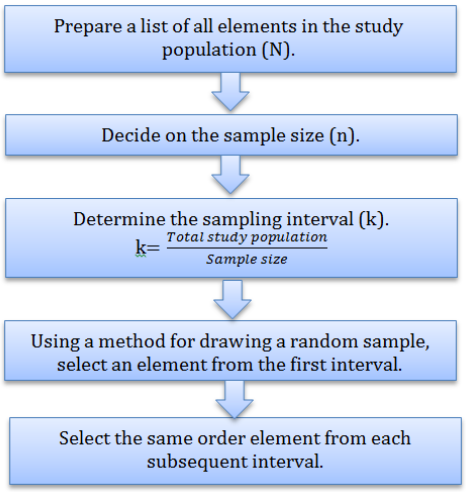


Figure 5: The procedure for selecting a systematic sample

III. *Stratified random sampling* implies dividing the study population into mutually exclusive groups called **strata**, and then a random sample is selected from each of the groups (strata). This type of sampling is based upon that if the heterogeneity in the study population can be reduced by some means, a greater accuracy of the estimation can be achieved. In stratified random sampling the study population is stratified in such a manner that the population within a stratum to be homogeneous regarding the characteristic on the basis of which the study population is being stratified. It is very important that the characteristics chosen to be the stratification factor are clearly identifiable in the study population and to be related to the main variable that is explored in the study. There are two types of stratified sampling:

- Proportionate stratified sampling – the numbers of people selected from the groups are proportional to their sizes in the study population.
- Disproportionate stratified sampling – the numbers of people selected from the groups are not proportional to their sizes in the study population.

For example, the stratification is made based on gender, and the study population consists in 70% male and 30% female and the desired sample size is 100. For proportional stratified sampling there will be randomly selected 70 males and 30

females. For disproportional stratified sampling is possible to randomly select 50 males and 50 females.

IV. *Cluster sampling* is used when the study population is large (city, state, country or regions). A **cluster** is a collective type of unit that includes multiple elements. There are three types of cluster sampling:

- Single stage cluster sampling – clusters are randomly selected and all the elements in the selected clusters constitute the sample;

Example: If 5 hospitals were randomly selected, then all patients in those 5 hospitals will be included in the sample.

- Double stage cluster sampling – clusters are randomly selected, and a random sample of elements is drawn from each of the selected clusters;

Example: If 5 hospitals were randomly selected, then from each hospital a random sample will be selected, 50 patients from every hospital floor.

- Multi-stage cluster sampling – when more than two stages are desired.

Example: If 5 hospitals were randomly selected and from each hospital are randomly selected 3 hospital floors, then from each hospital floor a random sample will be selected, 10 patients from every medical ward located on the hospital floors selected.

5.2.2 Non-random or non-probability sampling designs

The non-random sampling designs are used when the number of elements in the study population is either unknown or cannot individually identified, and so the selection of elements is dependent upon other considerations.

There are five most used non-random designs:

I. *Accidental or convenience sampling*–There are chosen the individuals most available and willing to participate in the study. The biggest disadvantage of this method is that is possible that the chosen individuals don't present the characteristic interested in study.

II. *Quota sampling*–A researcher will set **quotas** which represent some characteristics of the study population (gender, race, religion). The sample size is decided and in the sample will be included the first individuals that present the chosen characteristic, from a convenient location.

Advantages:

- Cheap;
- No need of a sampling frame, total number of elements in the study population;
- Guarantees the inclusion of the type of individuals that are needed in the study.

Disadvantages:

- The findings cannot be generalized to the study population;
- The most accessible individuals might present characteristics unique to them so in this case they will not be truly representative for the study population.

III. *Judgmental or purposive sampling* – It is based on the researcher judgment as to who will provide the best information in order to achieve the objectives of the study. This type of sampling is used when is desired a phenomenon description or the development of something about only few things are known.

IV. *Expert sampling* – It is similar with the judgmental or purposive sampling, except in this type of sampling are chosen only the individuals that are experts in the future research field.

V. *Snowball sampling* – Each research participant is asked to identify other potential research participants who present a certain inclusion characteristic. This process of selecting a sample uses networks. It starts with a few selected individuals, from whom the information is collected, then they are asked to identify another people and those people selected by them will be included in the sample. This process continues until the required number of individuals is reached (Figure 6).

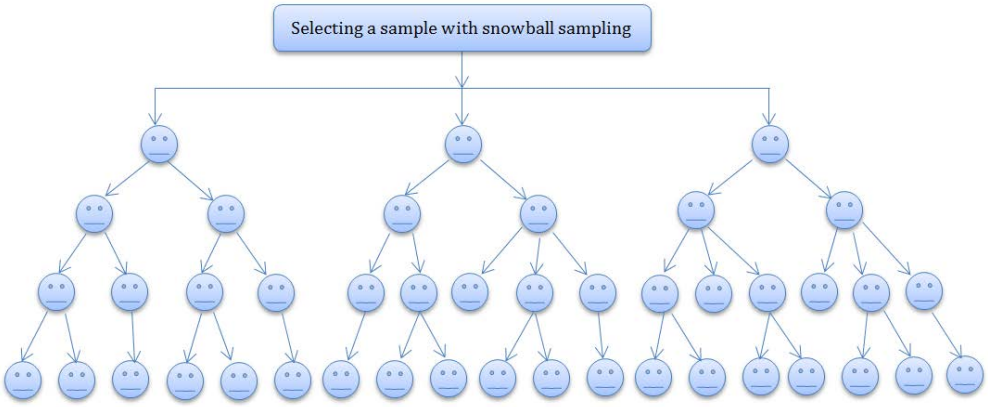


Figure 6: The process of snowball sampling

Disadvantages:

- The choice of the majority of the sample rests upon the choice of the first selected individuals;
- Difficult to use when the sample becomes very large.

Chapter 6 – Methods of data collection

Data can be collected from primary sources or from secondary sources. Data collected from primary sources are called primary data and the ones' collected from secondary resources are called secondary data.

There are several methods for primary and secondary data collection (Figure 1).

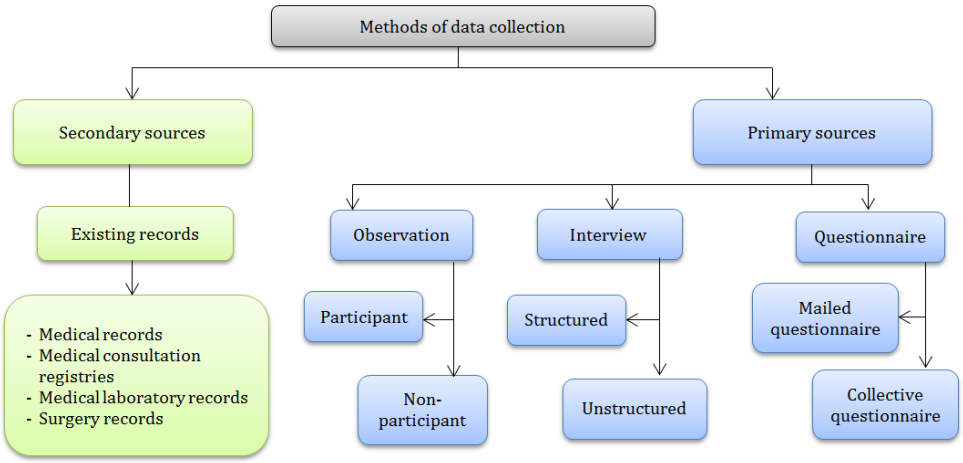


Figure 1: Methods of data collection

6.1 Methods for primary data collection

6.1.1 Observation

Observation is the best approach to collect information when the researcher is most interested in the behavior of the individuals included in the study. There are two types of observation:

- Participant observation – when the researcher participates in the activities of the group that is being observed, in the same manner as the members of the group.
- Non-participant observation – when the researcher does not participate in the activities of the observed group, but remains a passive observer.

The recording instrument for data collection must be adequatefor the protocol designed by the researcher. Data collected should be consistent, collected in the same manner. The records must be systematic and well standardized and control observations are necessary in order to discover any deviation due to external factors. Also ethical issuesconsiderations are mandatory.

Using observation as a method of data collection, several problems might arise:

- Hawthorne effect – when individuals that are being observed change their behavior because they are aware that they are observed;
- Observer bias – when an observer is not impartial;
- The interpretations drawn from the observation might be different from an observer to other;
- Incomplete observation due to the method of recording.

Methods of recording an observation:

- Narrative recording – the researcher records a description of the interaction in his/hers own words.
- Using scales – the development of a scale in order to rate various aspects of the interaction. Depending on the purpose of the observation and the type of data that is needed to be collected, the researcher can develop a one-, two- or three-directional scale.
- Categorical recording – the observations are recorded using categories (Good/Bad; Always/Sometimes/Never).
- Recording on electronic devices

6.1.2 The interview

*The interview*is a method of data collection very often used. The interviews are classified in two categories regarding the flexibility of the interview:

- Unstructured interviews – the researcher is free to formulate his own questions, ask them in any sequence he/she wishes and raise issues on the spur of the moment.
- Structured interview –the questions asked are predetermined and the researcher must use the same wording and sequence of the questions as they are specified in the interview schedule.

Regarding the number of respondents, the interview can be categorized as:

- In-depth interview (individual interview) – "face-to-face" encounter between the researcher and only one individual.
- Group interview – the researcher encounters with a group of individuals.

Advantages of the interview:

- The most appropriate approach when a complex and sensitive area is studied;
- Information can be supplemented by observation of nonverbal reactions;

- The questions can be explained;
- Has a wider application (can be used with almost any type of individuals: children, the handicapped).

Disadvantages of the interview:

- Time consuming and expensive;
- Quality of data depends upon the quality of the interaction;
- Quality of data depends upon the interviewer;
- The researcher may introduce biases coming from the interpretation of responses.

6.1.3 Questionnaire

The *questionnaire*is a written list of questions. The structure of the questionnaire must include:

- Introduction: must specify a brief presentation of the research topic, why is important to perform the research study, how will conduct the study and how the results will be used.Also in introduction must be described the following:
  - The purpose and the objectives of the study;
  - The importance of correct answers;
  - The answers confidentiality;
  - Anonymity.
- Questions:
  - Useful: provide the needed information;
  - Relevant: adequate to the research theme;
  - Adapted to the level of knowledge of the respondents;
  - Neutral forms: must not lead the respondent to answer in a certain direction;
  - Clear and easy to understand, using simple words;
  - In a logical order from simple to complex.
- Questionnaire graphics:
  - Interactive style;
  - Spaces between questions;
  - The sequence of questions should be easy to follow;
  - The layout should be such that the questionnaire is easy to read and pleasant to the eye.

Types of questions:

- Open-ended questions.The possible responses are not given, giving the respondent the possibility to express freely and to feel more comfortable about expressing their opinions (Figure 2).

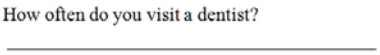


Figure 2: Open-ended question

- Semi-open questions. Some possible answers are given and offer to the respondent the possibility to complete the answer if it is not listed in the presented categories (Figure 3).

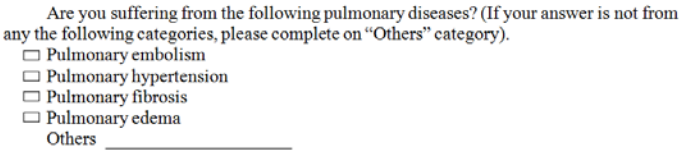


Figure 3: Semi-open questions

- Closed questions.The possible answers are set out in the questionnaire and so is ensuredthat the needed information is obtained (Figure 4). This type of question has a big disadvantage, through the given responses pattern, the respondent thinking might be conditioned and so the truly respondent opinion might not be provided.

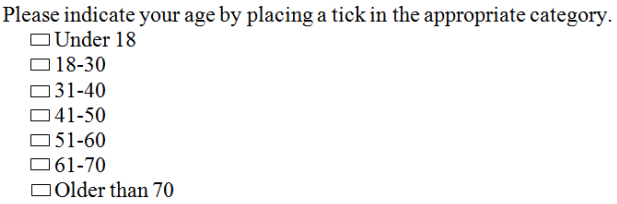


Figure 4: Closed questions

When formulating effective questions must be taken into account the following considerations:

- The phrasing must be kept simple and use everyday language;
- Ambiguity must be avoided;
- No double-barreled questions. A double-barreled question is a question that touches upon more than one issue, yet allows only for one answer;
- Leading questions must not be used. This questions by their contents, structure or wording will lead the respondent to a certain answer;
- Questions based on presumptions must also be avoided. This type of questions can be used only after the researcher is certain that the respondent fits into the enquiry category.

There are two ways of administering a questionnaire:

- The mailed questionnaire is a very effective way to administer a questionnaire, but the response rate is low.
- Collective questionnaire implies the administration of the questionnaireto a captive audience (students in a lecture hall, people assembled in one place). In this case the response rate is very high.

Advantages:

- Inexpensive;
- Offers great anonymity
- Useful when sensitive questions are asked

Disadvantages:

- Application is limited (cannot be used on children, very old or handicapped persons);
- Response rate is low;
- It is not possible to clarify issues;
- The response to a question may be influenced by the answer of another question;

6.2 Methods for secondary data collection

The *existing records*are most commonly used in medical research. Due to the fact the desired information already exists, the researcher only has to structure the data according to the purpose of the study. These sources are represented by:

- Medical records
- Medical consultation registries
- Medical laboratory records
- Surgery records

Chapter 7 – Research studies

Epidemiological studies can be classified as:

I. Descriptive
II. Analytical

*Descriptive studies*describe a health issue based on existing data and depending on time, place and person. Regarding **time**, data collection is realized per year/season or month and offers the possibility to recognize the disease variation in time. Accordingly to **place**, data collection is based on geographic distribution of the disease, risk factors or mortality. Regarding **person**, the individual characteristic used as descriptive criteria are age, sex, race orethnicity. Descriptive studies offer data concerning disease distribution pattern and disease natural history. It evaluates and monitors health status on population level. It can formulate a hypothesis, but it does not verify it. Usually a descriptive study is conducted when a researcher wants to investigate **who** is more predisposed to become ill the frequency of the disease, **what**is the frequency of the disease and **when** is the moment in time when the frequency of the disease is at maximum.

Descriptive studies can be categorized as:

- Individual
  - Case report
  - Case series
- Collective
  - Correlational study
  - Transversal (Cross-sectional) study

7.1 Case report

In a case report is described one particular case that cannot be explained by known diseases or cases that show an important and unprecedentedvariation of a disease or condition. Also a case report often describesunexpected eventsor a particular aspect of a clinical manifestation of a disease.

Case reports represent the first line of evidence, even though they are considered to present the lowest level of evidence. This type of study is used to report a unique and new case of a disease, any clinical problems (results of a screening test), unusual events (adverse drug reactions) or disease natural history.In a case report the patient should be described in detail, allowing others to identify patients with similar characteristics.

If multiple case reports describe something similar, the next step might be ananalytical study to determine if there is a relationship between the relevant

variables.A case report cannot demonstrate causalityor argue for the adoption of a new treatment approach.

7.2 Case series

A case series is a medical research descriptive study that tracks a group of patients who present a similar diagnosis or who are undergoing the same treatment procedure over a certain period of time. It can be retrospective or prospective and usually include a smaller number of patients. Based on the results of case series hypotheses can be generated that are useful in designing further analytical studies, but no causal inferences should be drawn from this type of study regarding the efficacy of the investigated treatment. In case series the disease frequency cannot be calculated and results cannot be generalized.

7.3 Correlational (ecological) studies

Correlational studies compare the frequency of a risk factor in a population with the disease prevalence and incidence, starting with known data. Usually this type of study is used to suggest disease causation, to evaluate disease control measures or to describe different social and cultural attributes affecting health. The unit to analyze is population groups.It compares populations from different countries, geographical locations in the same time frame orit compares the same population on different time periods.

Advantages

- Inexpensive
- Easy to carry out
- It relies on existing data
- Able to generate hypotheses

Disadvantages

- Poor validity (internal and external);
- Unable to control the confusion “bias”;
- It can lead to systematic errors;
- Information regarding the relationship between the levels of the risk factor levels and disease in individuals is not provided;
- Association observed between variables on group level does not necessarily represent the association at an individual level.

7.4 Cross-Sectional Studies

Cross-sectional studies are descriptive studies that can assess the prevalence, the amplitude, the regional distribution and the repercussions of a disease allowing the establishment of the health state diagnosis. In this type of study both exposure and disease outcome are assessed simultaneously for each individual.Also it performs a cross-section at a specific point in time, concerning the morbidity and the mortality of a disease. Data are collected from the individuals included in the study only once and during a single and usually a short period of time.

After is established the disease and the risk factor, the next step in a cross-sectional study is defining the study population. In this population, the presence or absence of disease and the presence or absence exposure are determined (Figure 1).

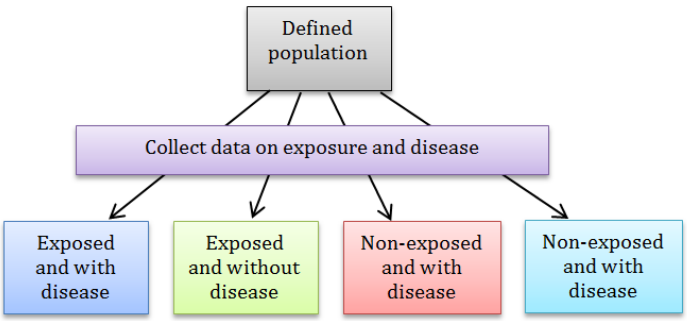


Figure 1: Cross-sectional study design

Based on collected data it can be calculated the prevalence of disease in persons exposed to the risk factor, prevalence of disease in persons non-exposed to the risk factor and prevalence ratio (PR) in exposed and unexposed.If P_0 is prevalence of disease in persons non-exposed and P_1 is prevalence of disease in persons exposed, then prevalence ratio is calculated dividingby P_1 with P_0. This prevalence ratio must not be confused with relative risk which is a measurement of the risk and involves the incidence of disease.

Cross-sectional studiescan also analyze some characteristics on population level (weight, height) and allow the start of some health programs orthe evaluation of the efficiency of some health programs.

Advantages

- Easy to carry out;
- Low cost;
- Requires a short time to complete;
- Allow assessment of the association between a disease and a risk factor, due to possibility to compare the prevalence of disease in the two groups (exposed and non-exposed);

- It generates causal hypothesis, but those are not verified;
- The first step in the research of epidemics;
- It allows the study of several diseases or risk factors at the same time;
- Are useful for investigating exposures represented byindividuals characteristics (blood type, personal characteristics, race).

Disadvantages

- Unable to asses a temporal relationship (if the exposure precedes the disease);
- Biases can occur;
- Unable to estimate the incidence of the disease;
- Does not allow risks calculation;
- Won't get results in rare diseases due to the necessity of a large number of persons to be included in the study.

7.5 Cohort studies

The cohort study is **an analytical and observational study**. This type of studies is **longitudinal** due to a long period of time for subjects follow up. A cohort study is conducted to investigate a possible **causal relation**between a risk factor or multiple risk factors and a disease.

The most well-known cohort study is the Framingham study of cardiovascular disease. The study began in 1948 in Massachusetts a town near Boston and the proposed follow-up time was 20 years. In the studywere included persons with age between 30 and 62 years. The reasoning for including this specific age group was that people younger than 30 years were most unlikely to present the studied cardiovascular disease and most people older than 62 years would already present the disease. The researchers recruited 5,209 men and women with no cardiovascular diseases at the time of the entry in the study. Among the exposures investigated in the Framingham study were smocking, obesity, elevated cholesterol levels, elevated blood pressure and low levels of physical activity.Data was collected by extensive physical examinations and lifestyle interviews the subjects continued to return to the study every two years for a detailed medical history, physical examination, and laboratory tests.

In 1971, in the studywere included 5,124 persons representing a second generation of participants. These subjects were represented bythe adult childrenof the original participants' and their spousesand similar examinations were performed. In 1994, due to the fact that the community was more diverse than it was in1948, the first Omni cohort (a more diverse sampling) was enrolled. In 2002 in the study was included a third generation of participants, represented by the grandchildren of the persons from the Original Cohort and in 2003 a second Omni cohort was enrolled.

Due to Framingham study fundamental contributions to the epidemiology of cardiovascular disease were accomplished.

7.5.1 Design and structure

The main objective is to evaluate the possible effects of different external or internal factors on the risk of developing a disease.

A cohort study is one that pursues a defined group (cohort), representing the sample of the study, over a given period of time.The study begins with **healthy or unaffected** subjectswho are divided in two groups: exposed and non-exposed to the risk factor. The subjects are followed up in order to determine the **incidence of disease** in both groups (Figure 2).If there is an association between the exposure and the disease, the proportion of subjects in whom the disease developswill be greater in the exposed group (incidence in the exposed group) than in the non-exposed group (incidence in the non-exposed group). In other words the cohort study compares the outcome in an exposed group and in a non-exposed group.

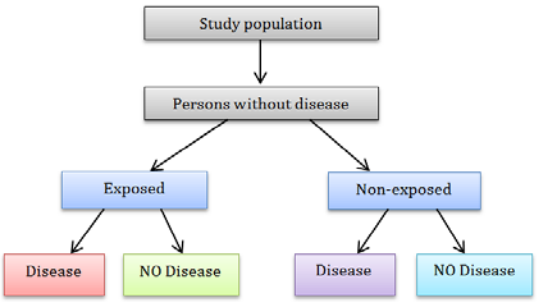


Figure 2: Cohort study design

Cohort studies can be prospective or retrospective.

In the prospective cohort study data is collected in the future, on pre-established time intervals. Exposure status is established as it occurs during the study. The groups are followed up for several years into the future and outcome is assessed.

Hypothetical example: A study starts in 2014. In the first step the population is selected. The subjects are followed up in time and in 2024, after the exposure and non-exposure are ascertained, the subjects are divided in two groups (exposed and non-exposed). After another follow-up period, in 2034, the disease status is determined (whom in the exposed group developed or not the disease and who in the non-exposed group developed or not the disease) (Figure 3).

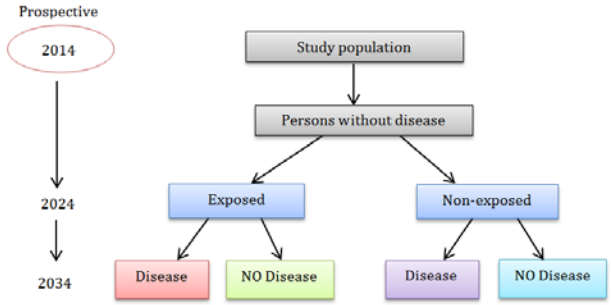


Figure 3: Prospective cohort study

In the retrospective cohort studies also called historical cohort study data is collected from the past, the study population and exposure statusare established from past records and outcome (disease occurrence or not) is ascertained at the time the study is begun.

Hypothetical example: A study starts in 2014. The study population is selected based on existing records from the past (1994). The included subjects were followed up through the years, also based on existing records. In 2004 the exposure status is ascertained and the subjects are divided in two groups (exposed and non-exposed). The disease status (development or no development of disease) is established in 2014 when the study starts (Figure 4).

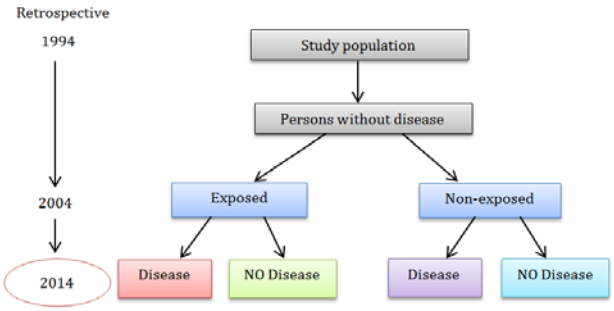


Figure 4: Retrospective cohort study

7.5.2 Selection of study population

The study population can be created either by selecting groups based on the exposure status (exposed and non-exposed), or by defining a population, based on characteristics not related to the exposure (neighborhood or community).This can happen before any members of the population become exposed or before their exposures are identified.

For each subject in the study there must be mentioned the follow-up period, the eligible criteria, diagnostic methods and measures that must be taken to avoid subjects'

losses during the study. Also for the subjects exposed, the methods used to assess the exposure must be described.

Matching must be ensured (see section 7.6.2). The groups are followed in identical way, present the sameinclusion criteria and the same diagnostic method is used to assess the disease.

Potential biases in cohort studies

A number of potential biases might occur when conducting cohort studies. These biasesshould be avoided or at least taken into account.

Bias in assessment of the outcome is one type of bias that can occur in a cohort study. This bias emerges when the investigator or the person who assess the disease status on each subject also has knowledge that the investigated subject was exposed to the risk factor and is aware of the tested hypothesis. In this case his judgment regarding the disease development might be biased.

Information bias occurs more often in retrospective cohort studies and represents differences in the amount and quality of data collected in the exposed and non-exposed group (see section 8.5.2).

Biases from nonresponse and losses to follow-up appear due to subjects' withdrawal or leaving the community and so it is impossible to be further followed-up.

Analytic bias occurs when the person responsible for data analysis has strong preconceptions and may insert his biases in results calculation.

7.5.3 Effect measurements

The effect measurements are based on estimates of disease **risk**. Usually the concept of risk is used when is intended to describe an association between a risk factor and the probability for adisease to develop. The importance of a risk factor is established by determining the proportion of the exposed and the proportion of the non-exposed who actually become ill, during a given period of time. This will allow the estimation of the risk of disease associated with the exposure to the risk factor. The confidence interval (see Chapter) gives confidence of the effect estimates. Usually in research studies the 95% confidence interval (95%CI) is used.

Effect measurements consist in the calculation of:

I. **Relative risk** (RR) is the ratio of the incidence of the disease in people exposed and unexposed, namely the risk of disease in exposed and unexposed.

$$Relative\ risk = \frac{Incidence\ in\ exposed}{Incidence\ in\ non-exposed}$$

All data regarding disease occurrence and the exposure to a potential risk factor are displayed in a table, a contingency table (Table I).The incidence of disease in the two groups is calculated based on this table.

Table I: Contingency table 2x2.

	Disease		
Exposure	Affected	Healthy	Total
Exposed	a	b	a+b
Non-Exposed	c	d	c+d
Total	a+c	b+d	

In exposed group (*a+b*) there are*a* persons with disease so incidence of disease or the risk in exposed group is *a/(a+b)*.

In non-exposed group (*c+d*) there are*c* persons with disease so incidence of disease or the risk in non-exposed group is *c/(c+d)*.

The relative risk is the ratio between risk in exposed group and risk in non-exposed group, so we can state that:

$$RR = \frac{(\frac{a}{a+b})}{(\frac{c}{c+d})}$$

The relative risk indicates how much greater is the risk of developing the disease in exposed group compared to non-exposed group. In other words a subject exposed to the risk has an*x*(where x is the obtained value of relative risk) more chances to develop the disease compared to a subject non-exposed.

The interpretation of risk relative value:

- If relative risk is greater than 1 means the risk in exposed is greater than the risk in non-exposed, there is an association between the risk factor and the disease.
- If relative risk is equal to 1, it means the risk in exposed equals the risk in non-exposed, the studied risk factor is an indifferent factor for the disease, thus there is no association between the risk factor and the disease.
- If relative risk is less than 1 means the risk in exposed is less than the risk in non-exposed, the studied risk factor is a protective factor for the disease, thus there is no association between the risk factor and the disease.

For each calculated value of relative risk it is calculated also the confidence interval. If there is or not an association between the risk factor and the disease depends also on the obtained confidence interval. If the value of relative risk is greater than 1, but the confidence interval includes value 1 (the lower limit is less than 1) there is no association between the risk factor and the disease. That is because the obtained

value of relative risk is an estimation of the true value of the risk of disease in population and the true value can take any values included in confidence interval.

II. **Attributable risk** for the exposed group is used to calculate the incidence of disease that is attributable to the exposure in the exposed group.

As is seen in the contingency table (Table I), even if the subjects were not exposed to the risk factorthere are *c*subjects with disease, so they have some risk of disease. This risk is termed *background risk*. Each subject exposed or non-exposed, due to the fact that was selected from a defined population, present this background risk. So the risk of disease in the exposed group is the sum of the background risk and the risk attributed to exposure to the investigated risk factor.

Attributable risk (AR) can be calculated as follows:

$$\text{AR} = (\text{Incidence in exposed}) - (\text{Incidence in non-exposed})$$

$$AR = \left(\frac{a}{a+b}\right) - \left(\frac{c}{c+d}\right)$$

The proportion of the risk in exposed due to exposure (attributable risk proportion) can be calculated as follows:

$$\text{ARp} = \frac{(Incidence\ in\ exposed) - (Incidence\ in\ non-exposed)}{Incidence\ in\ exposed} \text{x}100$$

$$ARp = \frac{\left(\frac{a}{a+b}\right) - \left(\frac{c}{c+d}\right)}{\left(\frac{a}{a+b}\right)} x100$$

There is an important difference between relative risk and attributable risk.

Relative risk is a measure of the strength of the association between the risk factor and the disease, while attributable risk expresses the potential in reducing the risk of disease if exposure can be eliminated.

III. **Population attributable risk** (PAR)represents the risk of disease in total population that is attributed to a certain exposure and can be calculated as follows:

$$\text{PAR} = (Incidence\ in\ total\ population) - (Incidence\ in\ non-exposed\ group)$$

To be able to calculate the population attributable risk first it must be calculated the incidence of disease in total population. If incidence of disease in total population is not known, it can be calculated by using the values of incidence in exposed (I_E) and of incidence in non-exposed(I_{NE}), and the proportion of the exposed and non-exposed in total population, as follows:

$$\text{Incidence in total population} = (I_E)\text{x}(\%\ \text{Exposed in population}) + (I_{NE})\text{x}(\%\ \text{Non-exposed in population})$$

The proportion of the risk in population due to exposure (population attributable risk proportion) can be calculated as follows:

$$PARp = \frac{(Incidence\ in\ total\ population) - (Incidence\ in\ non-exposed)}{Incidence\ in\ total\ population} x100$$

Usually the incidence of disease is calculated per 1,000 people (see section 8.1.2).

Advantages

Due to the fact that the study starts with unaffected individuals exposed and non-exposed to the risk factor and follows them in time through the development of disease, the temporal relationship (a risk factor precedes the disease)is established.Even if the cohort study can investigate multiple risk factors, there are intuitively easy to understand due to a well-defined design. Is well accepted the fact that prospective cohort studies in comparison with other types of research studies,give a more valid and less biased result. The incidence of disease and the risk of disease can be calculated. A cohort study is the best choice in investigating risk factors that are stable over time or rare exposures and in evaluating diseases with a high incidence rate. Also it can follow-upthe late effects of the disease.

Disadvantages

A cohort study requires a large number of subjects included in the study and a long follow-up time period that is why it isrelatively costly. It cannot be applied for assessing diseases with a low incidence rate because it may be impossible to form a sufficiently large study group. Due to changes that might appear regarding the participants to the study (change of residential area) the exposure assessment can be inflexible.

7.6Case – Control Studies

A case – control study is **an analytical and observational study**. This type of study is **longitudinal** due to a long period of time for subjects follow up and usually is **retrospective** data being collected from the past. Like the cohort study, the case control study is used to examine the possible relation of an exposure to a certain disease.

7.6.1 Design

The design of the study implies the selection, from the study population, of a group of individuals with the disease in question (called cases) and of a group of people without that disease (called controls), in order to compare them (Figure 5).If there is an association between the exposure and the disease, the proportion of subjects exposed to the risk factor will be greater in the case group than in the control group. In other words the prevalence of exposure will be greater in individuals who

present the disease (cases) than in individuals who do not present the disease (controls).

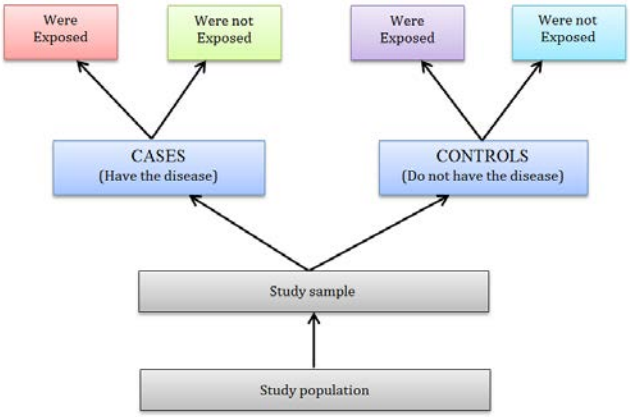


Figure 5: Case – control study design

Data regarding exposure is collected by reviewing past records (medical records or medical laboratory records) or by interview.

7.6.2 Selection of cases and controls

Selection of Cases

Cases can be selected from a variety of sources, including hospital patients, patients in physicians' private practices, or from private clinics. If the patients are selected only from one hospital is possible that the risk factor identified to be unique for that hospital. In this case the study results cannot be generalized to the study population.

In a case – control study the eligibility criteria for subjects selection most be very clear specified.

Selection of Controls

Controls may be selected from non-hospitalized persons living in the same community or from other hospitalized patients who do not present the studied disease.

Non-hospitalized persons as controls can be selected frominsurancelists of the health insurance agencies, company or school registers. Also it is possible to select for each case a control from the same neighborhood or a "best friend" control. Each individual included in case group is asked to name his/hers best friend and so the best friend becomes a control. Instead of a best friend, also a family member of the case can be selected.

Hospitalized persons as controls are often used due to the fact that these individuals are clearly identified and can be reached more easily being a "captive population", and also the costs are relativelysmaller when using such controls.

Matching

A major problem in a case – control study is that it might be differences between cases and controls regarding different characteristics or exposures. The controls might be exposed to other risk factors than the risk factor under study. This will lead to questioning the association between the risk factor and the disease. To avoid this problem the process of matching should be used in controls selection.

Matching is the process of selecting the controls in such way that they are similar to the cases regarding certain characteristics, like age, race, sex, socioeconomic status, occupation and religion.

Matching may be of two types:

Group matchingrefers to selecting controls so they are similar to the cases regarding the proportion of certain characteristic. For example if in the case group there are 43% males, in the controls will be selected that in the control group will be also 43% males.

Individual matching is when for each case is selected a control that is similar to the case in terms of several characteristics that are of concern. For example if a 25 years old white woman was included in the case group, than in the control group will be selected also a 25 years old white woman.

There two typesof problems regarding matching:

- *Practicalproblems* appear is when too many characteristics are used in matching process. In this case it might be difficult or impossible to find a control that is similar to a case tacking into consideration all the characteristics of concern.
- *Conceptual problems* occur when the matching process is made based on characteristic that is desired to be studied. Such a characteristic should not be used when matching a control with a case.

Potential biases in case – control studies

A major bias that might occur in a case – control study is the recall bias. This happens because in most case-control studies, data regarding the exposure status is collected through interviews.

Due to the fact that actually all humans present varying degrees of ability in remembering different information, limitations in recall occurs. **Limitation in recall** regarding the exposure is an important problem, and can lead to a misclassification of exposure status if it affects all subjects included in the study to the same extent.

Recall bias occurs when a possible relevant exposure is recalled by a person with the disease and forgotten by a person who does not present the disease (see

section 8.5.2). This type of bias in rather infrequent in case – control studies, but the possibility that it might occur must always be kept in mind.

7.6.3 Effect measurement

In a case – control study the effect measurement is based on the calculation of ***Odds Ratio***. The odd for an event to happen is the ratio between the number of ways the event can occur and the number of ways the event cannot occur. For example when choosing a race horse to bet on, the odd that the horse will win the race can be calculated. If the chosen race horse has 70% probability of winning (*P*), than it has also 30% probability of losing (1-*P*). The odd that the horse will win the race can be calculated as follows:

$$\text{Odd} = \frac{\textit{Probability that the horse will win}}{\textit{Probability that the horse will lose}} = \frac{70\%}{30\%} = 2{,}33{:}1$$

So the odd that the horse will win the race is 2,33.

In a case control study the odds ratio is defined as the ratio between the odd of the disease to develop in an exposed individual and the odd of the disease to develop in a non-exposed individual.

Like in the cohort study, all data regarding disease and the exposure status are displayed in a contingency table (Table II).

Table II: 2x2 contingency table

	Disease		
Exposure	Affected	Healthy	Total
Exposed	a	b	a+b
Non-Exposed	c	d	c+d
Total	a+c	b+d	

To calculate the value of odds ratio, based on data in the contingency table it is determinedthe proportion of the cases that were exposed (*a/(a+c)*) and the proportion that were not (*c/(a+c)*). The odd of disease in the exposed group (the probability of the disease to develop in an exposed person) is calculated as follows:

$$\text{Odd of disease in exposed} = \frac{\frac{a}{a+c}}{\frac{c}{a+c}} = \frac{a}{c}$$

Also it is determinedthe proportion of the controls that were exposed (*b/(b+d)*) and what proportion were not (*d/(b+d)*). The odd of disease in the non-exposed group (the probability of the disease to develop in a non-exposed person) is:

$$\text{Odd of disease in non-exposed} = \frac{\frac{b}{b+d}}{\frac{d}{b+d}} = \frac{b}{d}$$

The odds ratio (OR) is the ratio of the odd of disease in exposed to the odd of disease in non-exposed, and is calculated as follows:

$$\text{Odds ratio} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{ad}{bc}$$

For each calculated value of relative risk it is calculated also the confidence interval.

Advantages

One of the main advantages of case–control studies is that they are relatively simple to carry out and inexpensive. Also it requires a relatively small number of people included in the study. Unlike the cohort study, this type of research study is reasonably rapid, being retrospective, the exposure status is assessed from the past, so it does not require a long time to follow-up the subjects included in the study. It can be applied for assessing diseases with a low incidence rate and also multiple risk factors. The case control – studies can be organized as multicenter studies.

Disadvantages

The major disadvantage of the case–control study is that data regarding exposure status is collected after disease has developed so it is difficult to establish the temporal relationship between the exposure and the disease. Due to the fact that these studies are retrospective and data is collected from existing records or by interview, biases can occur. They cannot be applied in rare exposures and study only one effect (one disease). Also the incidence of disease cannot be determined.

The major differences between case-control studies and cohort studies are presented in the following table (Table III).

Table III: Differences between a case – control study and a cohort study

	Case-Control studies	**Cohort studies**
Population	Start with affected subjects	Start with healthy subjects
Incidence	No	Yes
Prevalence	No	No
Association	Odds ratio	Relative risk

7.7Randomized clinical trials

Usually the term "clinical trial" is applied to any type of study which presumes a planned medical experiment that involves humans. A clinical trial is conducted in order to determine which treatment is best for patients with a certain diagnosis. Also this type of study may be applied in order to study other types of treatment, like surgery procedures or radiation therapy.

Randomized clinical trial (RCT) is considered the ideal design for evaluating the effectiveness of new methods of treatment and to determine the possible side effects of the new forms of intervention.

When a new drug is developed, a long series of clinical trials are needed to assess the effect of the new treatment andthe clinical part usually occurs in four phases (Table IV).

Table IV: The four phases in the development of a new drug

Phase I	Toxicity and side-effects
Phase II	Dose–response
Phase III	Comparison with standard treatment or placebo
Phase IV	Post marketing

Phase I represents the study of toxicity and side-effects. In order to accomplish this, a clinical trial is conducted including only a small group of healthy and volunteer individuals.

In *phase II* another clinical trial is conducted on also a small group of patients to study the dose-response relationship (effect of different doses of the drug).

If there are promising results in the first two phases, in *phase III*, a clinical trial will be conducted to compare the new drug with the standard treatment or a placebo.

All these clinical trials are conducted before the new drug is released on the market.

Phase IV consists in long-term studies of safety and marketing studies.

Usually the term "clinical trial" refers to phase III trials.

Before any clinical trial starts a detailed protocol of the study must be developed. In the protocol must be stated the background and aim of the study, how the study will be conducted, subjects' inclusion and exclusion criteria, how many subjects will be included (estimated sample size), for how long they will be followed (patient follow-up) and the randomization methods used (randomization, blinding). Also must be described the treatment that is investigated and how data will be processed (statistical analysis). The expected outcomevariables and possible adverse

events must be mentioned. Most important before inclusion in the study each subject must sign an informant consent statement (ethics) (see section 10.4).

Randomized clinical trials are **analytical** and **experimental** studies. This type of is **prospective**.

7.7.1 Design

The design of randomized clinical trials allows the assessment of the effectiveness of a pharmaceutical product (drug or vaccine), evaluation of therapeutic procedures, comparing a new drug with the treatmentalready in use or a placebo and reporting and measurement of adverse reactions to administered products, and reporting different variations among patients.

The design of randomized clinical trials implies defining the study population represented by all patients that presents the disease in question. Next step is choosing the study sample. The patients included in the sample must present similar characteristics as patients in study population. Then the study sample is divided in two groups, experimental and control group. The best approach is to randomly assign the patients in the two groups. Further, the patients would be followed-up and different outcomes can be assessed (disease status, death, complications or relapses) (Figure 6)

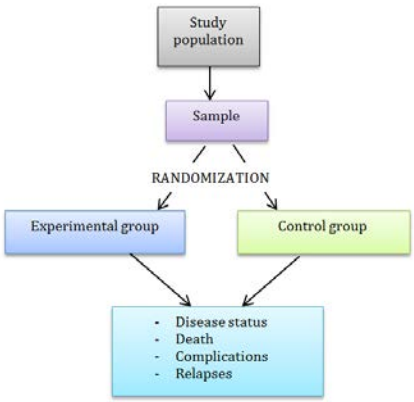


Figure 6: Randomized controlled trials design

7.7.2 Selection of controls

There are a few methods that can be used incontrols selection.

Historical controls are represented by the patients who previously received a standard treatment.In this case is quite difficult to ensure a correct comparison of treatments. When using historical controls data is collected from existing medical records, and the quality of data might not be same for the two groups.

Nonrandomized controls imply choosing simultaneous controlsthat are not selected by using a randomizationmethod. For example, the patients are assigned accordingly to the day of the month on which the patient was admitted. If a patient is admitted on an odd-numbered day he/she will be in experimental group and if the admission is on an even-numbered day, the patient will be in the control group.

The problem with this assignment system is that is predictable and might lead to occurrence of selection bias.

Randomization is the best approach in the design of a clinical trial. It involves patients assignment in such way that no systematic differences can occur between the two groups. The most important is that randomization ensures the unpredictability of the next assignment.

Methods of randomization:

One-to-onepresumes that in both groups are assignedequal numbers of patients. This methodcannot be applied in studies with a placebo group.It is desired that as few patients as possible receive placebo.

Random number tablesare usually the most used to assign patients in the two groups. There are several ways to use a random number table. One of the randomization methods that use a random number table is described in Chapter 5 (section 5.2.1). Another way to use such a table is the method with odd and even numbers.

Hypothetical example: A study is conducted in which are two groups, a group with therapy *X* and a group with therapy *Y*. First is decided which number (odd-number and even number) is assigned to the two groups. Let say an odd-number will be assigned to therapy *X*group and an even-number to therapy *Y* group. Next step is to close the eyes and place a finger on a random number table. The pointed row and column and also the direction of the movement on the table are written down. Let presume that was pointed the number "1" from column 5 and row 4. Then the first patient has an odd-number and will be assigned to therapy *X*group. If the direction of the movement is horizontally to right, next number is "8" an even-number, so the next patient will be assigned to therapy *Y* group and so on until all patients are assigned in one of the two groups (Figure 7).

	1	2	3	4	5	6	7	8	9	10
1	24697	87654	53367	64294	16436	81682	32700	39613	24127	69368
2	10917	19667	64436	28133	95333	35941	80423	19908	43043	42521
3	46768	11186	30081	63026	54008	11433	10983	47335	13544	66830
4	88611	47447	72546	14502	18353	78972	45857	79801	61375	45715
5	84444	66482	91751	78369	21489	58290	32364	22607	54812	59017
6	80672	19090	37005	39590	38536	58064	44617	24548	19584	58403
7	97765	80082	53100	66714	44238	78566	55492	64464	52660	24065
8	23621	98363	66452	30792	20937	24367	72160	16059	10958	94134
9	93950	83314	64793	98547	48587	55625	19982	65910	91782	83309
10	43494	58259	90223	84217	21321	30269	78896	14028	81523	18271

Figure 7: Method with odd and even numbers that use a random number table

Stratified randomizationrefers to similar patient groups regarding different prognostic factors, factors that might influence the studied therapy.These factors are represented by the age, sex orenvironment. This method of randomization can be resource demanding and is considered unnecessary in larger studies.

Blinding is used because the patients' assignment using randomization methods does not always ensure an unbiased comparison of treatments. Whenever the physician knows which patient has received which treatment systematic bias might occur.

There are three blinding methods:

- Simple blind – when the subjects do not know the details of the treatment given.
- Double blind – when neither the patient nor the treating physician knows the details of the treatment given.
- Triple blind – when including the evaluator does not know the details of the treatment given.

Blinding clinical trials is sometimes difficult due to the fact that a treatment might be so toxic that it is very important to know when it is given to a patient or dosing regimens differ considerably or a treatment has specific side-effects that it is very clearly to whom it is administered.

7.7.4 Surveillance of patients

For both groups, experimental and control group, the surveillance will be done in the same manner and for the same period of time and it will refer to:

- Improvement, stationary status or worsening of the disease
- Number of deaths
- Occurrence of complications
- Occurrence of relapses

7.7.3 Data analysis

The data analysis must provide answers to questions referring to the effectiveness of the intervention. In this case incidence rates for the "experimental" and "control" group will be calculated. To evaluate the strength of the effect induced by the intervention factor the value of relative risk is calculated as follows:

$RR=R_0/R_1$

Where: $R_0$is the risk of disease in the "experimental" group;

$R_1$is the risk of disease in the "control" group.

Interpretation of Relative Risk:

- If the value of relative risk is greater than 1, then the intervention is associated with an increased risk of disease occurrence;
- If the value of relative risk equals 1, then there is no evidence of the effect;
- If the value of relative risk is smaller than 1, then the intervention decreased the risk of disease.

Advantages

The randomized clinical trials are considered the "gold standard" of study designsdue to their ability to minimize the confounding and selection bias.Double-blind method decreases the possibility of biasesintroduced by the investigator. This type of study presents the ability to assess multiple outcomes.

Disadvantages

A randomized clinical trial implies a long time period of patients follow-up and is expensive to carry out. Also it requires the voluntary participation of subjects therefore these are not always representative of the population of interest for the study and in this case the results cannot be generalized.

7.8 Screening and diagnostic tests

In medical practice a good screening or a diagnostic test determines an accurate diagnosis, confirmed by the connection between the test valuesand disease. Also allows a better therapeutic decision.To be able to apply such a test in medical practice, first a screening or diagnostic test must be evaluated to determine the advantages and disadvantages of the proposed test.

A screening or a diagnostic test must ensure **reliability** before is used in medical practice. To assess the reliability of test,reproducibility (precision) and accuracy should be evaluated.

7.8.1 Ensuring reproducibility of test used

The **reproducibility** is the degree of a test to produce the same or a very similar result by repeated measurements.

Test reproducibility can be inter-observational and intra-observational. Inter-observational reproducibility (inter-observational variability) refers to an observation made by two different observers at the same time. Intra-observationalreproducibility (intra-observational variability) refers to an observation made by the same researcher at various intervals of time.

The reproducibility of the method can be quantified by the coefficient Cohen's kappa coefficient (*K*), is a statistical measure of the consistency between the two examinations.

$$K = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Where: Pr(a) is the relative observed agreement (the same result) among the two examinations.

Pr(e) is the hypothetical probability of chance agreement (the probability of two examinations to show the same result).

Cohen's kappa coefficient (*K*) interpretation

- If the two examinations show the same results then κ = 1.
- If the two examinations do not show the same results then κ = 0.

7.8.2. Evaluation of test accuracy

Accuracy (validity) is the ability of a test to identify between patients who present and who do not present the disease.

The accuracy of a test is evaluated by comparing it with a "golden standard test"(a reference test) in individuals with a known disease status.There are four types of results that can be obtained following the comparison of a two tests (true positive, true negative, false positive and false negative). The results obtained can be displayed in a contingency table. Value *a*represents the number of persons with disease and with positive test; value *b* represents the number of persons who do not present the disease, but tested positive; value *c*is the number of persons with disease, but tested negative; and value *d* is the number of unaffected persons tested negative (Table V).

Table V: Test results

	Disease		
Test	Present	Absent	Total
Positive	True positive *a*	False positive *b*	Individuals with positive test *a+b*
Negative	False negative *c*	True negative *d*	Individuals with negative test *c+d*
Total	Individuals with disease *a+c*	Individuals without disease *b+d*	

Based on the contingency table, accuracy is assessed by determining the **sensitivity** and **specificity** of the test.

Sensitivity (S_e) is the ability of the test to identify individuals who present the disease. It can be also defined as the probability of an affected person to be tested positive. Sensitivity is represented by the ratio between affected individuals with positive tests and all persons with disease.

$$S_e = \frac{Affected\ persons\ with\ positive\ test}{Total\ number\ of\ affected\ persons} = \frac{a}{a+c}$$

Specificity (S_p) is the ability of the test to identify individuals who do not present the disease. It indicates the proportion of subjects with a negative test of the total unaffected persons.

$$S_p = \frac{Unaffected\ persons\ with\ negative\ test}{Total\ number\ of\ unaffected\ persons} = \frac{d}{b+d}$$

An ideal test is one with sensitivity and specificity of 100% each. In medical practice this situation is not possible. The more test sensitivity increases, the more its specificity will decrease and vice versa.

Other two commonly used parameters to assess the accuracy of a diagnostic test are false rates (false positive and false negative rate).

False positive rate (FPR) represents the probability of a positive tested person to be unaffected. It is expressed as the ratio between the number of unaffected persons that were tested positive and all unaffected persons.

$$FPR = \frac{Unaffected\ persons\ with\ positive\ test}{All\ unaffected\ persons} = \frac{b}{b+d}$$

False negative rate (FNR) is the probability that an affected person is testednegative. It is expressed as the ratio between the number of affected persons with negative tests and the total number of affected persons.

$$FNR = \frac{Affected\ persons\ with\ negative\ test}{All\ affected\ persons} = \frac{c}{a+c}$$

When the test variables are quantitative (continuous variables) a threshold should be selected to allow subjects classification as affected or unaffected. If it is decided that the subjects with a specific value that is above threshold have a positive test and the subjects with a specific value that is below threshold have a negative test, by moving the threshold, its limits will influence the value of sensitivity and specificity:

- The moredown threshold is moved the greater will be the number of identified affected persons, but also the higher will be the false positive rate.

• If the threshold is moved upwards there will be a greater number of identified unaffected persons but also a higher false negative rate.

7.8.3 Evaluation of test performance in medical field

To evaluate a test performance in medical field it must be assessed the test's predictive power. To achieve this, the predictive values (positive and negative) must be calculated.

Positive predictive value (PPV) is the probability of a subject who tested positive to present the disease.It indicates the proportion of affected individuals from of all those with positive tests and it is expressed as the ratio between the number of affected persons with positive tests and all persons with positive tests.

$$PPV = \frac{Affected\ persons\ with\ positive\ test}{All\ persons\ with\ positive\ test} = \frac{a}{a+b}$$

Negative predictive value (NPV) represents the probability that an individual with a negative test does not present the disease. This value indicates the proportion of unaffected persons from all individuals with negative test. It is expressed as the ratio between the number of unaffected persons with negative test and all persons with negative test.

$$NPV = \frac{Unaffected\ persons\ with\ negative\ test}{All\ \ persons\ with\ negative\ test} = \frac{d}{c+d}$$

When a high sensitivity is associated with high positive predictive value and a low false negative rate, or a high specificity is associated with a high negative predictive value and a low false positive rate, a good validity of the diagnostic test is provided.

Chapter 8 – Data analysis and results interpretation

All medical data that are collected from the individuals included in research study are called variables. The variable is considered a function and it can take different values for each sample or study population element.

Variables are classified in two groups:

I. **Quantitative variables** (which can be measured):
- *Continuous variables* are measurable variables which can take an infinite number of values (cholesterol values, blood pressure values);
- *Discontinuous variables* are represented by thevariables which can takeonly integer values (APGAR score).

II. **Qualitative variables** (which can't be measured)
- *Nominal variables* represents groups of elements which can't be organized (hair color);
- *Ordered nominal variables*are represented by groups of elements organized inranked categories (for example the treatment efficiency can be categorized as: very good/good/bad);
- *Binary variables*are those elements that present only two possibilities(ill/healthy, YES/NO), usually labeled zero and one.

To acknowledge the health status of a population several indicators must be calculated. Proportions and rates are used to quantify morbidity and mortality. From these proportions the presence of the risk in different groups can be also quantified.

Epidemiological measurements can be classified in three categories (Table I):

I. *Disease occurrence measurements* describe causal relationships, evolution of disease occurrence or mortality over time. These measurements are represented by rates and proportions. A **rate** shows **how** fast the disease develops in a population. A **proportion** indicates **what** part of the population is affected.

II. *Association or effect measurements* show the strength of the association between a disease and a risk factor presumed to be the cause of the disease in question. These measurements are represented by the risk or the probability for an event to occur. The measures for demonstrating the association between a risk factor and a disease are represented by the relative risk and odds ratio (see Chapter 7, section 7.5.3 and 7.6.3).

III. *Importance or implication measurements* indicate the effect that a disease or a risk factor might have on a population.

Table I: Examples of different epidemiological measures

Disease occurrence	Association	Importance
Incidence	Relative risk	Excess risk
Prevalence	Odds ratio	Attributable risk
	Incidence rate ratio	
	Regression coefficient	

8.1 Measures of morbidity

The morbidity measurements are represented by prevalence, incidence, attack rate and duration of disease.

8.1.1 Prevalence

Prevalence(proportion) is the number of affected persons present in the population at a specific time divided by the number of persons in the population at that time.

$$P=\frac{No.of\ cases\ of\ a\ disease\ present\ in\ the\ poplation\ at\ a\ specified\ time}{No.of\ individuals\ in\ the\ population\ at\ that\ specified\ time}x1{,}000$$

The above formula shows the calculation of disease prevalence per 1,000 people.By multiplying it with 10,000 or 100,000, then it will be calculated the disease prevalence per 10,000 people, respectively per 100,000 people. Always, in the text, it must be specified the number of people that the prevalence is reported to.

The prevalence, in other words, represents an“image”of the study population at a point in time showing which individuals have the disease.

Often in medical literature the word “prevalence” is used in two ways:

- Point prevalence – the probability for an individual to be ill in a population at a specific moment in time;
- Period prevalence – the probability that an individual will ever have been ill during a period of time. The time period is arbitrarily selected and can represent one month, one year or 10 years. Usually the disease prevalence in one year is calculated.

Prevalence is an important and very useful measure to detect if a disease represents a health problem in a population.Therefore prevalence is a valuable indicator for the necessity of health services planning.

8.1.2 Incidence rate

The incidence rates are used to explore why diseases vary between groups in the population. The incidence is defined as the number of new cases of a disease that occur during a specified period of time in a population at risk for developing the disease.

$$I = \frac{\substack{No.of\ new\ cases\ of\ a\ disease\ occuring\ in\ the\ population \\ during\ a\ specified\ period\ of\ time}}{\substack{No.of\ individuals\ who\ are\ at\ risk\ of\ developing\ the\ disease \\ during\ that\ period\ of\ time}} x1{,}000$$

Like in the prevalence formula, the above incidence rate formula shows the calculation of disease incidence per 1,000 people, if it is desired to be calculated per 10,000 or 100,000 people, and then it must be multiplied by 10,000, respectively 100,000.When it is multiplied by 10,000 or 100,000, it must be specified in the text that incidence of disease was calculated per 10,000 people, respectively per 100,000 people. Also the period of time must be specified. The choice of the time period is arbitrary, it can be one week, one month, one year or 8 years, but when choosing the time period must be ensured that all individuals in the group represented by the denominator have been followed up and have been observed at risk through that entire period of time. The incidence rate calculated using a period of time during which all individuals in the population are considered at risk for the outcome is called cumulative incidence.

Incidence is a measure of events and so it can be stated that incidence is a measure of risk.

8.1.3 Attack rate

The attack rate is applied to a narrow population segment in a short period of time. It is used more in epidemics and represents the number of new cases of disease, occurring in the population at risk over a period of time, divided by the population at risk at the beginning of the study.

8.1.4 Duration of a condition

The duration of a disorder will, together with the incidence rate, determine how large a proportion of the population is affected by the disorder at any time, that is, the prevalence.

$$\text{Prevalence} = \text{Incidence x Duration of a disease}$$

Short duration of disease implies decreasing prevalence. An increased duration willlead to the increasing of prevalence even if the incidence rates remain constant.

8.2 Measures of mortality

8.2.1 Mortality rates

The mortality rate measures the frequency of death in a defined population and in a specified period of time. There are several types of mortality rates:

a. Annual mortality rate for all causes (AMR)

$$AMR = \frac{Total\ no.of\ deaths\ from\ all\ causes\ in\ 1\ year}{No.of\ individuals\ in\ the\ population\ in\ that\ year} x1{,}000$$

When the annual mortality rate for all causes is related to 1,000 people, the above formula is used. If the rate is desired to be related to 10,000 or 100,000, it must be multiplied by 10,000 respectively 100,000.

b. Age-specific mortality rates (ASMR)

When a restriction is placed on a rate, it is called a specific rate. If the restriction is represented by age, then the specific rate will be called age-specific mortality rate. Gender or race can also be selected as restrictions.

$$ASMR = \frac{\frac{No.of\ deaths\ from\ all\ causes\ in\ 1\ year\ in\ children\ younger\ than\ 10\ years\ of\ age}{No.of\ children\ in\ the\ population\ younger\ than\ 10\ years\ of\ age\ in\ that\ 1\ year}}{} x1{,}000$$

The restriction must be applied to both the numerator and the denominator.

c. Cause-specific mortality rates (CMR)

A diagnosis can be selected as a restriction. In this case the rate will be limited to deaths from a certain disease and will be called disease-specific rate or cause-specific rate.

$$CMR = \frac{No.of\ deaths\ from\ lung\ cancer\ in\ 1\ year}{No.of\ individuals\ in\ the\ population\ in\ that\ year} x1{,}000$$

Time must always be specified in any mortality rate.

d. Infant mortality

Infant mortality is the number of deaths among children under the age of one year per 1,000 live births during the same period.

e. Perinatal mortality

Perinatal mortality is the number of stillborn and deaths during the first week after birth per 1,000 born, including stillborn in the same period. Stillbirths are the number of dead babies per 1,000 babies born, including the dead during the same period.

8.2.2 Case-fatality rates (CFR)

This rate shows the percentage of people diagnosed as having a certain disease who die within a certain time after diagnosis.

$$CFR=\frac{No.of\ individuals\ dying\ during\ a\ specified\ period\ of\ time\ after\ disease\ diagnostic}{No.of\ individuals\ with\ the\ specified\ disease}x100$$

The difference between mortality rate and case-fatality rate is that in the first, the denominator is represented by the entire population at risk of dying from the disease (both those who have and who do not present the disease), while in the former, the denominator is limited to only the individuals who already have the disease.

Case-fatality rate is a measure for the severity of a disease.

8.2.3 Years of potential life lost (YPLL)

This mortality index is used to measure the premature mortality or early death. It is calculated as the sum of the differences between life expectancy and the age at which death occurs, if it is premature. The life expectancy is calculated as the difference between a predetermined age at death and the age that the individual has died. Usually the predetermined age at death is set at 65 years. For example, if a 5 years old boy dies, his life expectancy is 65-5=60, meaning he lost 60 years of life. The YPLL rate is the years of potential life lost divided by population with age under 65 years and multiplied by 10^n.

8.3 Validity

In science and statistics, validity represents the extent to which a concept, conclusion or measurement is well-founded and in accordance to the reality; validity could be defined as “true” or “reality”. In scientific research field validity refers to the probability of a study to be able to scientifically answer the questions it is intended to answer.

*External validity*represents the probability that the results obtained from a sample analysis (estimator) can be generalized to the entire study population. The external validity can be fulfilled only if the study presents internal validity.

Internal validity is accomplished if the measured data reflect the reality about the studied subjects. In other words, the results are not due to chance, bias, confusion or a wrongly chosen study protocol.

8.4 Interpretation of the statistical results

If the null hypothesis is accepted only because it is not rejected, the acceptance of the null hypothesis might prove to be a wrong decision due to different reasons; one example would be the small amount of data.In this case even if the obtained *p*-value islower than the alpha significance limit (0.05), the researcher must decide if "statistically significant implies or not a scientific significance". For an accurateinterpretation of the statistical results, besides value p (see Chapter 21), another decision tool was introduced: the confidence interval.

The confidence interval

The sample based mean will never be equal with population mean. The size of the difference between the two means depends on the size and variability of the sample. If the sample is small and presents a high variance (the group included in the sample is heterogeneous), the calculated sample mean might be very different from the true population mean (it is not representative for the entire population). If the sample is large and with low variance (the group included in the sample is homogeneous), the calculated sample mean can approximate more accurately the population mean.

The statistical analysis uses sample size and variability (standard deviation) to generate a confidence interval (CI) for the population mean. If the significance level of the value p (α) is chosen 0.05, then the confidence interval level is 95% (see Chapter 20).

Based on the value p and CI obtained after applying the proper statistical tests (see Chapter 21) the decision that the results are "scientifically significant" will made as follows:

- If the obtained results showa value $p<0.05$ and a wide confidence interval (confidence interval limits, lower and upper limit, are distant from the calculated sample mean), then statically significance does not necessarily imply scientific significance.
- If the obtained results show a value $p<0.05$ and a narrow confidence interval (confidence interval limits, lower and upper limit, are close to the calculated sample mean), then there is a scientific significance.
- If the obtained results show a value $p>0.05$ and a wide confidence interval, then it can be stated that there is no scientific significance.
- If the obtained results show a value $p>0.05$ and a narrow confidence interval, then a scientific significance might be considered.

If the obtained p value is greater than 0.05 could lead to a debate on scientific significance of the results. This is due to possible biases involved in sampling, data collection and measurement.

8.5 Research study bias

Epidemiological studies measure different population characteristics. The calculated parameters can be represented by rates, indicatorsof prevalence, incidence and measurements of association between exposure and disease. Due to the fact that the research studies involve human beings, constraints regarding study subjects, ethical aspects, or study design, lead invariably to bias (errors).

8.5.1 Accidental bias

Accidental bias represent inconsistencies due to chance between measurements calculated based on a sample and the actual population parameters. There are three sources of accidental bias: individual biological variability, sampling error and measurement error. These errors cannot be completely eliminated but can be minimized by careful measurement of exposure and outcome, increasing sample size or by validating measuring instruments.

8.5.2Systematic bias

Systematic bias occurs when there is a tendency in obtaining results that are systematic different from the true values. The most important biases from this category are: selection bias and information bias.

Selection bias represents errors that occur in selecting subjects within groups the study. For example, in the research studies that investigate a possible association between a certain exposure and a disease, if the subjects included in the study groups were selected in such a manner that an apparent association is observed, even if in reality, the exposure and disease are not associated, then the apparent association is the result of a selection bias. Selection bias results from systematic error in selecting individuals in one or more of the study groups.

Another type of selection bias is the *exclusion bias*. It occurs when researchers, studying an association between an exposure and a disease, intentionally exclude from the study those individuals who were exposed to the risk factor but do not present the disease.

Information bias occurs when the methods of data collection are not properly chosen and the information gathered are incorrect or the quality and the amount of data is different for the study groups. When the subjects are included in the wrong study groups due to inaccuracies in the methods used to gather information *misclassification bias* is introduced in the study. For example, a misclassification bias occurs in a case control study, due to the methods of data collection, like inadequate or incomplete medical records, or the use of an inaccurate diagnostic test. This will lead to

a misclassification of the study subjects, the individuals with disease (cases) will be misclassified as controls and individuals without disease (controls) will be misclassified as cases. Also a misclassification bias can occur in assessing the exposure. An individual may be misclassified as unexposed when in fact he was. This happens more often when the exposure status is assessed by interview, because subjects might not know that they were exposed or believe that the exposure did not occur.

Misclassification bias can be classified as:

- Differential misclassification happens when the misclassification rate is greater in one of the study groups. This usually occurs when the exposure status is assessed, and will lead to a misclassification of the individuals with disease. Due to this bias, they will be considered as being more often exposed to the risk factor.This happens because the individuals with the disease tend to remember even the events that are not related with the disease (recall bias).
- Non-differential misclassification results from inaccuracies in the obtained information on any study groups due to inadequatedata collection methods used.

Recall bias occurs when people with a disease enhance information recall regarding the exposure. These individuals tend to remember more events or phenomena than individuals without the disease, not related to the exposure, but which might justify their disease status.

A related type of bias is represented by the *reporting bias*. This type of bias occurs when a person is aware of the fact that he/she is exposed, but hesitates in reporting the exposure due to different beliefs or perceptions.

8.5.3 Confounding

*Confounding*often emerges in the research studies that investigate a possible association between a risk factor and a disease. Confounding implies the presence of other risk factor, besides the studied one, that might be associated with the disease in question. For example, in a research study is investigated the association between the risk factor “R” and a disease “D”, a third factor “F” is a confounder if the following statements are true:

- The factor “F” is a known risk factor for disease “D”;
- The “F” is associated with factor “R”, but is not a result of the factor “R”.

Confounding might lead to false results and drawing the wrong conclusions like for example a certain risk factor is actually a protective one.

Chapter 9 – Displaying results

Displaying results is an important part in scientific research. All results obtained in a research study must be displayed in an adequate manner for several reasons:

- The results helps to communicate the study conclusions:
  - If the study hypothesis will be confirmed or rejected;
  - If the conducted study led to the discovery of something with a great significance in medical science.
- Based on the results other researches can assess the conducted study.

Methods of communicating and displaying analyzed data

Text

The text is the most common method used to communicate and display the analyzed data. The wording of the text must keep a balance between academic and scientific rigour and the understanding level of the readers. In the text the most important results must be highlighted.

Tables

The tables offer a useful mean of presenting large amounts of detailed information in a small space. It is mandatory for a table to be self-explanatory. The title and the column headings must be brief. If numerical data are displayed the unit of measurement must also be displayed. A table has five parts (Figure 1):

- *Title* indicates the table number and describes the type of data the table contains.

It is essential to number the tables, in this way when interpreting and discussing results in text, it will be much easier to refer to a certain table.The tables can be numbered:

  - Sequentially as they appear in the text. Tables can be numbered by using Arabic numeralsor Roman numerals (Table 1, Table I).
  - By the chapter number followed by the sequential number of the table in the chapter (Table 1.1, Table 1.2).

The description of data follows the table number (Table 1.1: Respondents by age). In the title it must be specified the variable about which the information is given in the table. If a table represents data about two variables, in the table title the dependent variable must be mentioned first.

Example: Attitudes towards flu vaccine administration (dependent variable) by age (independent variable).

- *Stub* represents the subcategories of a variable, listed along the y-axis (the left-hand column of the table). Usually in the stub are presented the

subcategories of the dependent variable.Lists the items about which information is provided in the horizontal rows to the right.

- *Column headings* represent the subcategories of a variable, listed along the x-axis (the top of the table). Usually in the column headings are presented the subcategories of the independent variable.

In univariate tables the column heading is usually the "Number of respondents" and/or the "Percentage of respondents".

In bivariate tables the column headings contain the variable and its subcategories.

- The body of the table is represented by the cells containing the analyzed data. Usually the number of decimals is limited to maximum two.
- Supplementary notes or footnotes are displayed at the bottom of the table. There are few types of footnotes:
    - Source notes - The source is labeled by the word "Source:" and must specify the author and the book the table was taken from.
    - Other general notes
    - Notes on specific parts of the table: statistical parameters and abbreviation explain.

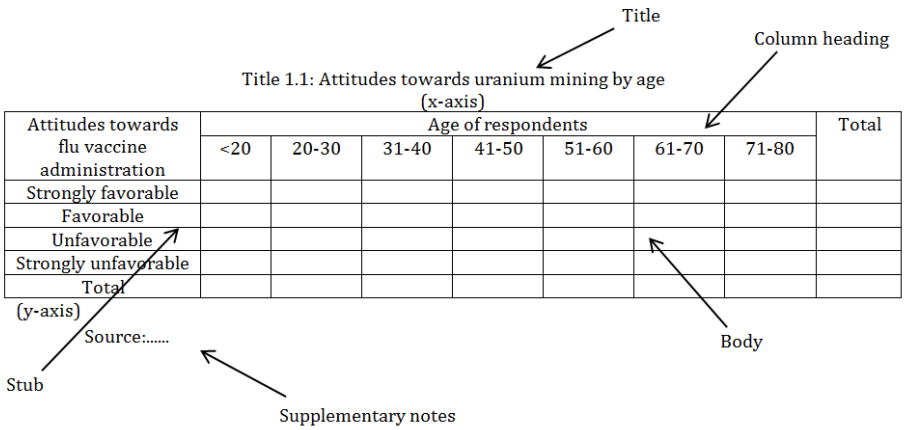


Figure1: The structure of a table

The tables can be categorized by their structure or by the type of contained data.

Accordingly to structure, there are three types of a table:

- Univariate (also known as frequency tables) – contains information about one variable

Age	No. of respondents	Percentage of respondents
<20 years	10	10%
20-30	20	20%
31-40	25	25%
41-50	25	25%
51-60	10	10%
61-70	5	5%
71-80	5	5%
Total	100	100%

- Bivariate (also known as cross-tabulations) – contains information about two variables

Attitudes towards flu vaccine administration	Age of respondents							Total
	<20	20-30	31-40	41-50	51-60	61-70	71-80	
Strongly favorable	5	5	2	1	2	2	4	21
Favorable	2	8	7	5	3	3	1	29
Unfavorable	2	4	12	14	4	0	0	36
Strongly unfavorable	1	3	4	5	1	0	0	14
Total	10	20	25	25	10	5	5	100

- Polyvariate or multivariate – contains information about more than two variables

Attitudes towards flu vaccine administration	Number of respondents										
	<20		20-40		41-60		61+		Total		
	F	M	F	M	F	M	F	M	F	M	T
Strongly favorable	3	2	5	2	1	2	3	3	12	9	21
Favorable	1	1	9	6	6	2	3	1	19	10	29
Unfavorable	2	0	7	9	7	11	0	0	16	20	36
Strongly unfavorable	0	1	3	4	2	4	0	0	5	9	14
Total	6	4	24	21	16	19	6	4	52	48	100

Accordingly to type of contained data, tables can be categorized as:

- Containing frequency distributions - It includes distinct values and their frequencies

Blood glucose levels (mg/dL)	Number of investigated patients
95	11
115	20
120	30
127	15
126	12
100	8
130	4
Total	100

- Containing frequency distribution classes

Blood glucose levels (mg/dL)	Number of investigated patients	Percent	Cumulative percent
95	11	11%	11%
115	20	20%	31%
120	30	30%	61%
127	15	15%	76%
126	12	12%	88%
100	8	8%	96%
130	4	4%	100%
Total	100	100%	

- Contingency tables are used to display the results in research studies that investigate a possible association between an exposure and a disease. This type of tables is used to calculate the value of relative risk and odds ratio.
    - 2x2 (two rows and two columns)

	With disease	Without disease	Total
Exposed	73	38	111
Non-exposed	45	64	109
Total	118	102	220

    - Larger (multiple rows and columns)

	With disease	Without disease	Total
<20 years	10	36	46
20-40	47	51	98
41-60	63	86	149
61-80	12	45	57
Total	132	218	350

Graphic display

There are two ways the results can be graphically displayed:

1. Illustrations and photos
2. Graphs

*Illustrations and photos*must present onlythe most significant results. If the picture illustrates a microscopic image, it must be sufficiently enlarged so the studied aspects are easily observed. Also it is mandatory to draw arrows to indicate the most important elements.Microscopic slides illustrated must be numbered.For all images clear accurate explanations must be given.

Graphs are computer-assisted data representation,which allow a visual assessment of the studied items.A proper graph presentation shows the study results in such a way that are easy to understand and interpret.

A graphic presentation is constructed in relation to two axes:

- Horizontal axis is called the "abscissa" or x-axis
- Vertical axis is called the "ordinate" or y-axis

When a graph presents:

- One variable – usually the subcategories of the variable are displayed along the x-axis and the frequency or count of that subcategory along the y-axis.
- Two variables – one is displayed on each axis

A graph must have a title that describes its contents and also the axes should be labeled. A graph should be drawn to an appropriate scale (not to small, nor too large). The choice of scale should result in the spread of axes proportionate to one another.

The type of graph depends upon the type of data that are displayed:

- For qualitative variables – only bar charts, histograms or pie charts;
- For quantitative variables – in addition to the above, line or trend graphs, scatter plot.

Types of graphs

a. The **histogram** is a particular type of graph. It represents a series of rectangles drawn next to each other without any space between them, representing the frequency of a category or subcategory.This type of graph is used to present the number of elements in a class by an area. There two histogram types, two-dimensional histogram and three-dimensional histogram(Figure 1):

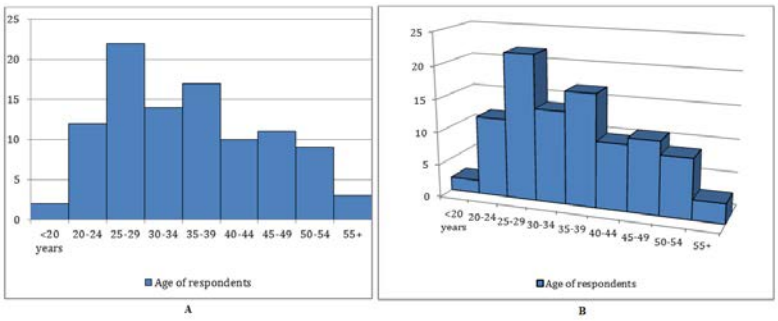


Figure1: A. Two-dimensional histogram; B. Three-dimensional histogram

b. The **bar chart** is identical to a histogram, except that in a bar chart the rectangles are spaced, thus indicating that the displayed variables are qualitative. This type of graph is useful to show two variables by comparison. It can present

- One variable over several stages using horizontal or vertical columns (Figure 2).

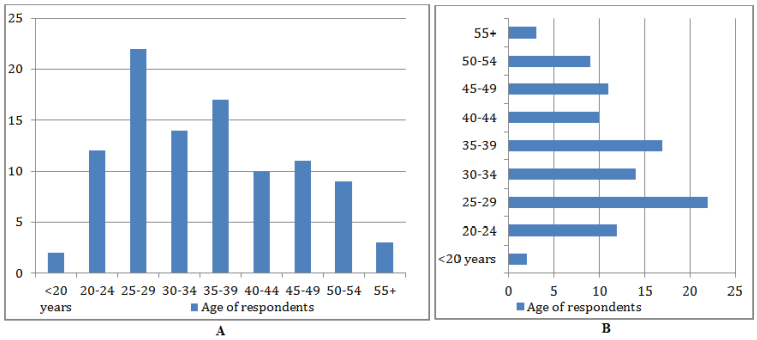


Figure 2: A. Bar chart with horizontal columns; B. Bar chart with vertical columns

- Two or more variables showing so any differences between the studied variables (Figure 3A). Also the obtained data can be presented (Figure 3B).

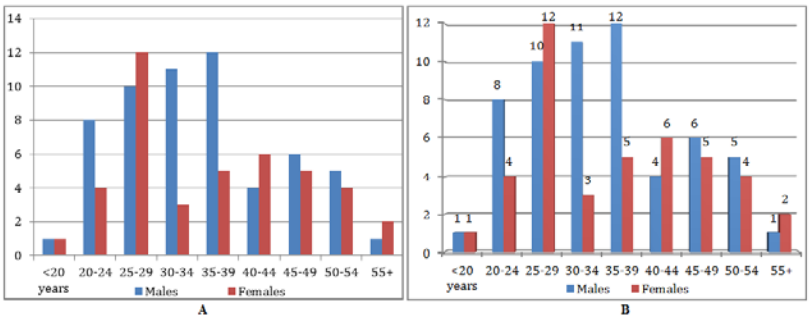


Figure 3: A. Bar chartpresenting two variables; B. Bar chartpresenting obtained data

Like histograms, bar charts can be drawn in two or three dimension.

c. In the **stacked bar chart** the section of a bar shows the proportion of the variables they represent in relation to one another (Figure 4). It can present two or more variables.

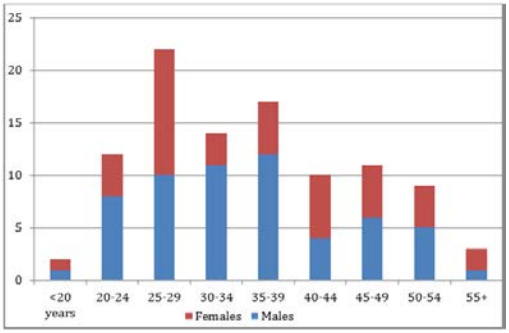


Figure 4: A stacked bar chart

d. In the **100 per cent bar chart** the subcategories of a variable are converted into percentages of the total population.A bar represents 100% and its division represents the percentage of each subcategory of the variable (Figure 5).

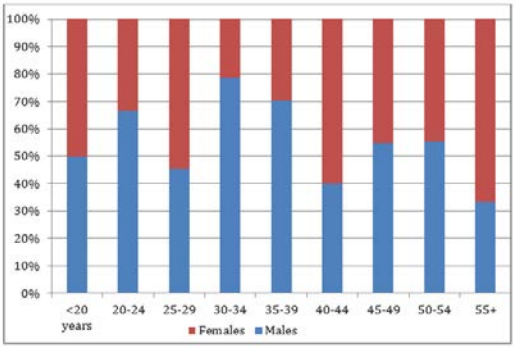


Figure5: A 100 per cent bar chart

e. The **pie chart** shows components of a whole. It is usually used to present qualitative variables (especially binary variables), but can be drawn for quantitative variables if data are grouped into categories. The pie chart represents data as a circle. The circle is divided into sections (slices) accordingly with the size of each subcategory of a frequency distribution (Figure 6). It can present numerical data or percentages. This type of graphic representation can be two or three dimensional.

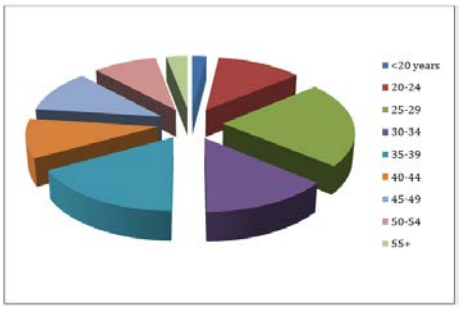


Figure6: A pie chart

f. The **area chart** is used to highlight the total magnitude of the subcategory in relation to other subcategories (Figure 7). This type of graph is used to present variables measured on an interval or a ratio scale.

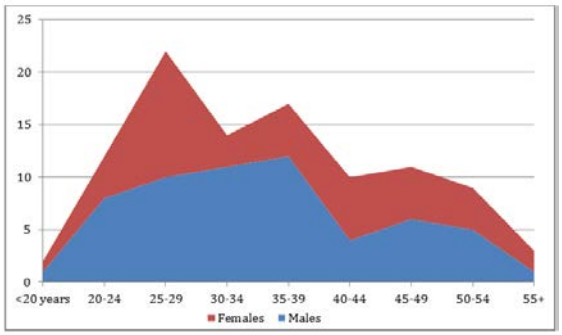


Figure7: An area chart

g. The **area map**shows the location of a disease/event/behavior by place using different colors to represent the various levels of the disease/event/behavior. It is used more often by epidemiologists to show the magnitude of a pandemic (Figure 8).

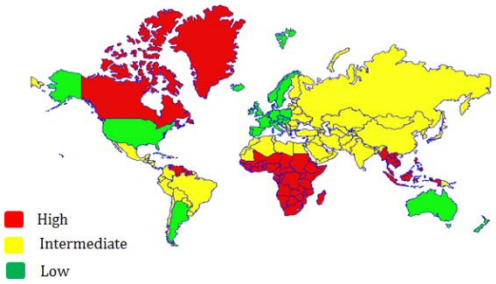


Figure 8:Global pandemic for hepatitis B

h. The **line diagram or trend curve** is useful to show the disease/event evolution in time. It can be drawn for data collected at a specific point in time (Figure 9) or over a period of time (1980-1985, 1986-1990, 1991-1995). Each frequency is marked as a dot, and then dots are then connected with straight lines.

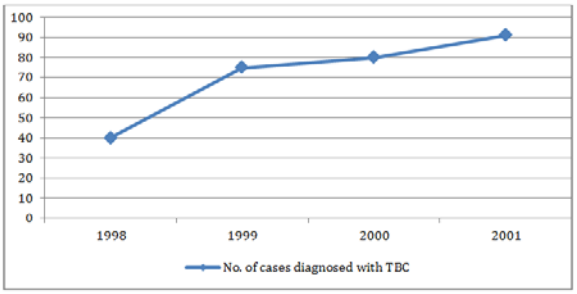


Figure9: The line diagram or trend curve

i. The **scatter plot** represents a useful summary of the association between two numerical variables. This type of graph is used only to present quantitative variables. Data for both variables are displayed as dots correspondent to their values on both axes (Figure 10).

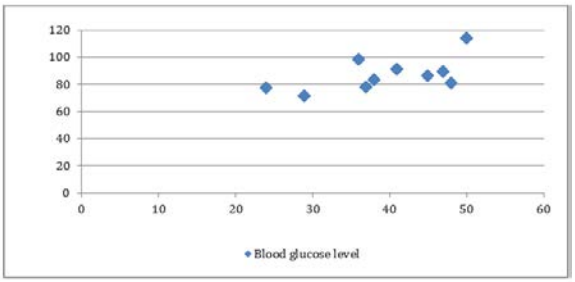


Figure 10: A scatter plot

Chapter 10 – Structure of a scientific paper

A scientific paper should respect a specific structure thus all published articles will present uniformity, a better dissemination of medical information, an understandable and accessible message. The most used structure when writing an article is the IMRAD structure. The acronym comes from the articles headings (Introduction, Material, Results And Discussions). To this structure are new headings were added (Conclusions and References). The majority of the scientific journals prefers this structure and requires that all articles sent for submission to be written using the IMRAD structure.

The IMRAD structure refers to the structure of the actual scientific work, but the article must also present a title, authors' namesand institution affiliation, abstract and keywords.

A scientific paper should be structured as follows:

- Title
- Authors and institution affiliation
- Abstract, keywords
- Actual scientific work

*Title*should be interesting,brief, easy to understand and without abbreviations. The title should be very accurate and conciseand to announce the content of the article. Based on the title an article will or not be read.

Authors' names and institution affiliation

In the byline the authors first and last name is listed. For men, first is written the last name followed by first name initial (example: Anderson P) and for women, first is written the first name followed by last name (example: Maggie Peterson). Usually there are no more than 7 authors on a scientific article.

The institution (academic) affiliation for each author is indicated by a superscript number on each author name (Anderson P[1], 1 – Department of Anatomy, University of Medicine and Pharmacy, Bucharest, Romania).

The order of authors in the byline must be accordingly to the importance, volume and contribution of each author to the research. To be able to take responsibility for the article each author should have participated sufficiently in the study. Ethics of scientific research does not support the inclusion of authors who do not fulfill the criteria for authorship.

Abstract and keywords

The ***abstract*** is an important part of a scientific article, based on it is decided if the actual scientific work will be read.The abstract should have a similar structure as the actual scientific work: Aim, Material and method, Results and Conclusions. In the

situation when only the abstract of the article is available, based on this part, a reader should be able to determine why and how the study was conducted, the most important results and conclusions. When writingthe abstract short sentences should be used andthe verbswill be in the past tense.The abstract must not contain abbreviations, tables, figures and references and should be kept short. The medical journals where the article is going to be submitted indicates in "Instruction for authors" section, how many words an abstract should contain, most often not more than 300 words.

*Keywords*are the words that express the key elements of the research. They are placed below the abstract and usually there areno more than 6 words. Most of medical journals require that the keywords should be MeSH (Medical Subject Headings) terms. MeSH is a comprehensive controlled vocabulary of the United States National Library of Medicine's (https://www.nlm.nih.gov/mesh/MBrowser.html). This vocabulary is used indexing journal articles and books in the life sciences. It consists in sets of terms organized in a hierarchical structure that allows a user to search at various levels of specificity.

*Actual scientific work*is structured as follows*:*

• *Introduction* should give the reader the opportunity to make a clear and concise idea of the research subject. That it is why the introduction must be to be precise and informative, so the reader will understand why the study was conducted. Also in this part it must be stated the current state of knowledge of the studied topic. The introduction can be divided in two parts. In the first part should be presented the general aspects of the studied subject and eventually a brief history regarding the studied theme. The second part will present the particular aspects of the problem studied and the last phrase will be represented by the study aim.

When quoting another author the verbs used will be in the past tense and when there are described general aspects and already known facts, the verbs used will in simple present tense.

• In *material and method*it is described how the research study was conducted. A qualified reader should be able to repeat the study in the same manner using the information given in this part. This part provides details on type of study, sources, methods of data collection, study subjects (sample, representativeness, inclusion-exclusion criteria), type of variables (what are we study), used techniques, reagents, statistical methods applied (results validation) and statistical programs used in data analysis.

No results and discussion should be presented in material and method. The verbs used must in the past tense.

• *Results*representthe final stage of the research and the basis for further discussions. The study results should be presented systematic, concise and precise, but without explanation of the results. The results can be displayed using graphs, tables

and pictures. Results should give answers to the study objectives. The negative results or the results that invalidate the scientific hypothesis should not be removed.The results obtained following the statistical analysis (significant, insignificant p) must be mentioned.

The verbs used should be in the past tense.

• In the*discussion*part it is interpreted the work performed. First it is statedif the research aim was reached, if the used methods were properly chosen and if the sample was representative (number of patients). In this part the results are interpreted and explained. The obtained results should be compared with other research results. Also it should be mentioned if the hypothesis was accepted or rejected (commenting on the statistical results) and the theoretical and practical involvements of the study findings and the personal contribution of the authors. This part should not contain affirmations unsustainable by the results of the study, criticism regarding other authors' scientific articles and repeating results (graphs, tables). Discussions part should not exceed 50% of the length of the article and it is better to use personal pronouns "we" or "I"when presentingone's own work to avoid confusion with the results of other researchers.

• *Conclusions* must be clear and concise, based on conducted research and highlight the findings of the study. The results will not be repeated (no tables, graphs), the evidence for each conclusion will be summarized. The conclusions should not be in an exaggerate number.

References

The references are a set of articles and books on the subject and aims to justify any fact stated in the paper. It must include the most important and recent works, only consulted and published papers. For the cited papers in abstract must be specified that those are in abstract.

Essential elements:

• For the **articles**: the name of the authors, title of the article – in original language, journals name (Index Medicus), year of publication, volume of the journal, number of pages

• For the **books**: the name of the authors/editors, if they are more than 3, the first 3 are written and then*et al.*, book title, edition number, publishing house, city, year of the publication, pages

• For the **book chapters**: the name of chapter authors, chapter title, In: book authors, book title, year of the publication, pages

Reference in text citing and the elaboration of the references list can be done by using one of three systems:

1. Vancouver style: in the text, references are numbered with Arabic numerals and placed in square brackets. If there are multiple references on the same paragraph, their correspondent numbers are placed in the square brackets in

ascending order with commas between them. If the same reference is cited multiple times, it keeps the correspondent number from the first citation in text. In references list the authors are listed sequentially as they appear in text.

2.Harvard style: in text, references are cited by placing in round brackets the name of the author and the year of publication.In the reference list authors are listed in alphabetical order without serial number.

3.Alphabetic-numeric system: in text the references are numbered with Arabic numerals and placed in round brackets sequentially as they are cited and in the reference list the authors are listed in alphabetical order.

Chapter 11 – Ethics in scientific research

11.1 Ethical theories and frameworks

Although moral frameworks hold that values are universal,still permit leeway in application.

The main differences of opinion in ethical theories are about whether the ethical evaluation is actually an evaluation of acts, states of the people or of the world and whether the rightness or goodness of an act is essential in the evaluation of it.

The three most often used ethical theories or frameworks are:

Consequentialism is the simplestethical theory.It holds that goodness is primary and defines the right act as that which maximizes goodness.The main criticism of consequentialist theory is that in some situations, it can justify actions that most people think are wrong. For example the sacrifice of the patients interests, perhaps even lives, in order to obtain important scientific knowledge.

Deontological ethics are in some ways the opposite of consequentialist theories.Deontological theories of ethics claim that the main concern is if an act is right, not if it has good consequences. The problem is that various deontological theories differ when it comes to determine how actions are established as to be right, or when it is no right act how to choose between two wrong acts.

*Virtue ethics*focuses on the person performing an act and not on the act itself. A morally right action is defined as the action performed by a virtuous person.The main criticisms of this theory are that there is no unanimously approved list of virtues and that the theory does not recognize that even a morally evil individual sometimes is capable to perform a good act.

11.2 Scientific worknorms

Knut Erik Tranøy a Norwegian philosopher, with major contributions in ethical and especially medical and science fields argued that a scientific work requires three different kinds of norms:

- Internal norms:
  - Epistemic norms that should guide the activity of each researcher. In this type of norms, Tranøy includes truth seeking, simplicity,coherency, consistencyand testability.
  - Social norms focus on the collaboration between researchers or between research groups. In the social norms Tranøy includes openness, open-mindedness and honesty.
- Linkage norms refer to utility, fruitfulness and relevance.

- External normsrepresent the limits society places on scientific conduct.

11.3 Scientific misconduct

Scientific misconduct is defined in accordance with the US Office of Research Integrity (ORI) definition as:

"Fabrication, falsification, plagiarism, and other practices that seriously deviate from accepted standards when proposing, conducting and reporting research".

Eric T. Poehlman,a scientist in the field of human obesity and aging, was the first biomedical scientistin the United States sentenced to 366 days in prison "because his actions led to a loss to the government, obstruction of justice, and abuse of a position of trust". When investigated by the ORI (Office of Research Integrity) he had been found to have falsified or fabricated data in at least 12 publications and 19 grant applications. AlsoMr.Poehlman was forbidden for life to apply for or to receive federal research grants.

Scientific misconduct is most severe when it alters the truth claims of scientific findings, as when itcompromises the cumulative nature of scientific work and development and may lead to applications in medical practice that are harmful to patients.

Fabrication

Fabrication refers to the invention of data and without doubt is the most serious form of misconduct that affects truth claims. It ranges from the invention of allreported data to the invention of just a part of it.

Theintentionally biased analysis of data to obtain the "desired" results or the suppression of unwanted results represents other types of misconduct from this category. Both lead to misleading information in scientific records.

Plagiarism

Plagiarism refers to claiming another person work to be one's own. In published papers plagiarism may be:

- Total when fully copied or translated works of other authors who have been previously published elsewhere, are submitted for publication as one's own;
- Limited when copied parts of the works of other authors, either in original form or slightly modified, are integrated into one's work

For exampleElias Alsabti, an Iraqi medical researcherwas suspected of having published about 50 to 60plagiarized full articles.

*Theft of ideas*is another form of plagiarism. An example of theft of ideas is the situation where a peer reviewer stalls an article, during which he studies the idea of the article in question and publishes it elsewhere as an original study.

Authorship issues

Disputes about authorship are probably the most common of conflicts within research groups.

The International Committee of Medical Journal Editors (ICMJE) promulgated "Ethical Considerations in the Conduct and Reporting of Research: Authorship and Contributorship"a list withauthorship rules that must be followed in order to prevent disputes among authors.The rules regarding the author status states:

1.An author to be listed in an article byline must have had major contributions in conception and study design, data collection, data analysis and results interpretation. Also an author should be involved in the paper elaboration or in the critical review of the content or in ensuring the final approval of the version that will be published.

2.A person listed as an author should meet the above conditions. If the person fulfills only one of the conditions stated above does not present the qualification to be an author.

Authorship issues

Exclusion from authorshiphappens when a person is excluded from the byline even though that person desires to be listed and fulfills the criteria for authorship. An unjustified exclusion from authorship is considered theft.

Gift authorshipis when someone even if has not fulfilled the criteria to be listed as an author is still offered authorship.

Authorship achieved by coercionhappens when a person, usually a senior researcher claims to be author to all articles that are about to be published, not taking into account if his/hers criteria for authorship are fulfilled.

Unsolicited authorshipis whena personwithout their knowledge or consent is listed as an author in the byline.Usually it happens simultaneously with ghost authorship, when the person who actually has the right to be listed as an author is not.

Refusal to accept responsibility as an author when other misconduct is detected

When a person is listed as an author to scientific paper must also accept responsibility for the published paper. Sometimes, however, when a serious form of misconduct is discovered in the paper, some authors renounce responsibility.

Salami, imalas and duplicate publication

Salami publication is when a researcher divides obtained results insuch small pieces so they could be published separately.

Imalas publication is the sequential publishing of actually the same results each time but with a just few data added in the analysis.

11.4 Basic principles of human research ethics

The main principles of research ethics state that the potential harm of the subjects included in a research study must be minimized. Also all participants in a study must be volunteers and understand correctly what research project is about and to be allowed to withdraw from the study if this is their wish.

There are four ethical principles for medical research on human subjects which lists all valid values and ethical standards:

I. *The principle of interest and benefit in research* refers to maximize benefits and reduce risks. A research should provide benefits to the population as a whole and these benefits should be reported to the risks to which subjects may be exposed. The benefit shall be assigned to the study subjects.

II. *The principle of risk – benefit* refers to the balance between benefit and risk (the risk-benefit). Both benefits and risks should be clearly defined at the beginning of the research.The maximum risk in a research should be at least equivalent to which that patients are exposed in the daily clinical practice.

III. *The principle of respect for persons* states that the participants in a medical research must be considered as autonomous agents. The study protocolshould specify the procedures that will be applied to ensure personal privacy and data confidentiality both during and after data collection.

IV. *The principle of justice* refers to a proper and fair distribution of both benefits and risks.

All participants on a study must sign the informed consent before their inclusion in the research. The informed consent must describe the aim of the research, inclusion and exclusion criteria, what methods will be used to assess the disease and the exposure status, the benefits and the potential risks, the right to withdraw from the study and voluntarism. Researchers must to ensure that all study participants understood the information presented in the informed consent before they signed it.

Persons unable to consent

There are situations in which future research participants are legally or factually unable to consent.

Research on children

When minors are future participants to a research study, their parents or guardians can sign the consent, if the research project is not against the best interests of the minor. Even so, the child must be informed using informational materials and

techniques accordingly to the child age and his/herswish to participate or not in the study should be taken into account.

Older children should be asked to sign the informed consent and in case of a refusal this must be accepted contrary to any opinions of the parents or legal guardians.

Research on permanently incapacitated adults

Sometimes the permanently incapacitated adults do not have a legal designated person to decide in their behalf. These personscan be either people that previously were competent or they have never been competent. Inclusion in a research study can be made only if these individuals have a legal designated person who will decide for them only after this person understands the best interests of the incapacitated person. If the incapacitated person was previouslycompetent his/hers previously expressed wishes or values should be taken into account.

Vulnerable research participants

There are situations when research participants either do not have the proper education to understand the informed consent, or are socially powerless or somehow dependent on the researchers. In this case a clear justification of the reason why these individuals should be included in the research study.

Research ethics committees (REC)

The research ethics committees must ensure that the research ethics are respected. Researchers must submit the research study protocol, including the informed consent signed by the study participants, to a research ethics committee for approval before the research study starts.

The members of the research ethics committee will assess the scientific validity of the research study proposed, any ethical problems contained by the study protocol and if the informed consent is signed by each study participant. After the evaluation of the study protocol, the research ethics committee members will decide if the research project can start. In many countries a research study cannot begin without the approval of the research ethics committee. Also there are scientific journals that require this approval in order to publish the research.

11.5 Ethics in animal research

The two major problems in the ethical evaluation of animal research refer to if it is always justifiable to use animals in research, and if not which are the condition that allow the use of animals.

The official regulation of animal research has two aims:

- To ensure that the suffering level of the animals included in a research is minimum;
- To evaluate research projects to ensure that in each case there is an acceptable balance between animal suffering and benefits

For reaching these aims is the 'Three Rs' approach:

- "Refinement" refers to the improvement of all aspects of the animals' lifetime experience in a research, to minimize suffering and improve welfare.
- "Reduction" refers to the any strategies that useas fewer animals as possible in each experiment without compromising scientific outcome and the quality of biomedical research, and also without compromising animal welfare.
- "Replacement" refers to the use of methods that stillallowto be achieved a given scientific purpose without conducting experiments on living animals.

In animal research, is still debated if whether a classification of different kinds of protection for different kinds of animals is required and which are the basis characteristics for an increased protection.

A typical classification of increasing protection can be:

- Not protected:
    - Non-vertebrates.
- Increasingly protected:
    - Vertebrates;
    - Mammals;
    - Non-human primates;
    - Great apes.

Many countries do not approve or strongly discourage research on great apes.

Chapter 12 – Population. Samples.

Mathematical statistics is a science specialized in drawing general conclusion using a limited number of observations. Thus, with a small amount of data the researcher who uses statistical calculations can explain or describe mass phenomena, which influences a large number of people. This process is called extrapolation from a *sample* to a *population.*

A *population* is represented by all the individuals of interest relative to a problem or phenomena which is targeted by a study or experiment.

A *sample* represents a subset of the population, a small number of individuals selected from that population which usually are intended to represent the entire population with regard to the problem or phenomena under study.

In the case of medical studies, the population is often determined as all the individuals that are suffering from a certain disease and live in a certain region, a country for example. Other approaches narrow down the population of a medical study to all the patients of a certain hospital or medical facility. Regardless of the definition we use, selecting a sample from that population allows us to work with a limited amount of data and draw conclusions for a larger group.

In order for such conclusions to be correct, the sample must be *representative* for the population. A sample is representative for a population if every element of the population has the same chance of being selected in the sample as any other element. This process is called random selection and it assures the representativity of the selected sample.

A simple example might illustrate the above mentioned concepts. Let's assume that we wish to measure the average height of all the students in a university. In order to do that we either measure each and every one of the students – that can be impractical – or we select a representative sample of individuals and measure this smaller group. The easiest way to do this would be probably to walk into one classroom and measure the height of all the students that are attending a lecture. But this would violate the principle of random selection, because the chances of students not attending that particular lecture to be included in the sample are clearly not equal to the chances of those who attend. The correct way of selecting a sample would be to obtain a list with all the students, and have a computer randomly select a number of those individuals. After selecting the sample, we can measure the height of the selected students, calculate an average height of the sample and then *generalize* the result to the entire group.

Chapter 13 - Variables.

When conducting a research, scientists are interested in specific characteristics of the individuals selected in the study sample in order to generalize a conclusion regarding the whole population. These characteristics can have different values for different individuals or they can change in time or as a result of an external influence. Such characteristics are called *variables.*

There are two types of variables depending on the values they can take.

Discrete variables represent separate indivisible categories. There can be no other values between two adjacent categories. For example the numbers on a dice represent a discrete variable. There can be no intermediate value between 3 dots and 4 dots on a dice.

Continuous variables present an infinite number of possible values between any two observed values. For example height or weight are continuous variables, if the measurement instrument used provides a sufficient level of detail. Weight can be measured in kilograms, grams, milligrams, fractions of milligrams etc. So theoretically there can be an infinite number of values between 45 kg and 46 kg.

13.1 Types of measurement scales.

In order to statistically process data we have to record it. Recording the data means assigning the members of our selected sample to various categories depicted by the variables that we observe. The possible values that can be measured for a variable represent its *measurement scale*. There are several types of measurement scales according to the relationships between the categories of a scale.

13.1.1 Nominal scales

These scales categorize a variable without indicating any sort of order or quantity. Such a scale simply names the possible categories for a variable.

Examples:

- Gender: male, female
- Civil status: married, divorced, widow, single
- Eye color: brown, green, blue

These types of variables can be coded as numbers (0 – female, 1 – male) but they cannot be processed as quantity or order between categories.

13.1.2 Ordinal scales

In this case the values represent hierarchies and offer information about the place or the order of a category within the scale. The values can be compared as "more", "less" or "equal".

Examples:

- The place of a certain competitor in a sports event: first, second, third, …, last
- Type of smoker: occasional, regular, heavy smoker
- Degree of illness severity: light, medium, severe

This type of scale allows us to determine whether there is a difference between two measurements or individuals and gives us information about the direction of the difference. However, no information is provided about the magnitude of the difference. The competitor that took second place in a race could have been very close to the performance of the one on the first place, or the difference could have been tremendous. We simply cannot know if we are using an ordinal scale.

Variables that are measured using nominal or ordinal scales are also called *qualitative variables* because they categorize the individuals under study but they lack any quantitative information.

13.1.3 Interval scales

An interval scale is an ordinal scale that has equal sized categories. That means that the difference between any two adjacent points of the scale is always the same. An important characteristic of interval scales is that the zero point of the scale is arbitrary and does not represent the absence of the variable being measured.

Examples:

- C

elsius temperature scale – the difference between 21° and 22° is the same as the difference between 32° and 33° but 0° does not mean that the temperature is absent, in fact measurements can go below that point.

- I

Q measurement – a scale of equal intervals, with an arbitrary zero point. In fact one practically cannot score 0 on an IQ test, the zero point is just for reference.

When using interval scales, we can compare the measured values and we can compare relative differences. For instance we know that the difference between 20°

and 24° is two times larger than the difference between 30° and 32°. But we cannot say that 30° is two times hotter than 15°, because the zero point is arbitrary.

13.1.4 Ratio scales

Ratio scales are the most common type of scale used in medical research. A ratio scale has all the characteristics of an interval scale and has a meaningful zero point which represents the absence of the variable being measured.

Examples:

- Height or weight – if we measure height in centimeters, we can say with certainty that someone measuring 180 cm is 12,50% taller than someone measuring 160 cm.

- Time (incubation time, reaction time etc.)

- Number of occurrences of a certain event

Having a meaningful zero value on this type of scale enables us to measure the absolute value of a variable and to calculate percentages when comparing two values.

A ratio scale can be converted to an interval scale. For instance we might convert the measured values of height to the series of differences between each subject's height and the average height of the group. In this case, if the average is 170 cm, the subject measuring 180 cm will be assigned the value +10 cm and the subject measuring 160 cm will measure -10 cm on the interval scale. It is worth mentioning that by this conversion we can still compare differences but calculation of ratio does not make sense any more. It is incorrect to say that the person measuring +10 cm is twice as tall as the person measuring -10 cm.

Variables measured by interval or ratio scales are also known as *quantitative* variables.

Chapter 14 – Descriptive statistics. Frequency distributions.

The first results usually yielded by a clinical or an experimental study are large volumes of raw data, representing the measured values of the variables defined in that study for each individual or event. The first step in processing this data should be an attempt to organize the values and present them in a comprehensible form. This is the main objective of descriptive statistics.

One of the most convenient form of organizing raw data is represented by frequency distributions. A *frequency distribution* is an organized table or graph presenting the number of individuals or events corresponding to each category on the measurement scale that has been used to record the data. This type of data presentation is particularly suggestive in the case of quantitative variables, when we can present the categories of the scale in a meaningfully ordered fashion.

Example:

Let's assume that we have conducted a study in order to measure the height of the students at the university. In order to accomplish this we have randomly selected a number of 33 students and recorded their height measurement. The results of these measurements are presented in Table I.

Table I – Raw results for height measurement study

No.	Height (cm):	No.	Height (cm):	No.	Height (cm):
1	123	13	157	25	179
2	126	14	158	26	182
3	129	15	163	27	183
4	131	16	165	28	187
5	138	17	166	29	190
6	139	18	166	30	192
7	142	19	169	31	201
8	146	20	169	32	205
9	148	21	173	33	209
10	150	22	175		
11	153	23	176		
12	154	24	177		

In order to make the data as readable as possible we have already ordered the values from lowest to highest. So we can see from this data that the shortest person measured 123 cm and the tallest measured 209. These extreme values are known as the *minimum* and the *maximum* values of the measurements. The difference between them, which in our case is equal to 86 cm is called the *range* of the data.

At a first glance, the min and max values seem fairly extreme for a random group of students. We might wonder how many short people and how many tall people we have in this group. In order to answer this question, we have to count the values in a particular range. Let's assume that we consider as "very tall" all the subjects that measured over 200 cm. We have three such persons in our group. Furthermore, let's consider as being "tall" people everyone between 190 cm and 200 cm. We have two such subjects in our sample. If we continue this reasoning, we end up with a number of categories, and a number of subjects in each category depicting the *frequency* of the measured variable relative to each category. In other words, we are representing the data using a frequency distribution.

A very important decision to make in this process has to do with selecting the right intervals for grouping the values. These intervals are known as *variation intervals* or *bins* (the process is similar to grouping items into bins).

In order to produce a comprehensive representation, a number of rules have to be followed:

- The number of intervals (or bins) should vary between 10 and 15

- The limits of each interval should be greater than the precision of the measurement instrument

- The intervals should be equal in order to facilitate further processing f the data

- The intervals should be mutually exclusive (no overlapping margins)

- The intervals should be ordered from lower values to higher values

After constructing the frequency distribution, we can present it as a table or a graph called a *histogram.* The frequency distribution for the above mentioned example is presented in Fig. 1. In this case the size of the "bins" was set to 10 cm. So we have 3 occurrences for values less than 130 cm, 3 occurrences for measurements taller than 139 cm but under 140 cm, 4 occurrences between 140 cm and 150 cm and so on.

It is interesting to emphasize that the largest number of occurrences is concentrated around the middle of the scale, around 170 cm. If we add the frequencies of the bins corresponding to 160, 170 and 180 cm we get 15 occurrences, which represents almost half the persons in our sample.

Bin	Frequency
130	3
140	3
150	4
160	4
170	6
180	5
190	4
200	1
210	3
More	0

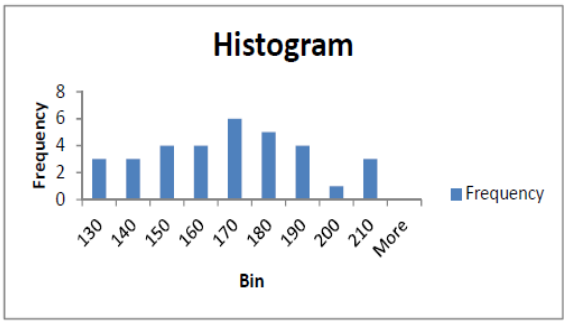


Figure. 1 – Frequency distribution of height measurement – bin width: 10 cm

Why is selecting the proper variation interval so important? Let's take a look at two other ways to present the same data. In Fig. 2 we can see the frequency distribution for the same data, but with large bins. In fact, we have only three variation intervals: lower than 130 cm, between 130 and 190 cm and between 190 cm and 250 cm. The representation is correct, but offers very little detail about the group being studied.

Bin	Frequency
130	3
190	26
250	4
More	0

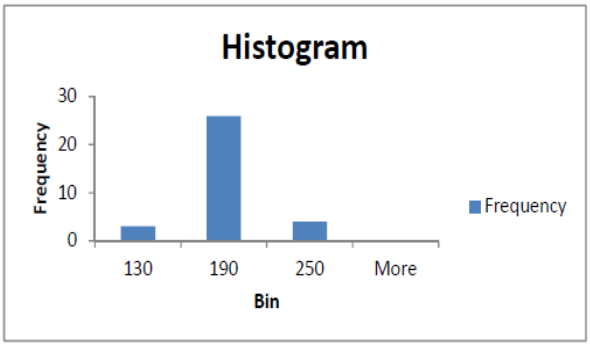


Figure. 2 – Frequency distribution of height measurement – bin width: 60 cm

Alternatively, we can decide on a very narrow variation interval, let's say 5 cm (Fig. 3). In this case the frequency values are smaller, the greatest value being 4 occurrences in the interval between 165 cm and 170 cm. This representation is also mathematically correct, but due to the presence of too much detail, the final picture is not very suggestive as to the structure of the sample group. Usually a bin width that will produce between 10 and 15 bars on the histogram ensures a comprehensive representation of the data.

Bin	Frequency
130	3
135	1
140	2
145	1
150	3
155	2
160	2
165	2
170	4
175	2
180	3
185	2
190	2
195	1
200	0
205	2
210	1
More	0

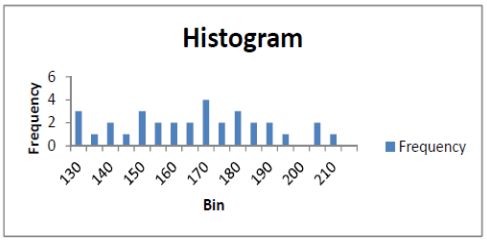


Figure. 3 – Frequency distribution of height measurement – bin width: 5 cm

In the above described example we have used the *absolute frequency* of the measured values by counting how many measurements fall into each variation interval. An alternate way of presenting the data is to calculate the relative frequencies for the occurrences by dividing the absolute frequency values to the total number of measurements (Table II).

Table II - Relative frequencies and percentages

Bin	*Frequency*	*Relative frequency = Frequency / N*	*% = Relative frequency(100)*
130	3	0,0909	9,09%
140	3	0,0909	9,09%
150	4	0,1212	12,12%
160	4	0,1212	12,12%
170	6	0,1819	18,19%
180	5	0,1515	15,15%
190	4	0,1212	12,12%
200	1	0,0303	3,03%
210	3	0,0909	9,09%
More	0	0	100,00%

The relative frequencies can be converted into percentages. These percentages offer a quick view of the distribution of the frequencies. It is easy to see how, in this case, the majority of the values are grouped around the center of the scale.

Chapter 15 – Frequency distribution types

According to the shape of the frequency distribution graph, the distributions can be classified into several types. In order to have more suggestive representations, especially when the measured variable is a continuous one, it is customary to represent the distribution graph not as a series of bars, but rather as a continuous curve. This smooth line depicts the relative changes that occur in the frequencies from one category or score to the next. An example of such a representation can be seen in Fig. 1.

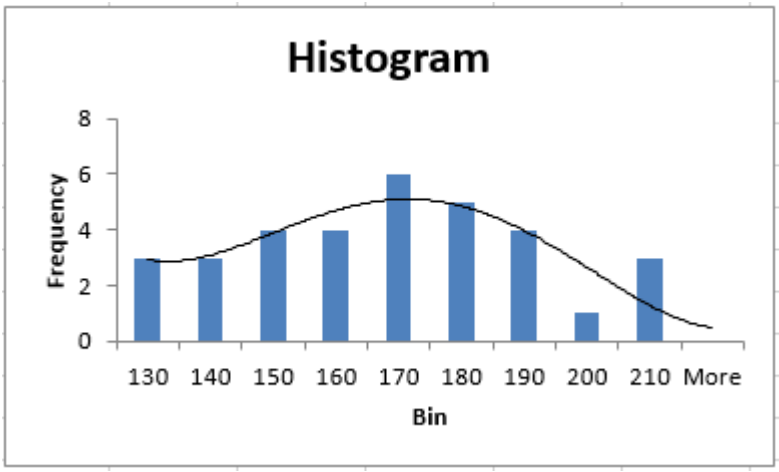


Fig. 1 – Histogram with a smooth trend-line

15.1 The normal (Gaussian) frequency distribution

The most common shape of such a curve is the *normal frequency* distribution, or *Gaussian frequency distribution*. This type of curve is characterized by symmetry, with lower frequencies at both ends, towards the extreme values of the measurement scale, and larger frequencies towards the center, as shown in Fig. 2.

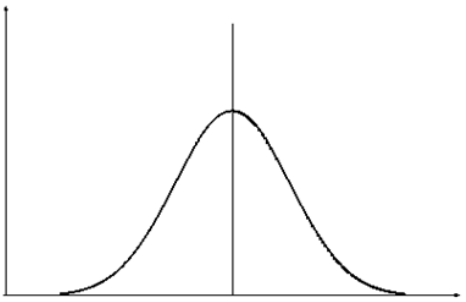


Figure. 2 – The normal distribution

The perfect normal distribution is described by an equation, but there can be a large number of variations of this bell-like shape that are still considered Gaussian. In order to precisely determine whether a set of measurements has a Gaussian

distribution of its frequencies statistical tests must be used. This issue will be discussed in a later chapter. For now, we will only look at the general shape of the curves.

15.1.1 Kurtosis

Some distributions gather the majority of the occurrences around a central point, others present a more wide spread of the values across the measurement scale. This characteristic is described by a statistical indicator named *kurtosis,* usually noted as "k".

A "pointy" distribution curve, which usually packs the majority of the measurements around a central tendency is called a *leptokurtic distribution* (k>0). A well balanced curve, which resembles closely the perfect bell-shape of the Gaussian distribution is called a *mesokurtic distribution* (k=0). Finally, a wide spread distribution, where the majority of the values are present across a longer range of the measuring scale, presents a flat curve and is called a *platykurtic distribution* (k<0). The three types of distributions discussed are presented in Fig. 3.

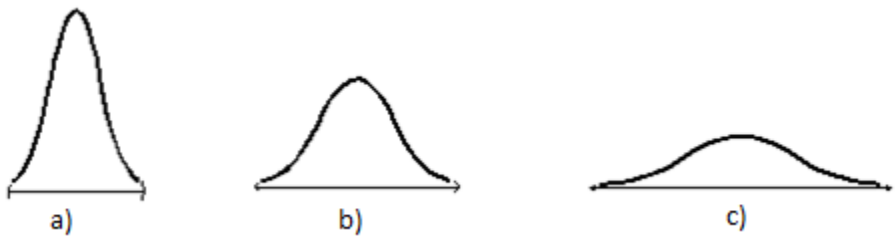


Figure. 3 – Leptokurtic (a), mesokurtic (b) and platykurtic (c) distribution curves

15.1.2 Skewness

The vast majority of the distribution curves based on real-life data are not perfectly balanced. Usually the curves lean towards one of the extremes of the scale in a certain degree. This characteristic of the distribution curves is measured by a statistical indicator called *skewness,* A positive skew means that the distribution curve presents a tail towards the higher values of the scale (the bell leans towards the lower values). A negative skew means that the curve presents an elongated tail towards the lower values of the measurement scale (the bell leans towards the higher values). Examples on positive and negative skew are resented in Fig. 4.

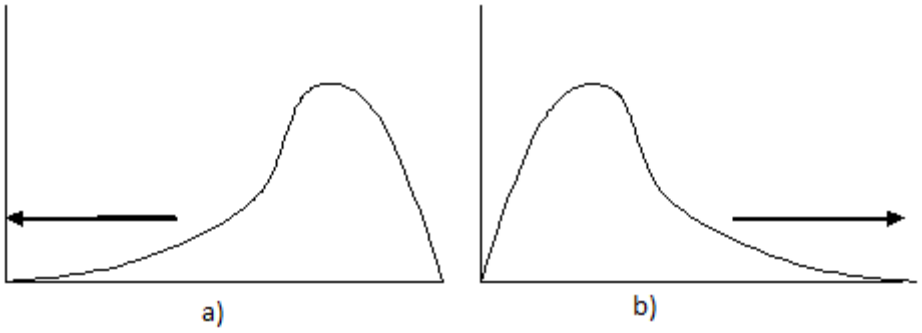


Figure. 4 – Negative skew (a) and positive skew (b)

15.2 Non-Gaussian frequency distributions

Usually, if we measure a natural variable on individuals or events randomly selected from a natural population the measurements will yield a Gaussian frequency distribution curve. However, Due to a number of reasons, which include sampling errors, measurement errors or simply the nature of the phenomenon under study, the frequency distribution can have other shapes. Some of these shapes are presented in Fig. 5.

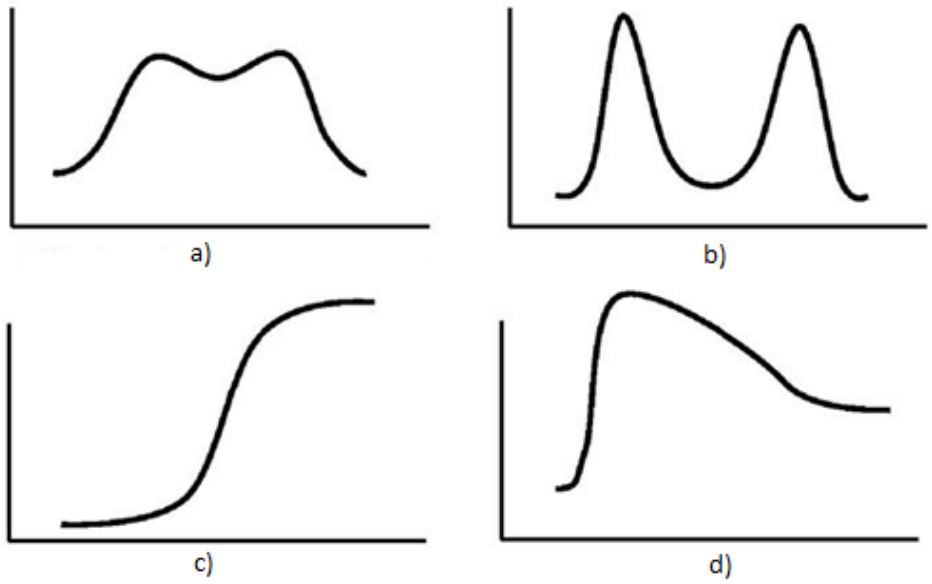


Figure. 5 – Non-Gaussian frequency distribution curves

For a thorough understanding of frequency distributions, let us imagine some situations which would yield such curves. Let's assume that we repeat the earlier described study regarding the height of the students, but in this instance we include in

our sample the women soccer team along with the male basketball team. What would the distribution frequency curve look like in this case? Well, if we have a group of people whose height is normally distributed around a low average mixed with a number of people who were selected into the sports team based on the fact that they are very tall, the distribution would look probably like the graph in section b) of Fig. 5. This type of curve is also called a *bimodal* frequency distribution.

Finally, let us imagine a situation for graph c) in Fig. 5. If the values measured would be exam grades, this would represent a very easy exam, where the majority of the students have high grades and very few fail the examination. If the examination would have been more difficult, the curve would have had a second tail towards the right as fewer students would have scored higher grades, thus resembling a normal frequency distribution curve.

Chapter 16 – Central tendency (mean, mode, median)

We have seen in the previous chapters that in order to describe a sample or a population according to a certain characteristic we can present the data as a frequency distribution and analyze its properties. This type of presentation is fairly elaborate though. If we wish to compare two or more samples or describe very succinctly a population according to a measured variable we need a simpler approach. The most common way to do this is by determining the central tendency of the measured values. This allows us to summarize and describe a distribution using a single value that is considered to be representative for the entire set of values.

In everyday language this type of synthetic description is known as the "average" value ("the average number of children in a family", "the average salary in an economy" etc.) or the most "popular" or "trendy" value ("the most popular restaurant in town", "the most trendy color of the year" etc.).

The central tendency of a set of data is defined as a statistical measure determining a single value which defines the center of the distribution. This central value is considered to be the most representative or most typical for the entire group.

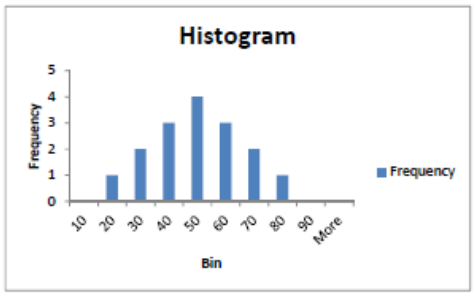
Unfortunately, there is no single formula to determine the central tendency. The way of determining it depends on the type of variable (qualitative or quantitative) and on the shape of the distribution itself. Let's look at the examples presented in Fig. 1. We can see three sets of data of equal length (n=16) and values ranging from 10 to 90. The frequency distributions however are very different. Let us try to determine the central tendency by looking at these distribution graphs.

In the case of dataset a) we have a tightly packed distribution around the value 50. This distribution is also symmetric. In this case it is very easy to determine that the central tendency of the group is 50.

a)

	Values:
1	20
2	30
3	30
4	40
5	40
6	40
7	50
8	50
9	50
10	50
11	60
12	60
13	60
14	70
15	70
16	80

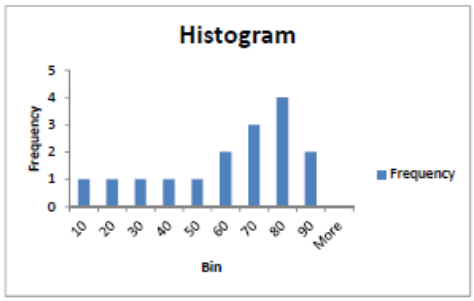
Bin	*Frequency*
10	0
20	1
30	2
40	3
50	4
60	3
70	2
80	1
90	0
More	0



b)

	Values:
1	10
2	20
3	30
4	40
5	50
6	60
7	60
8	70
9	70
10	70
11	80
12	80
13	80
14	80
15	90
16	90

Bin	*Frequency*
10	1
20	1
30	1
40	1
50	1
60	2
70	3
80	4
90	2
More	0



c)

	Values:
1	10
2	10
3	20
4	20
5	20
6	20
7	30
8	30
9	70
10	70
11	80
12	80
13	80
14	80
15	90
16	90

Bin	*Frequency*
10	2
20	4
30	2
40	0
50	0
60	0
70	2
80	4
90	2
More	0

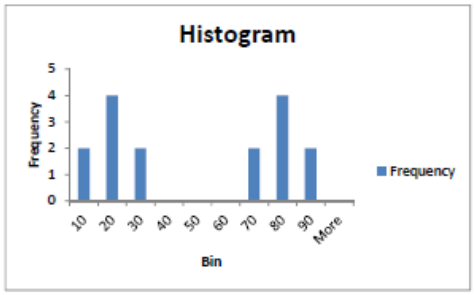


Figure1 – Different types of distributions having different central tendencies

Dataset b) presents a pronounced negative skew with the highest peak formed around the value 80. Which value would be a good descriptor of the central tendency for this distribution? Many people would be tempted to say 80, because more individuals have this vale than any other value and the peak in the distribution is very evident. However, if we take a closer look, we can see that 10 out of 16 values are under 80 – that's almost two thirds of the total number of measurements. Knowing this, it would probably be reasonable to determine the central tendency somewhere under the value 80.

Finally dataset c) offers a very interesting distribution shape. Like for dataset a) the distribution is symmetrical but instead of having one centered peak we have two distinct piles of frequencies, each with its own peak. If we consider the symmetry of the distribution, we can choose the value 50 as the central tendency as it is situated in the middle of the scale. But actually none of the subjects scored 50 in this dataset, so that would be a virtual center. On the other hand, if we consider each pile as a separate distribution, we might be tempted to say that this dataset has two central tendencies: 20 and 80. So, which center is the more representative in this case: the virtual center or the double center?

In order to properly determine the central tendency in different situations statisticians have developed three ways of measuring it: the mean, the median and the mode. Each of these indicators fit best for particular types of data and frequency distributions.

Chapter 17 – The mean

The mean is the most popular indicator of central tendency. It is computed by adding all the scores in a dataset together and dividing the sum by the total number of scores. This formula is also known as the *arithmetic average* of the scores. The mean is usually noted in statistical formulas with the letter M. The formula for the mean is:

$$M = \frac{\sum x}{n}$$

Where:

M – the mean

x – the measured values

n – the total number of values

An alternative notation for the mean is $\bar{x}$. This notation is used mainly in mathematical textbooks.

Thinking of the mean as the arithmetic average is a very accurate viewpoint when numbers are involved. But there is one other way of interpreting the mean, which offers a connection to the real world. After all, this is the world that statistics is trying to analyze.

The mean as a balance point

Let's take a look at Fig. 1. Here we have the frequency distribution of a simple set of measurements: 1, 2, 6, 6, 10. The mean of this dataset is calculated as M = 25/5 = 5. If we imagine that each occurrence of a score has the same weight as any other occurrence and that the frequency distribution sits on a board, the mean should be the position of the supporting point that balances out the board.

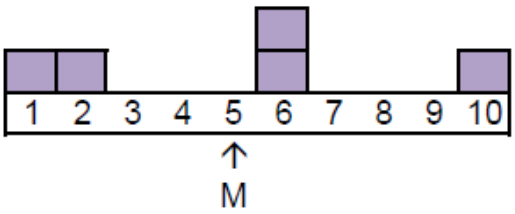


Figure 1 – The mean as a balance point

Notice that there are three boxes to the right of the mean and only two boxes to the left of the mean. So how can it be that the board stays in balance? This has to do with the distance of each box from the mean. Let's calculate those distances:

Score	Distance from the mean
1	4 points to the left of the mean
2	3 points to the left of the mean
6	1 point to the right of the mean
6	1 point to the right of the mean
10	5 points to the right of the mean

Now let us calculate the total distance of the boxes to the left and to the right of the mean:

Distance to the left = 4 + 3 = 7 points

Distance to the right = 1 + 1 + 5 = 7 points

The distances are equal. In other words, the mean balances out the distances of the scores to the central point. This is an important observation which will be useful in later chapters when dealing with variability. For now, let's just note that:

- The mean is a central tendency measure that takes into account all the values of the dataset
- The mean balances out distances

There are other characteristics of the mean that might be of importance when working with real-life data. Often a set of data undergoes some changes during processing, such as insertion of new values or deletion of existing ones or changing one value or all the values by applying a mathematical formula. What happens with the mean in these situations?

Inserting or deleting a score – this operation will usually change the mean of the distribution, with one notable exception: if the score that has been deleted or inserted is the same as the value of the mean, the mean will remain the same. The balance-board analogy is very suggestive in this case: imagine putting another weight on the board to the left or to the right of the mean, or lifting one of the existing weights. Alternatively imagine adding any number of weights on top of point 5 in Fig. 10.

Changing a score – this operation will always change the mean. Imagine shifting any weight to the left or to the right on the perfectly balanced board of Fig. 10. The balance will be lost and in order to restore it a new suspension point is necessary, hence a new mean will be calculated.

Changing all the scores by a formula – there are two usual operations that change all the scores of a dataset during statistical processing. One common operation is to add or subtract a constant from each score. In this situation the mean will change by the value of the constant. Another usual operation is to multiply or divide the values by

a constant. In this scenario the mean will change in the same way as the values, being multiplied or divided by the same constant. This technique is often used to change the measurement unit (for example from meters to centimeters by dividing by 100).

Chapter 18 – The median

Another way to measure the central tendency of a distribution is to determine its median point. *The median* of a distribution is represented by the point that divides the ordered set of scores in two equal parts. If we order the set of scores from smallest to largest, we can determine the median by counting the values, starting from the smallest, and stopping at the first point on the scale that is greater than 50% of the values measured.

In other words, the median is the midpoint of the list. The median will split the list of measurements into two sub-lists having an equal number of elements. Notice that the median does not take into account the relative distances between the elements of the list. In order to compute the median, we only need to order the scores and count them. Also note that the median is determined empirically, without the use of a mathematical formula.

There are two distinct cases for calculating the median: odd number of scores and even number of scores.

When *n* is an odd number, determining the median implies ordering the scores and selecting the middle score from the list. In this case the median will always be a value from the set of scores, as shown in the following example:

Scores: 1, 3, 4, 7, 17, 23, 24 – the median is 7

In the case of an even number of scores, the median will be a value between two adjacent scores from the ordered list. In order to determine the median, we have to order the list of values from smallest to largest, split in in two equal parts and determine the highest score from the first 50% of the list as well as the lowest score from the second 50% of the list. The median will be the average of these two scores, as shown in the following example:

Scores: 2, 4, 5, 8, 11, 15 – the median is AVERAGE(5, 8) = (5+8)/2 = 6,5

There are some similarities and some differences between the median and the mean. As in the case of the mean, the median takes into account all the scores from a distribution. However, the median does not take into account the actual values of the measurements, only the number of the scores. Let's examine the example presented in Fig. 1. For this distribution, the "balance point' of the board is situated under the value 4, which is the mean. However, the majority of the values (5 out of 6) are to the left of the mean. The median of the distribution is situated at the value 2,5. If we are to support the board with the weights at the point 2,5 the board will go out of balance because the median does not take into account the distances of the values to itself. Instead the median indicates the middle of the distribution as defined by the number of scores situated to the left and to the right of it.

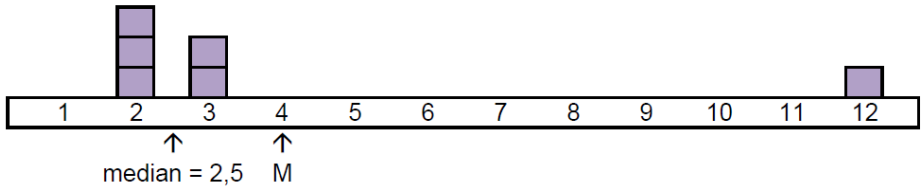


Figure 1 – The median and the mean as indicators of central tendency

As a conclusion, when relative distances among the scores are important, the mean is probably the most representative indicator for the center of the distribution. On the other hand, if the number of scores is of more importance, then the median becomes relevant as the measure of central tendency.

Chapter 19 – The mode

The third indicator of central tendency mentioned earlier is the mode. The word "mode" is also used as a synonym for "fashion" – something trendy or very popular. This meaning is not far from what "mode" means when used to analyze a frequency distribution. *The mode* of a distribution is the value that has the highest frequency in the set of scores. We could say that this represents the value that is the most "popular" or "trendy" in the group.

We might ask the question "why do we need a third way to measure central tendency"? Are there situations when the mode is a better descriptor of the center as the mean or the median? The answer is yes. For instance, the mode is the only relevant central tendency measure for nominal scales. Let's look at the following example:

Favorite color	*Frequency*
Red	35
Green	9
Blue	26
Black	14
Yellow	4
Grey	12

This is the frequency table for the answers of a number of one hundred students about their favorite car color. Notice that our measurement scale consists of eight different colors, so we have a nominal scale. There is no meaningful order or any way to measure distance between these elements, so computing a mean or a median is impossible. What would be the central tendency in this distribution? It is only reasonable to look at the most or "trendy" value, which is:

mode = Red

If we look at the shape of a frequency distribution, it is very easy to identify the mode. The mode is the value on the measuring scale that sits directly below the peak of the distribution.

Each distribution has one computed mean and one determined median. But can a distribution have more than one modes? Let's look back at the example about the height of the students, specifically at the distribution in Fig. 5 b, chapter 14) – the mixed group among the women soccer team and the men basketball team. What is the central tendency of such a distribution? If we compute the mean, it would probably be a value situated at the middle of the scale, but we will have registered very few scores close to that value. The median should be close to the mean judging by the shape of the

distribution. What about the mode? If incidentally the two peaks are exactly of the same height, will we have two modes? The answer is yes, and this is the reason why we called this type of distribution a *bimodal* one. If the two peaks are not exactly the same height but the difference is small, the distribution can still be described as bimodal, having in this case a *major mode* and a *minor mode*.

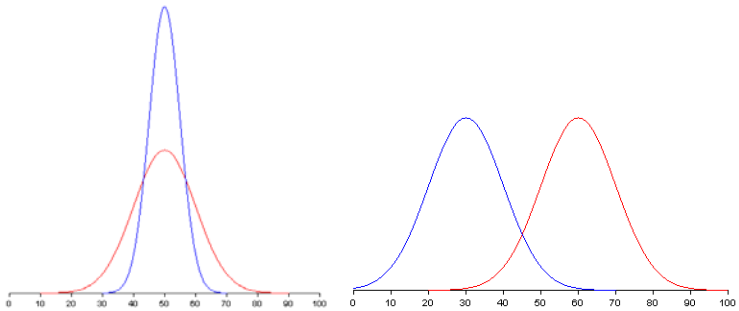
The mode also has another interesting characteristic: it corresponds by definition to one of the values in the list of scores. The mean and the median can incidentally have the same value as one of the scores, but often they fall between two adjacent scores. If we look back at Fig. 1, we can observe that for this distribution the mean equals 4 and the median equals 3,5, none of these values appearing in the actual list of scores. The mode however is equal to 2, the most common score in the list.

Chapter 20 – Variability

Variability in statistics has the same meaning as in everyday language. To say that scores or values are variable means that they differ from each other, they are not all the same. A low level of variability means small differences, a great level of variability means large differences between measured values. Thus, *variability* represents the quantitative measure of the differences between scores in a frequency distribution. This indicator provides information about the degree to which the values are packed together or spread across the measurement scale.

Variability is illustrated in Fig. 1. Here we can see two pairs of distribution curves. In Fig. 1a) the red and blue curves are centered on the same value of 50, thus having the same central tendency, the same mean in this case. But what about the variability? It is easy to notice that the scores yielding the blue distribution are more tightly packed around the center then the scores that produced the red one. It is safe to say that the red distribution has a greater variability than the blue one.

In Fig. 1b) the central tendencies of the two distributions are different. The blue one is centered on the value 30 whereas the red one has a higher mean equal to 60. Judging by the shape of the distributions, we could say though that their variability is similar. Indeed if we were to shift the curves so that they overlap, we would have a perfect fit.



a) b)

Figure 1 – Variability vs. central tendency

Notice that by providing information about the scores being spread or packed together, variability involves distances between scores or distances between scores and the central tendency. Low variability means that we can expect short distances between scores whereas great variability means larger distances between the measured values.

If we use the balanced board analogy (Fig. 2), we can see that both boards are in balance both having the same suspension point at the value 5, but the weights on board b are more evenly distributed from one end to another than the weights on board a). This means that distribution b) has higher variability than distribution a).

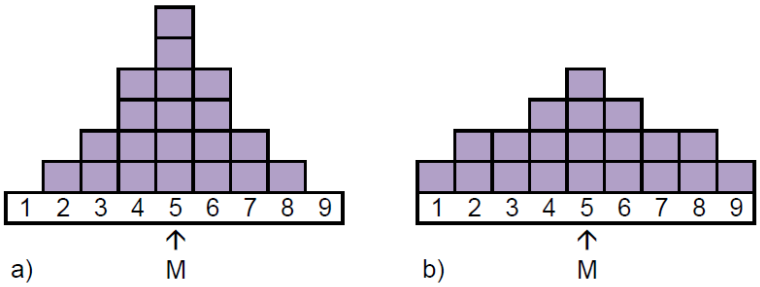


Figure2 – Different variability around the same mean

In order to measure variability statisticians have developed a number of indicators. We will present three of them: the range, the variance and the standard deviation.

20.1 Range

The range of a distribution is the distance between the smallest score and the highest score of the dataset. In Fig 2., for example, distribution a) has a range of 8-2 = 6 whereas distribution b) has a range of 9-1 = 8. In this case the range is a fairly good measure of variability and allows us to compare the two distributions.

The range has some limitations though. Let's consider the distribution in Fig 3.

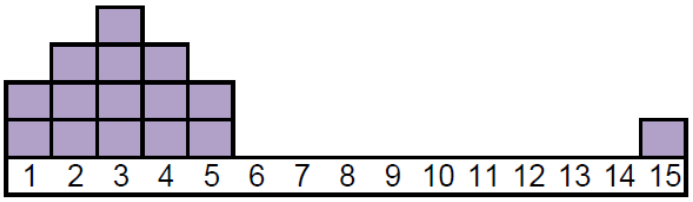


Figure3 – A distribution with one unusually large score

Here we have 14 out of 15 scores packed relatively tight around the value 3 and one unusually high value which is 15. By analyzing the shape of the distribution we can say that the majority of the values are not widely spread across the scale. However if we calculate the range, the result is 15 which suggests a high degree of variability.

Being determined only by two scores in the dataset, the smallest and the largest, this indicator completely ignores the values and the distribution of all the other scores. For this reason the range is considered to be a crude indicator for variability.

An accurate indicator of variability would be one that takes into account all the scores. Furthermore, the spreading of the values seems to be happening around the central tendency. The more packed the values are around the central tendency, the less variability there is in the distribution and vice-versa. So a good measure of variability

would have to somehow incorporate the distances of the values to the central tendency. Let's try to determine such a value in the next few pages.

20.2 Variance

Let' consider the set of values: 2, 3, 3, 5, 7, 7, 8. The central tendency for this dataset is the mean m = 35/7 = 7. Now let us compute the distance from this mean, also called *the deviation,* for each element:

x	x-mean(X)
2	-3
3	-2
3	-2
5	0
7	2
7	2
8	3

So, if we are trying to determine a measure of the variability of this group based on each element's distance to the mean, we could simply add all these distances. If we add the distances for this example, the result is zero, because the negative values counterbalance the positive ones. For any dataset we might imagine, some of the values will be situated below the mean (having negative distances to that mean) and some will be above the mean (having positive distances), so we have to come up with a way of compensating for this. The most usual way to do that in mathematical calculus is to square each value. This way we will compute the "*sum of squares"* for this distribution:

Sum of squares = $(-3)^2 + (-2)^2 +(-2)^2 +0^2 +2^2 +2^2 +3^2 = 9 + 4 + 4 + 0 + 4 + 4 + 9 = 34$

This value can be an indicator of the variability of our sample. But could we use this as a universal indicator and compare different samples in terms of variability?

Let' analyze the example in Table I. Here we present two variables, x and y. Variable x has a number of N(x) = 9 recorded values, whereas variable y has a number of N(y) = 18 scores. Notice that the scores for y are exactly the same as the scores for x, only each score has double the number of appearances for y compared to x. If we analyze the distributions of the two sets of data, presented in Fig. 4, we can see that the central tendency, described by the mean, is identical: m(x) = m(y) = 4. Furthermore, based on the shape of the two distributions, we could speculate about the variability. Most people would agree that these two sets of scores should probably have the same

variability. Let us test the sum of squares against this hypothesis. As shown in the table, the sum of squares for x is 12 whereas the sum of squares for y is 24. This would suggest that dataset x is twice as compact around the mean as dataset y, which is hard to believe when looking at the shape of the distributions.

Table I – Two datasets having equal means and equal variability

x	x-m	$(x-m(x))^2$
2	-2	4
3	-1	1
3	-1	1
4	0	0
4	0	0
4	0	0
5	1	1
5	1	1
6	2	4
		12

y	y-m(y)	$(y-m(y))^2$
2	-2,00	4,00
2	-2,00	4,00
3	-1,00	1,00
3	-1,00	1,00
3	-1,00	1,00
3	-1,00	1,00
4	0,00	0,00
4	0,00	0,00
4	0,00	0,00
4	0,00	0,00
4	0,00	0,00
4	0,00	0,00
5	1,00	1,00
5	1,00	1,00
5	1,00	1,00
5	1,00	1,00
6	2,00	4,00
6	2,00	4,00
		24,00

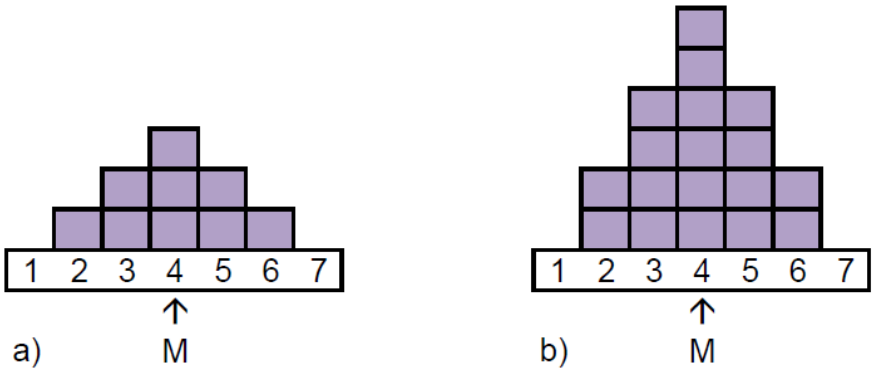


Figure4 – Distribution shapes having equal means and equal variability

Notice that if we simply add all the squared distances to the mean, the sum will be dependent on the size of the group. What we really want in this case, is a value that describes each group, without being influenced of how large or small the sample is. For this we can use the average squared distance from the mean, or the mean of these squared distances. This is also known as variance. *The variance* of a sample equals to the mean squared deviation, representing the average squared distance from the mean.

For the example presented above:

Variance(x) = 12/9 = 1,33

Variance (y) = 24/18 = 1,33

This confirms our initial impression based on the shape of the distributions that the two samples have equal variability.

20.3 Standard deviation

In order to calculate the variance of a set of scores, we had to square score's distance from the mean. This can affect the meaning of the data. For example is not very useful to know the squared value of a patient's blood pressure or the squared value of a hemoglobin count. If we were presented with such data, the next step in processing it would probably be an operation that eliminates the square and restores meaning to the scores. This can be done by extracting the square root from the provided values.

Similarly, in order to have a meaningful measure of variability, the last step in calculating this indicator is to extract the square root from the variance. The square root of the variance is called *standard deviation*:

$$Standard\ deviation = \sigma = \sqrt{Variance} = \sqrt{\frac{Sum\ of\ squares}{N}} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}}$$

The standard deviation is the most relevant statistical indicator to describe variability in a distribution.

20.4 Coefficient of variation

The standard deviation allows us to compare different distributions in terms of variability. What if we only work with one distribution? How can we interpret the absolute value of the standard deviation? One way to do this is to compare it to the mean. Remember that the standard deviation incorporates all the measured values and the value of the mean. For this we can calculate the *coefficient of variation*, defined as the ratio of the standard deviation to the mean:

$$CV = \frac{Standard\ deviation}{mean} = \frac{\sigma}{\bar{x}}$$

The coefficient of variation has no measurement unit and is often expressed as percentage. This allows us to compare the variability of sets of measurements taken in different units. For instance the variability of height measurements vs. the variability of weight measurements for a sample of the population.

Chapter 21 – Inferential statistics. Definition and terminology

Inference refers to drawing conclusions from data. Statistical inference can be defined as the drawing of these conclusions from different types of data (quantitative and qualitative) using statistical tests to describe and to test hypothesis. Using inferential statistics, a researcher can extend the conclusions drawn from the results obtained by studying a sample, to the entire population.

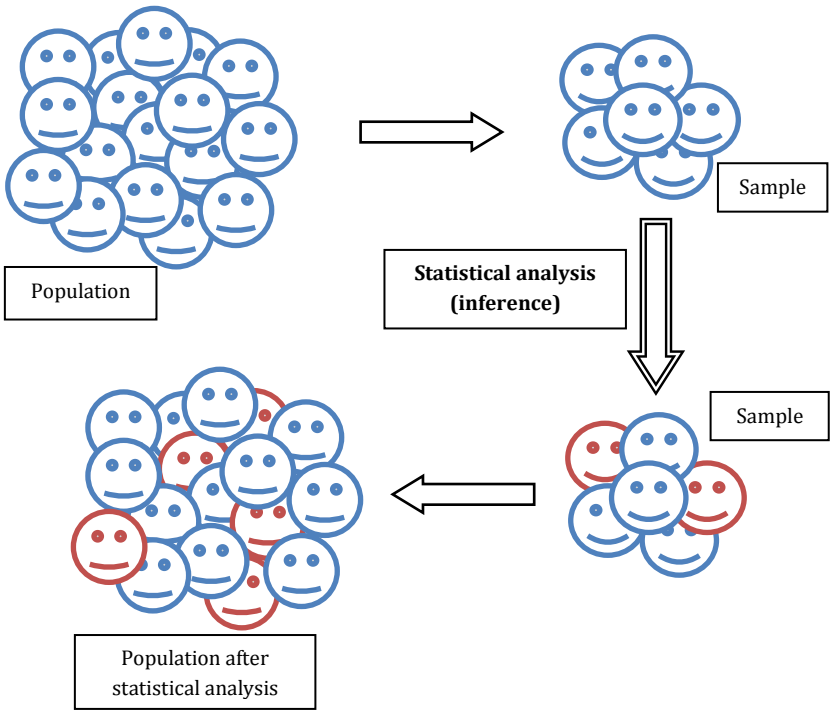


Figure 1 – Inferential statistics in biomedical research

Inferential statistics may include:

- point estimation: uses the sample data to calculate a single value, which approximates best an unknown population parameter (e.g. Relative Risk, RR=4.23);
- interval estimation: uses the sample data to calculate an interval of possible values for an unknown population parameter (e.g. CI 95% (1.5 - 6.2));

Sometime it is possible to use both, point and interval estimation, in order to make inferences about a parameter of the population through a sample extracted from it.

If we define the "true value" as the actual population value that would be obtained with perfect measuring instruments and without committing error of any type, we will have to accept that we may never know the true value of a parameter of the population. But, using the combination of these two estimators, we may obtain a certain level of confidence that the true value may be in that interval, even if our result (point estimation) is not necessarily identical with the true value.

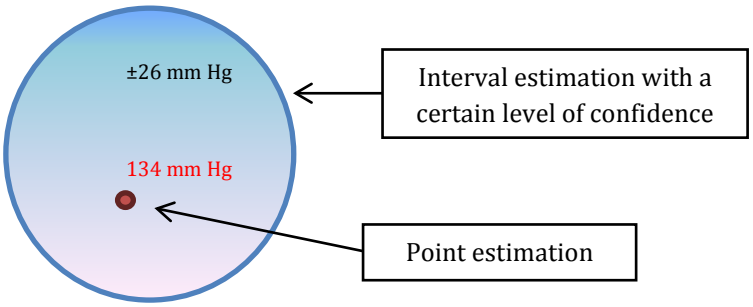


Figure 2 – The relationship between point and interval estimation

- prediction / forecast: a prediction is a statement that a particular event will occur in the future, in more certain terms than a forecast. Forecasting is the process of estimation in unknown situations.
- statistical hypothesis test: a method of making statistical decisions using experimental data. These decisions are almost always made using null-hypothesis tests. The null hypothesis (H_0) formally describes some aspect of the statistical behavior of a set of data; this description is treated as valid unless the actual behavior of the data contradicts this assumption. The null hypothesis is in opposition to another hypothesis, called „alternative hypothesis".

Chapter 22 – Steps of applying a statistical test

In order to properly apply a statistical test, certain steps must be followed. Ignoring these steps may result in distorted results

1. stating the relevant null and alternative hypothesis to be tested
2. choosing the significance level, represented by the greek symbol α. Popular levels of significance are 5%, 1% and 0.1% (0.05, 0.01 and 0.001).
3. computing the relevant test statistic
4. comparing the test statistic to the relevant critical values (obtained from tables in standard cases). The p value is obtained.
5. deciding to either fail to reject the null hypothesis or reject it in favor of the alternative hypothesis. The decision to reject the null hypothesis is based on the p value being smaller or equal to α ($p \leq \alpha$).

Steps 3 and 4 are usually performed by statistical analysis software.

22.1 Stating the relevant null and alternative hypothesis to be tested

A hypothesis is a statement of belief about the values of population parameters. In hypothesis testing, we usually consider two hypotheses: the null and alternative hypotheses.

The null hypothesis, denoted by H_0, is usually a hypothesis of *no difference*. It will state that there is no difference between the population parameter and its hypothesized value or set of values. The hypothesized values chosen for the null hypothesis are usually chosen to be uninteresting values. An example might be that in a trial, comparing two diabetes drugs, the mean values for fasting plasma glucose are the same for the two treatment groups.

In general, the researcher is interested in rejecting the null hypothesis. The alternative hypothesis, denoted by H_1, is a claim that the null hypothesis is false; e.g. the population parameter takes on a value different from the value or values specified by the null hypothesis. The alternative hypothesis is usually the scientifically interesting hypothesis that we would like to confirm.

22.2 Choosing the significance level

Before deciding on a significance level, we must familiarize ourselves with the types of errors a researcher can make.

Type I error

It is also known as an „error of first kind", a „α error" or a „false positive". This error consists in rejecting a null hypothesis when it is actually true. It occurs when we observe a difference when actually there is none. Type I error can be viewed as the error of „excessive credulity".

Type II error

It is also known as an „error of the second kind", a „β error" or a „false negative". A type II error consists in failing to reject a null hypothesis when it is in fact not true. It occurs when we fail to observe a difference when in truth there is one. Type II error can be viewed as the error of „excessive skepticism".

Table I: Outcomes of statistical hypothesis testing

	Real status of H_0	
Decision made	H_0 is true	H_0 is false
H_0 rejected	Type I error (false positive)	Correct decision
H_0 not rejected	Correct decision	Type II error (false negative)

Significance level (α) is defined as the probability of rejecting the null hypothesis when it is actually true. Popular levels of significance are 5%, 1% and 0.1%, corresponding to a confidence level of 95%, 99% and 99.9%.

22.3 Computing the relevant test statistic

The test statistic represents a numerical value obtained from the data set or sets. The formulas used in obtaining this value are specific to each statistical test. Based on this test statistic, a researcher can calculate the p value.

22.4 Comparing the test statistic to the relevant critical values

After the calculation of the test statistic, the researcher must compare this value to the critical values. This comparison is made using critical values tables. It is important that this comparison is made using the relevant critical value, corresponding to the number of degrees of freedom and the significance level chosen.

The number of degrees of freedom varies for each statistical test. It depends on the number of independent pieces of data used in calculating a parameter, e.g., for the t-student test the number of degrees of freedom is calculated by the formula N_1+N_2-2, where N_1 and N_2 are the number of records in each sample and „2" represents the number of samples.

22.5 Deciding to either reject or not the null hypothesis

After the comparison of the test statistic to the relevant critical value, decision of either rejection or acceptance of the null hypothesis can be made. If the test statistic is greater than the critical value, the null hypothesis can be rejected. If the test statistic is smaller than the critical value, one cannot safely reject the null hypothesis.

Usually, the statistical analysis software makes the calculation of the test statistic and the comparison to the critical value. More than this, the software can also estimate the p value, making it much easier for the researcher to make a decision. Thus, if $p \leq \alpha$, null hypothesis can be rejected. The p value is a number between 0 and 1. It represents the statistical significance of the test.

Chapter 23 – Outlier tests

Outlier tests can determine if an experiment result or a parameter measurement is due to the normal biological variability and would not take the experiment result away from the „true value" or is an outlier that appeared as an accident during the experiment. Outliers are extreme values, either very small or very large, that alter the sample mean significantly.

Often the experimental results contain some records that are very different compared to the others. An issue arises: should these outlying values be eliminated considering that those values deviate the sample mean significant. There is a tendency amongst researchers to dismiss those values without any further consideration. Doing such values elimination is wrong and the results could be misleading.

The outlier tests are optional. It is recommended to apply an outlier test when the data shows a high dispersion. Such high dispersion is assessed evaluating the descriptive statistics parameters (mean, standard deviation, sample variance). If sample variance or standard deviation is much higher than the mean, a researcher might consider applying an outlier test.

Since outliers elimination still represents a debate subject for statisticians, a researcher faced with a data sample that contains outliers must decide on keeping the values or eliminating them. This decision should be based on the researcher's knowledge of the experiment methodology. Thus, if the methods used are not error prone and there is no data collection error, the so-called outlier should not be eliminated.

The outlier test checks the furthest away value from the mean. Based on this result a researcher can safely eliminate that value. Each extreme value is checked one at a time.

One of the most commonly used outlier test is Grubbs test. The steps for applying Grubbs test are:

- Stating the hypothesis:
    - H_0: The checked value IS NOT significantly different from the sample mean.
    - H_1: The value IS significantly different from the sample mean.
- Choosing the significance level: $\alpha=0.05$
- Compute the relevant test statistic:

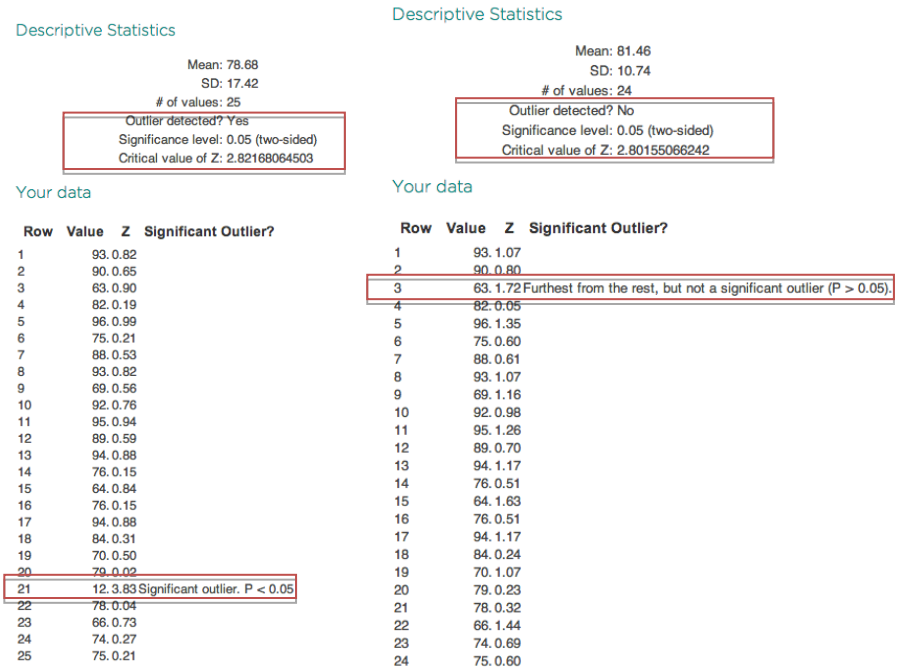
$$Z = \frac{|\bar{X} - X_i|}{SD}$$

- $\bar{X}$is the sample mean, X_i represents the value being checked and SD stands for standard deviation.

- The test statistic Z is compared to the critical value (CV). If Z>CV, H_0 is rejected and the value is considered an outlier.

Statistical analysis software returns the p value; if p≤α then H_0 is rejected and the checked value is considered an outlier and it can safely be eliminated from the data set.

If the analysis is performed on dependent samples, it is mandatory to eliminate outlier's pair also. The "paired data" characteristic must be preserved.The entire procedure is applied for each extreme value until no more outliers are detected.



Descriptive Statistics

Mean: 78.68
SD: 17.42
# of values: 25
Outlier detected? Yes
Significance level: 0.05 (two-sided)
Critical value of Z: 2.82168064503

Your data

Row	Value	Z	Significant Outlier?
1	93.	0.82	
2	90.	0.65	
3	63.	0.90	
4	82.	0.19	
5	96.	0.99	
6	75.	0.21	
7	88.	0.53	
8	93.	0.82	
9	69.	0.56	
10	92.	0.76	
11	95.	0.94	
12	89.	0.59	
13	94.	0.88	
14	76.	0.15	
15	64.	0.84	
16	76.	0.15	
17	94.	0.88	
18	84.	0.31	
19	70.	0.50	
20	79.	0.02	
21	12.	3.83	Significant outlier. P < 0.05
22	78.	0.04	
23	66.	0.73	
24	74.	0.27	
25	75.	0.21	

Descriptive Statistics

Mean: 81.46
SD: 10.74
# of values: 24
Outlier detected? No
Significance level: 0.05 (two-sided)
Critical value of Z: 2.80155066242

Your data

Row	Value	Z	Significant Outlier?
1	93.	1.07	
2	90.	0.80	
3	63.	1.72	Furthest from the rest, but not a significant outlier (P > 0.05).
4	82.	0.05	
5	96.	1.35	
6	75.	0.60	
7	88.	0.61	
8	93.	1.07	
9	69.	1.16	
10	92.	0.98	
11	95.	1.26	
12	89.	0.70	
13	94.	1.17	
14	76.	0.51	
15	64.	1.63	
16	76.	0.51	
17	94.	1.17	
18	84.	0.24	
19	70.	1.07	
20	79.	0.23	
21	78.	0.32	
22	66.	1.44	
23	74.	0.69	
24	75.	0.60	

Figure 1 – Output example of an outlier test

Chapter 24 – Goodness-of-fit tests / normality tests

Goodness-of-fit tests are used to determine if the data sample we analyze follows a certain distribution type. Normality tests are a subset of goodness-of-fit tests and are used in determining if the distribution curve of the data set we are analyzing can be safely approximated as a normal (Gaussian) one. In statistical hypothesis testing, they will test the data against the null hypothesis that it is normally distributed.

Knowing that the distribution curve can be approximated as a normal one is extremely important since many statistical tests (e.g. t-tests, ANOVA and its variants), assume that we have sampled data from populations that follow a Gaussian (normal/bell-shaped) distribution. Tests that follow this assumption are called parametric tests and the branch of statistical science that uses such tests is called parametric statistics.

Parametric statistics assume that data come from a type of probability distribution (e.g. normal distribution) and make inferences about the parameters of the distribution. However, many populations from which data are measured - and biological data are often in this category - never follow a Gaussian distribution precisely. A Gaussian distribution extends infinitely in both directions and so includes both infinitely low negative numbers and infinitely high positive numbers and biological data are often naturally limited in range. Still, many kinds of biological data do follow a bell-shaped distribution curve that is approximately Gaussian.

Another branch of statistics, called nonparametric statistics, propose distribution-free methods and tests, which do not rely on assumptions that the data are drawn from a given probability distribution (in our case, the normal distribution). Such tests are named nonparametric statistical tests. We should be aware that almost every parametric statistical test has a correspondent nonparametric test.

Considering the importance of knowing the distribution type, especially if the data sample follows a normal distribution curve, the normality tests are mandatory in every statistical analysis protocol.

The most commonly used normality tests are:

Kolmogorov-Smirnov test compares the cumulative distribution of the data with the expected cumulative normal distribution and at the base at its p-

value stands the largest discrepancy. Its use is discouraged, due to the lack of sensibility. (Figure 1).

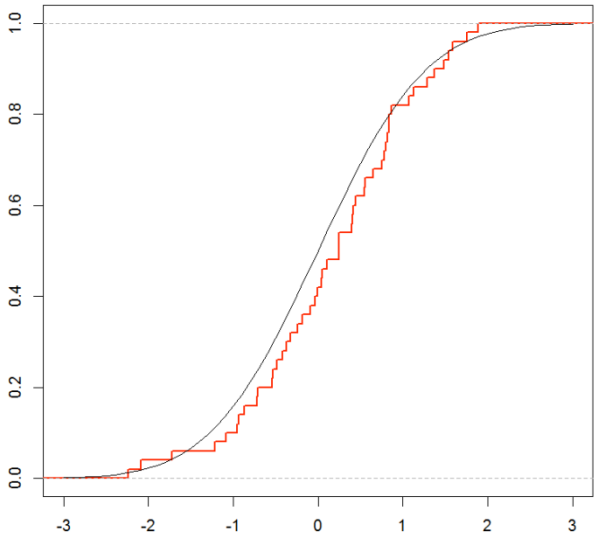


Figure 1: Cumulative distribution function for a normal distribution (black) and a 50 random numbers data set (red)

D'Agostino-Pearson normality test – which computes the skewness and kurtosis to quantify how far is the distribution from normality, in terms of asymmetry and shape. Then it calculates how far each of these values differs from the value expected with a normal distribution, and computes a single P-value from all these parameters. It is a versatile and powerful (compared to some others) normality test and is recommended by some modern statistical books.

There are available a relatively large number of other normality tests: Jarque-Bera test, Anderson-Darling test, Cramér-von-Mises criterion, Lilliefors test for normality (an adaptation of the Kolmogorov-Smirnov test), Shapiro-Wilk test, the Shapiro–Francia test for normality.

Steps needed to perform in order to apply a normality test are the same as any statistical test:

- hypothesis statement: H_0 – these is no difference between a normal distribution and the one to analyze; H_1 – there is a statistical significant difference between a normal distribution curve and the one to analyze;
- choosing the significance level: α=0.05;
- test statistic calculation – specific for each test;
- test statistic comparison to the critical values, according to the significance level – if test statistic is greater than the critical value, H_0 is rejected and the analyzed data sample cannot be approximated to a normal distribution.

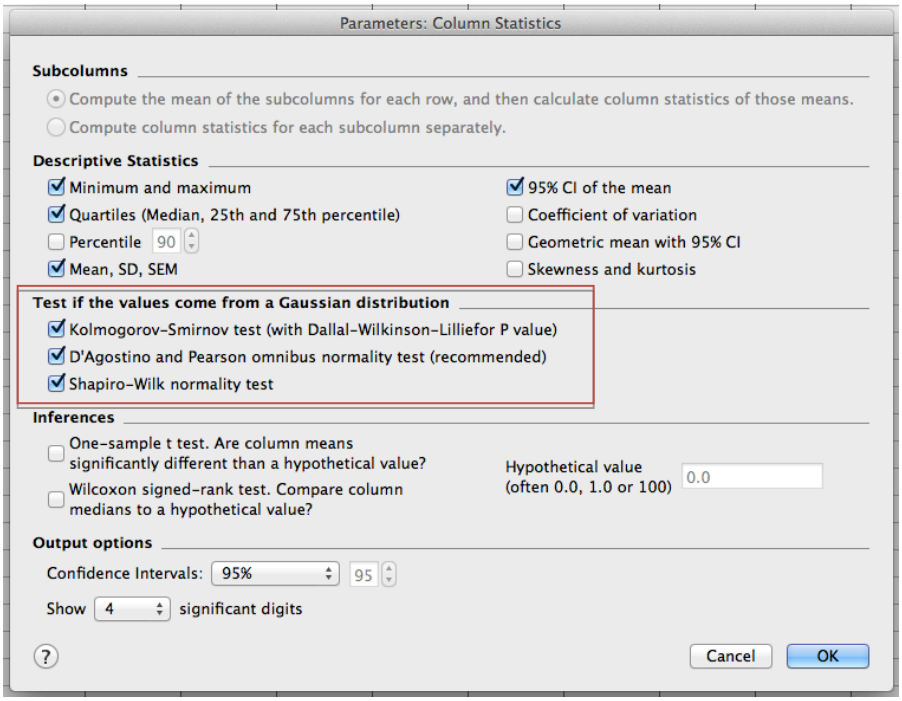


Figure 2: Normality test selection in Graph Prism 5 Trial

Col Stats		A
		Data
1	Number of values	200
2		
3	Minimum	61.80
4	25% Percentile	90.31
5	Median	99.93
6	75% Percentile	109.8
7	Maximum	138.0
8		
9	Mean	100.1
10	Std. Deviation	14.82
11	Std. Error	1.048
12		
13	Lower 95% CI of mean	98.03
14	Upper 95% CI of mean	102.2
15		

16	KS normality test	
17	KS distance	0.04414
18	P value	> 0.10
19	Passed normality test (alpha=0.05)?	Yes
20	P value summary	ns
21		
22	D'Agostino & Pearson omnibus normality test	
23	K2	1.090
24	P value	0.5798
25	Passed normality test (alpha=0.05)?	Yes
26	P value summary	ns
27		
28	Shapiro-Wilk normality test	
29	W	0.9940
30	P value	0.5950
31	Passed normality test (alpha=0.05)?	Yes
32	P value summary	ns

Figure 3: Results display of normality tests in Graph Prism 5 Trial

Statistical analysis software usually returns the p value and give the researcher the conclusion: the analyzed sample has a normal distribution or not, making it easier than test statistic calculation.

Chapter 25 – Significance tests

Usually, medical studies are based on performing biological parameters measurements on different samples and interpreting the differences amongst them. The researcher can conclude either that the differences are due to chance and in fact there is no difference between the samples studied or that the samples belong to different statistical populations and the difference between them is real.

A researcher can confront with two situations: the distribution curves of the samples to analyze are either Gaussian (normal) or not. The decision to apply either a parametric test (used for Gaussian distributions) or a nonparametric test (test for which the distribution types are not that important) must be made.

One might be tempted to apply a nonparametric test, since it does not rely on the assumption of data normality. To understand the difference between parametric and nonparametric tests, two more basic concepts in statistics are needed: robustness and power of a statistical test.

A robust statistical test is one that performs well enough even if its assumptions are somewhat violated. In this respect, nonparametric tests tend to be more robust than their parametric equivalents, for example by being able to deal with very small samples, where data are far to be normally distributed.

The power of a statistical test is the probability that the test will reject the null hypothesis when the alternative hypothesis is true (e.g. that it will not make a type II error). As power increases, the chances of a Type II error decrease. Nonparametric tests tend to be more robust, but usually they have less power. In other words, a larger sample size can be required to draw conclusions with the same degree of confidence.

A nonparametric test should be used in either of these cases:

- The outcome variable is a rank or score with fewer than a dozen or so categories (e.g. Apgar score). Clearly the population cannot be Gaussian in these cases.
- The same problem may appear when the sample size is too small less than 10 records).
- When a few values are off scale, too high or too low to measure with a specific measurement technique. Even if the population is normally

distributed, it is impossible to analyze the sample data with a parametric test (e.g. t-test or ANOVA). Using a nonparametric test with these kinds of data is easy because it will not rely on assumptions that the data are drawn from a normal distribution. Nonparametric tests work by recoding the original data into ranks. Extreme low and extreme high values are assigned a rank value and thus will not distort the analysis as would use of the original data containing extreme values. It won't matter that a few values could not be precisely measured.

- When we have enough "statistical confidence" that the population is far from normally distributed. A normality test should be applied.

Using normality tests seems to be an easy way to decide if we will have to use a parametric or a non-parametric statistical test. But it is not, because we should pay attention to the size of the sample(s) before using such tests. For small samples normality tests are not very useful. They have little power to discriminate between Gaussian and non-Gaussian populations. Small samples simply do not contain enough information to let us make inferences about the shape of the distribution of the entire population.

Another important aspect to mention is the paired or unpaired data characteristic. This derives from the methodology used in the study or experiment. The samples can be dependent or independent.

Paired data are recognized in the following study designs:

- The researcher measures a variable before and after an intervention or procedure on the same subject.
- The same laboratory experiment is performed several times; each time the researcher makes the determinations to the studied material and its control.
- Study subjects were recruited as pairs, based on a different criterion such as age or sex.
- The study involves some determinations for twins or pairs parent-child (genetic studies).

The list is not limited to these cases. The samples are dependent (paired data) when a value from one sample is correlated or is in some kind of connection to a specific value from another sample.

Usually, a significance test has two types; one can apply a one-tailed test or a two-tailed test. It is important to know the difference between them. A one-tailed test looks only for an increase or a decrease (a one-way change) in

the parameter whereas a two-tailed test looks for any change in the parameter (which can be any change - increase or decrease).

When comparing two groups, one must distinguish between one- and two-tail P-values. The two-tail P-value answers this question: Assuming the null hypothesis is true, what is the chance that randomly selected samples would have means as far apart (or further) as we observed in this experiment with either group having the larger mean?

To interpret a one-tail P-value, we must predict which group will have the larger mean before collecting any data. The one-tail P-value answers this question: Assuming the null hypothesis is true, what is the chance that randomly selected samples would have means as far apart (or further) as observed in this experiment with the specified group having the larger mean?

A one-tail P-value is appropriate only when previous data, physical limitations or common sense tell us that a difference, if any, can only go in one direction. Or alternatively, we may be interested in a result only in one direction. For example, if a new drug has been developed to treat a condition for which an older drug exists. Clearly, researchers are only interested in continuing research on the new drug if it performs better than the old drug. The null hypothesis will be accepted if the new drug performs the same or worse than the older drug.

So, the real issue here is whether we have sufficient knowledge of the experimental situation to know that differences can occur in only one direction, or we are interested only in group differences in both directions.

For all these reasons, especially for beginners, choosing the right two-tailed test instead of a one-tailed test is recommended, unless there is a good reason to pick a one-tailed P-value.

Chapter 26 – Fischer-Snedecor "F" test

The F-test is a parametric significance test used in comparing two samples variances. Since it's a parametric test, it can only be applied to normal distributions. Both samples must be tested for normality previous to using the F test and both samples must have a normal distribution.

To apply an F test the same steps are necessary, like any other statistical test:

- stating the hypothesis: H0 – "there is no difference between the variances of the two samples";
- choosing the significance level: α=0.05;
- calculating the test statistic: the ratio between the variances of the two samples. Let X_1, …,X_n and Y_1, …, Y_m be the samples from normal distributed populations. Samples means are calculated as

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{and} \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$$

for each sample. Let the variances be

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \text{and} S_Y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y})^2$$

for each sample. The test statistic is $F = \frac{S_X^2}{S_Y^2}$.

- comparing the test statistic to the critical value;
- rejecting the null hypothesis if the test statistic is greater than the critical value.

The F test is very sensitive to the assumption of normality. It should never be used on samples that do not follow a normal distribution. In practice, a researcher would use statistical analysis software that returns the calculated p value for the test. If p value is smaller than or equal to α (0.05), the null hypothesis is rejected in favor of the alternative hypothesis, concluding the inequality of the samples variances.

F test to compare variances	
F,DFn, Dfd	1.680, 5, 4
P value	0.6354
P value summary	ns
Are variances significantly different?	No

Figure 1 – F test results display in Graph Prism 5 Trial

Chapter 27 – Student's t-test

T-test was introduced by William Gosset to compare population means using small samples. He published his work under the pseudonym "Student" (hence the test name).

Student's t-test is a parametric test. It can only be applied to samples from normal distributed populations. T-test can be used on either one sample (one sample t-test) or two samples (two samples t-test).

There are 3 types of two samples t-test:

- t-test for paired data (paired t-test);
- t-test for unpaired data with equal variances (homoscedatic, Welch's corrected unpaired t-test)
- t-test for unpaired data with different variances (heteroscedatic, independent samples t-test)

One sample t-test compares a sample mean to a hypothetical value. The formula for the test statistic is:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where $\bar{x}$ represents the sample mean, μ_0 is the hypothetical value the sample mean is compared to, s is the sample standard deviation and n stands for the sample size.

Based on the test statistic, a researcher can either reject or accept the null hypothesis that states "there is no difference between the sample mean and the hypothetical value".

Two samples t-test is used in comparing the means of two samples taken from normal distributed populations. The null hypothesis states, "There is no difference between the means of the studied samples". In order to use the correct formula and to apply the correct type of t-test, the researcher must acknowledge the paired or unpaired character of the data (dependent or independent samples) and must test for samples variances equality (F test) if the data is unpaired.

The formulas used in the calculation of the test statistic are:

$$t = \frac{\overline{x_1} - \overline{x_2}}{S_d} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

where $\overline{x_1}$, $\overline{x_2}$ represent the samples mean, n_1, n_2 represent the sample sizes and S_d stands for standard error of difference between means.

$$S_d = \sqrt{\frac{\sum_{i=1}^{n_1}(x_i - \overline{x_1})^2 + \sum_{j=1}^{n_2}\left(x_j - \overline{x_2}\right)^2}{n_1 + n_2 - 2}}$$

Replacing the formula for standard error of difference between means, test statistic formula becomes:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sum_{i=1}^{n_1}(x_i - \overline{x_1})^2 + \sum_{j=1}^{n_2}\left(x_j - \overline{x_2}\right)^2}{n_1 + n_2 - 2}}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

It is unlikely that a researcher would calculate test statistic manually, using the formula. Usually, computer software (statistical analysis software) is used for these calculations. Amongst the results, the software returns the test statistic, the p value and other parameters needed. Based on the p value (if p value is smaller than chosen α), the null hypothesis is rejected and the conclusion would be that there is a statistical significant difference between the means of the studied samples.

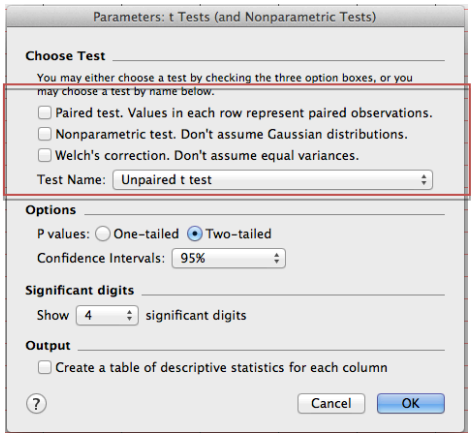


Figure 1: Choosing the appropriate mean comparison test in Graph Prism 5

Table Analyzed	Unpaired t test data
Column A	Male
vs	vs
Column B	Female
Unpaired t test	
P value	0.2613
P value summary	ns
Are means signif. different? (P < 0.05)	No
One- or two-tailed P value?	Two-tailed
t, df	t=1.199 df=9
How big is the difference?	
Mean ± SEM of column A	44.20 ± 5.669 N=5
Mean ± SEM of column B	55.00 ± 6.708 N=6
Difference between means	-10.80 ± 9.010
95% confidence interval	-31.18 to 9.582
R squared	0.1377
F test to compare variances	
F,DFn, Dfd	1.680, 5, 4
P value	0.6354
P value summary	ns
Are variances significantly different?	No

Figure 2: T-test analysis results display in Graph Prism 5; F test results simultaneously displayed

Chapter 28 – Wilcoxon test

Wilcoxon test is a nonparametric statistical hypothesis test. It is used for paired data (dependent samples), when one or both samples failed normality test (the distributions cannot be approximated as Gaussian). It is usually applied as an alternative to the paired t-test.

Since the samples distributions are different than normal, it is improper to formulate the hypothesis as a test for *mean* comparison. Instead, the null hypothesis to test is "There is no difference between the *medians* of the two samples"

The steps needed to go through, in order to apply a Wilcoxon test, are the same as any significance test: formulate the hypothesis, choose the significance level, calculate test statistic, reject or accept the null hypothesis.

Calculation of the test statistic:

For each pair, needs to be calculated the absolute difference between the values and a sign function for it. The pairs that have the difference between the values equal to 0 are eliminated.

The remaining pairs are ordered from the smallest difference between the values to the largest. Every pair receives a rank starting from 1 (the pair with the smallest difference).

The formula for the test statistic is:

$$W = \left| \sum_{i=1}^{n} \left[sgn\left(x_{2,i} - x_{1,i}\right) \cdot R_i \right] \right|$$

where $x_{1,i}$ and $x_{2,i}$ represents a pair, n represents the number of pairs left after the elimination of those whose values were identical and R_i represents pair's rank.

Alternatively, a statistical analysis software could be used and get both the W and p values.

Chapter 29 – Mann-Whitney U test

The Mann-Whitney U test is a nonparametric test used when the two samples to analyze do not follow a normal distribution and the samples are not dependent (unpaired data). Unlike t-test for unpaired data, Mann-Whitney test doesn't have two types (for equal and unequal variance). This happens because the use of variance in a sample that has a distribution other than normal is improper.

Test statistic calculation:

- all records from both samples are ordered, keeping track of what record came from which sample;
- consider U_1 = number of records from first sample that precede the *first* record from the second sample + number of records from first sample that precede the *second* record from the second sample + … + number of records from first sample that precede the *last* record from the second sample;
- U_2 is calculated in the same manner;
- test statistic (U) is the smallest between U_1 and U_2;
- the p value is obtained from critical values tables;
- if $p \leq \alpha$, H_0 is rejected.

Chapter 30 – CHI-squared test

CHI-squared test is a statistical test used when the analyzed data is categorical. It is used for comparing ratios or to search for an association between categorical variables. CHI-squared test applies to both absolute and relative frequencies. It can be used only if the smallest frequency from the data sample is greater than 10.

In order to apply the test, the data is summarized into a contingency table. The most common use in biomedical field is the study of an exposure factor and an outcome (presence of a disease). The rows of the table correspond to the exposure and the columns to the outcome.

If both the exposure and the disease are stored in binary variables, the table is called a 2x2 contingency table.

Table I: 2x2 contingency table example

	With disease	**Without disease**	Total
Exposed	**a**	**b**	a+b
Nonexposed	**c**	**d**	c+d
Total	a+c	b+d	a+b+c+d

For each cell, a "theoretical frequency" is calculated using the formula:

$$E_{i,j} = \frac{\left(\sum_{n_c=1}^{c} O_{i,n_c}\right) \cdot \left(\sum_{n_r=1}^{r} O_{j,n_r}\right)}{N}$$

The expected frequency is the product between the column total and the row total corresponding to the cell, divided by the sum of all cells.

The value of the test statistic is

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{i,j} - E_{i,j}\right)^2}{E_{i,j}}$$

The greater χ^2 is, the smaller the p value will be. If p value is smaller than chosen significance level, the null hypothesis ("the occurrence of the

outcomes is independent“ or “there is no association between the studied variables”) is rejected.

For contingency tables that contain values smaller than 10, a different version of the test would be used: CHI-squared with Yate’s continuity correction. Its interpretation is the same as standard CHI-squared. For tables that contain values between 1 and 5, Fischer exact test would be used. In order for a table to be analyzed, all values within must be greater than 0.

Using this test, a researcher can analyze not only 2x2 contingency tables, but also tables of any size (m x n).

Table format: Contingency		A	B	C
		Myocardial Infarction	No MI	Title
		Y	Y	Y
1	Placebo	189	10845	
2	Aspirin	104	10933	
3	Title			
4	Title			
5	Title			
6	Title			
7	Title			

Figure 1: Graph Prism contingency table structure

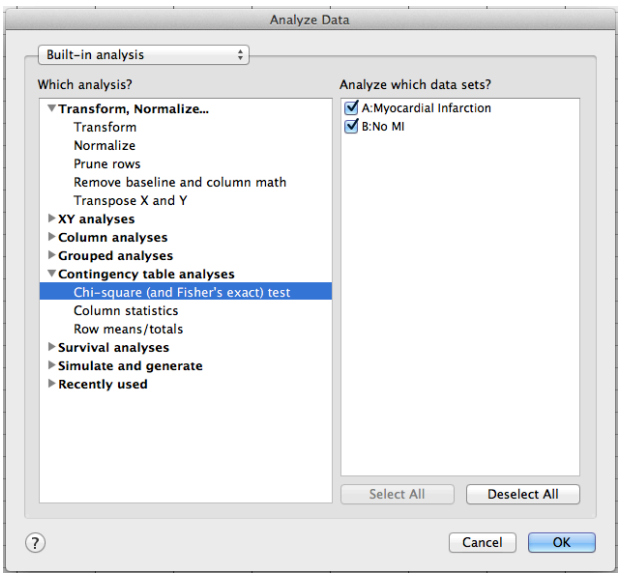


Figure 2: Graph Prism contingency table analysis selection

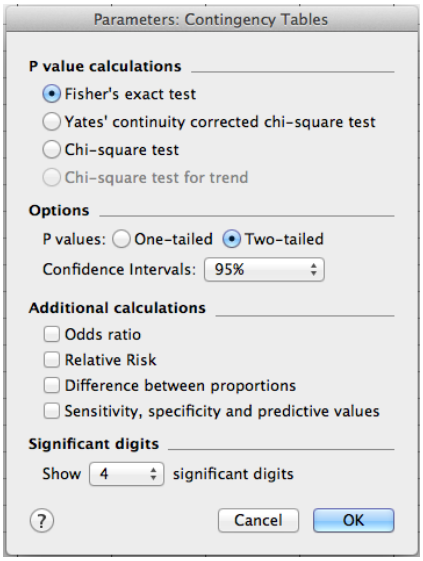


Figure 3: Graph Prism test selection

Table Analyzed	Prospective (aspriin and MI)		
Chi-square			
Chi-square, df	25.01, 1		
P value	< 0.0001		
P value summary	***		
One- or two-sided	Two-sided		
Statistically significant? (alpha<0.05)	Yes		
Data analyzed	Myocardial Infarction	No MI	Total
Placebo	189	10845	11034
Aspirin	104	10933	11037
Total	293	21778	22071

Figure 4: Graph Prism contingency table analysis results output

Chapter 31 – ANOVA

Test name ANOVA is an acronym from *analysis of variance*. It is used to test for difference between the means of three or more samples. The most used version of the test is one-way ANOVA, where a single variable is studied between several groups, version described further. It is an extension of the independent samples t-test. One might argue that a series of t-test can be used to compare two by two samples, instead of one ANOVA analysis. This could lead to errors, because every t-test has a 5% chance of Type I error. This would sum up to the each t-test performed, e.g. 3 t-test analysis would give a 15% chance of error on the same data set, while an ANOVA maintains its 5% error.

The null hypothesis tested by one-way ANOVA is $\mu_1=\mu_2=...=\mu_n$, where μ_1, μ_2, μ_n represent the sample means to analyze. The alternative hypothesis states that at least one sample is significant different to the others.

Before performing ANOVA, some assumptions must be verified:

- The populations from which the samples were extracted must be normally distributed; ANOVA is a parametric test. If this condition is not met, a different test should be considered, like Kruskal-Wallis.
- The populations' variances must be equal.
- The samples must be independent, meaning each subject would provide a single value included in one of the samples.

It is important to notice that an ANOVA analysis result will not specify which sample mean is different than the rest. If a p value smaller than α is obtained, the result will just read, "At least one sample mean is different." without any specification to which one. A *post-hoc* test is needed in order to get more details on which sample is different to the others. There are several post-hoc tests available for intergroup comparison. Some well-known are:

- Tukey's honestly significant difference (HSD);
- Scheffé post-hoc test
- Dunnett's C post-hoc test
- Games Howell test

A post-hoc test should only be applied when ANOVA analysis results show a significant difference ($p \leq \alpha$).

		A	B	C	D
		Control	Treatment 1	Treatment 2	Treatment 3
		Y	Y	Y	Y
1	GS	54	43	78	111
2	JM	23	34	65	99
3	HM	45	65	99	78
4	DR	54	77	79	90
5	PS	45	46	87	95

Figure 1: One-way ANOVA data structure

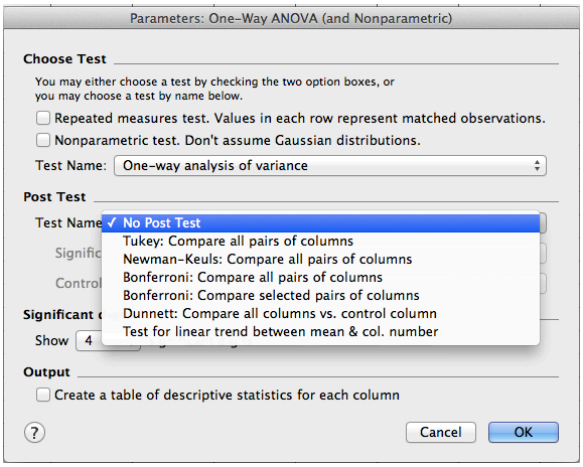


Figure 2: Post-hoc test selection in Graph Prism 5 Trial

1	Table Analyzed	Repeated measures one-way ANOVA data				
2						
3	One-way analysis of variance					
4	P value	< 0.0001				
5	P value summary	***				
6	Are means signif. different? (P < 0.05)	Yes				
7	Number of groups	4				
8	F	14.55				
9	R squared	0.7318				
10						
11	Bartlett's test for equal variances					
12	Bartlett's statistic (corrected)	0.7190				
13	P value	0.8687				
14	P value summary	ns				
15	Do the variances differ signif. (P < 0.05)	No				
16						
17	ANOVA Table	SS	df	MS		
18	Treatment (between columns)	8417	3	2806		
19	Residual (within columns)	3085	16	192.8		
20	Total	11500	19			
21						
22	Tukey's Multiple Comparison Test	Mean Diff.	q	Significant? P < 0.05?	Summary	95% CI of diff
23	Control vs Treatment 1	-8.800	1.417	No	ns	-33.93 to 16.33
24	Control vs Treatment 2	-37.40	6.022	Yes	**	-62.53 to -12.27
25	Control vs Treatment 3	-50.40	8.116	Yes	***	-75.53 to -25.27
26	Treatment 1 vs Treatment 2	-28.60	4.605	Yes	*	-53.73 to -3.474
27	Treatment 1 vs Treatment 3	-41.60	6.699	Yes	**	-66.73 to -16.47
28	Treatment 2 vs Treatment 3	-13.00	2.093	No	ns	-38.13 to 12.13

Figure 3: One-way ANOVA results output with Tukey's post test

Chapter 32 – Choosing the right statistical test

The previous chapters described in a succinct manner the most important and the most used statistical tests in biomedical sciences.

Knowing all this, a researcher should be able to design the statistical analysis protocol without doing major mistakes. The important issue to consider is choosing the appropriate statistical test, according to the data types and the conditions that need to be met for some test to be applied.

The entire protocol is resumed to an algorithm that needs to be followed Annexes 1 and 2 present this algorithm for one sample and two samples analysis. The entire algorithm is presented in the table below.

Table I – Choosing the right statistical test

Number of samples	**Dependent/independent samples**	**Normal distributions?/ Parametric (P) vs nonparametric (NP) test**	**Statistic test**	**Observations**
One sample	NA / one sample only	Yes / P	One-sample t-test	
		No / NP	Wilcoxon rank sum test	
Two samples	Dependent samples	Yes / P	Paired t-test	
		No / NP	Wilcoxon matched pairs test	
	Independent samples	Yes / P	Independent samples t-test	It assumes the samples have **equal** variances. F-test is needed to confirm.
		Yes / P	Welch's corrected unpaired t-test	It assumes the samples have **unequal** variances. F-test is needed to confirm.
		No / NP	Mann-Whitney U test	
Three or more samples	Dependent samples	Yes / P	Repeated-measures one-way ANOVA	Post hoc tests are available to test each possible pair for mean difference.
		No / NP	Friedman's test	
	Independent samples	Yes / P	One-way ANOVA	
		No / NP	Kruskal-Wallis test	

References

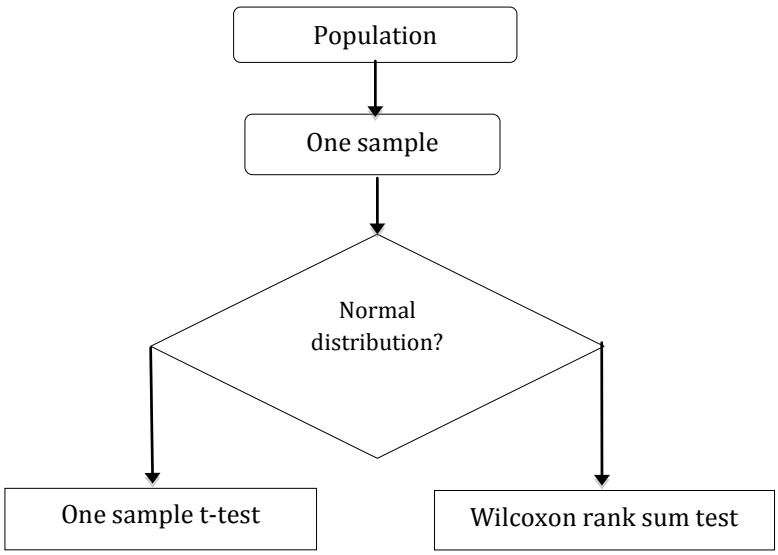
1. Armitage P: Before and after Bradford Hill: Some trends in medical statistics. J Roy Stat Soc A 1995, 158:143-153.
2. Armitage P: Fisher, Bradford Hill, and randomisation. Int J Epidemiol 2003, 32:925-928.
3. Bacârea V, Măruşteri M, Monica Sabău, Anca Bacârea. Metodologia cercetării ştiinţifice medicale. 2009, University Press, TârguMureş.
4. Bjørn Hofmann, Søren Holm and Jens-Gustav Iversen. Philosophy of science. In: PetterLaake, Haakon BreienBenestad, BjørnReino Olsen. Research Methodology in the Medical and Biological Sciences. Academic Press, 2007, pp: 1-24.
5. Brock, Thomas. Robert Koch: A life in medicine and bacteriology. ASM Press: Washington DC, 1999. Print.
6. Chalmers I: Fisher and Bradford Hill: theory and pragmatism? Int J Epidemiol 2003, 32:922-924.
7. Dag S.Thelle and PetterLaake. Epidemiology: concepts and methods. In: PetterLaake, Haakon BreienBenestad, BjørnReino Olsen. Research Methodology in the Medical and Biological Sciences. Academic Press, 2007, pp: 241-278.
8. Daston, Lorraine. "Scientific Error and the Ethos of Belief". Social Research 72 (Spring 2005): 18.
9. Doi, Suhail A. R. (2012). Understanding evidence in health care: Using clinical epidemiology. South Yarra, VIC, Australia: Palgrave Macmillan. ISBN 978-1-4202-5669-7.
10. Doll R: Sir Austin Bradford Hill: A personal view of his contribution to epidemiology. J Roy Stat Soc A 1995, 158:155-163.
11. Donald S. Baim, Cardiac Catheterization History and Current Practice Standards. In: Donald S. Baim, Grossman's Cardiac Catheterization, Angiography, and Intervention, Volume 1, 7th edition, 2006, Lippincott Williams & Wilkins, Philadelphia, USA.
12. Elizabeth R. Cluett. Evidence-based practice. In: Elizabeth R. Cluett, Rosalind Bluff. Principles and Practice of Research in Midwifery, second edition, 2006, Elsevier, 34-52.
13. Epling J, Smucny J, Patil A, Tudiver F (October 2002). "Teaching evidence-based medicine skills through a residency-developed guideline". Fam Med 34 (9): 646–8. PMID 12455246
14. Evidence-Based Medicine Working Group (November 1992). "Evidence-based medicine. A new approach to teaching the practice of medicine". *JAMA* **268** (17): 2420–5.
15. Feinstein S (2008). Louis Pasteur: The Father of Microbiology. Enslow Publishers, Inc. pp. 1–128. ISBN 9781598450781.
16. Fredericks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. Clin. Microbiol. Rev. 1996, 9(1):18.
17. Gordis L. Epidemiology, 4th Edition, 2009, Saunders Elsevier, Philadelphia, USA, pp:37-247.

18. Guyatt G, Sackett D, Sinclair J et al 1995 Users' guides to the medical literature 9. A method of grading health care recommendations. Journal of the American Medical Association 274:1800–1804.
19. Hempel CG (1965) Aspects of scientific explanation. In: Aspects of Scientific Explanation and Other Essays. The Free Press, New York, pp. 331–496.
20. Hill AB: The environment and disease: Association or causation? Proceed Roy Soc Medicine – London 1965, 58:295-300.
21. Höfler M. The Bradford Hill considerations on causality: a counterfactual perspective Emerging Themes in Epidemiology 2005, 2:11
22. Hofmann B, Holm S, Iversen JG. Philosophy of science. In: PetterLaake, Haakon BreienBenestad, BjørnReino Olsen. Research Methodology In The Medical And Biological Sciences. Academic Press, Elsevier, 2007, pp:1-24.
23. Jamtvedt G, Young JM, Kristoffersen DT, Thomson O'Brien MA, Oxman AD (2003). "Audit and feedback: effects on professional practice and health care outcomes". In Jamtvedt, Gro. Cochrane Database Syst Rev (3): CD000259. doi:10.1002/14651858.CD000259. PMID 12917891
24. Kerlinger, Fred Foundations of Behavioral Research, 3rd. Edition. Orlando, FL: Harcourt Brace & Company, 1986. Read Ch 2, pp 15—25.
25. Koch, R. 1891. UberbakteriologischeForschungVerhandlung des X InternationalenMedichinischen Congresses, Berlin, 1890, 1, 35. August Hirschwald, Berlin. (In German.) Xth International Congress of Medicine, Berlin.
26. Kooistra B, Dijkman B, Einhorn TA, Bhandari M. How to design a good case series.J Bone Joint Surg Am. 2009, 91 Suppl 3:21-6. doi: 10.2106/JBJS.H.01573.
27. Larry B. Christensen, R. Burke Johnson, Lisa A. Turner. Research Methods, Design, and Analysis. 11th edition, Prentice Hall, New Edition.
28. Mackie J (1974) The Cement of the Universe. Clarendon Press, Oxford.
29. Marusteri M, Bacarea V. Comparing groups for statistical differences: how to choose the right statistical test? BiochemiaMedica 2010;20(1):15-32
30. Marusteri M, Notiunifundamentale de biostatistica: note de curs, University Press, TarguMures, 2006
31. Mendis S.The contribution of the Framingham Heart Study to the prevention of cardiovascular disease: a global perspective. ProgCardiovasc Dis. 2010;53(1):10-4.
32. Motulsky HJ. GraphPad Prism - Statistics Guide. GraphPad Software Inc., San Diego California USA, 2007, Available at: www.graphpad.com. Accessed: January 8, 2014
33. Nobel Lectures, Physiology or Medicine 1901-1921, Elsevier Publishing Company, Amsterdam, 1967.
34. OCEBM Levels of Evidence Working Group*. "The Oxford Levels of Evidence 2". Oxford Centre for Evidence-Based Medicine. http://www.cebm.net/index.aspx?o=5653
35. Parkes J, Hyde C, Deeks J, Milne R (2001). "Teaching critical appraisal skills in health care settings". In Santesso, Nancy. Cochrane Database Syst Rev (3): CD001270. doi:10.1002/14651858.CD001270. PMID 11686986.
36. Ranjit Kumar. Research Methodology: A Step-by-Step Guide for Beginners. 3rd edition, 2011, Sage Publications Ltd.

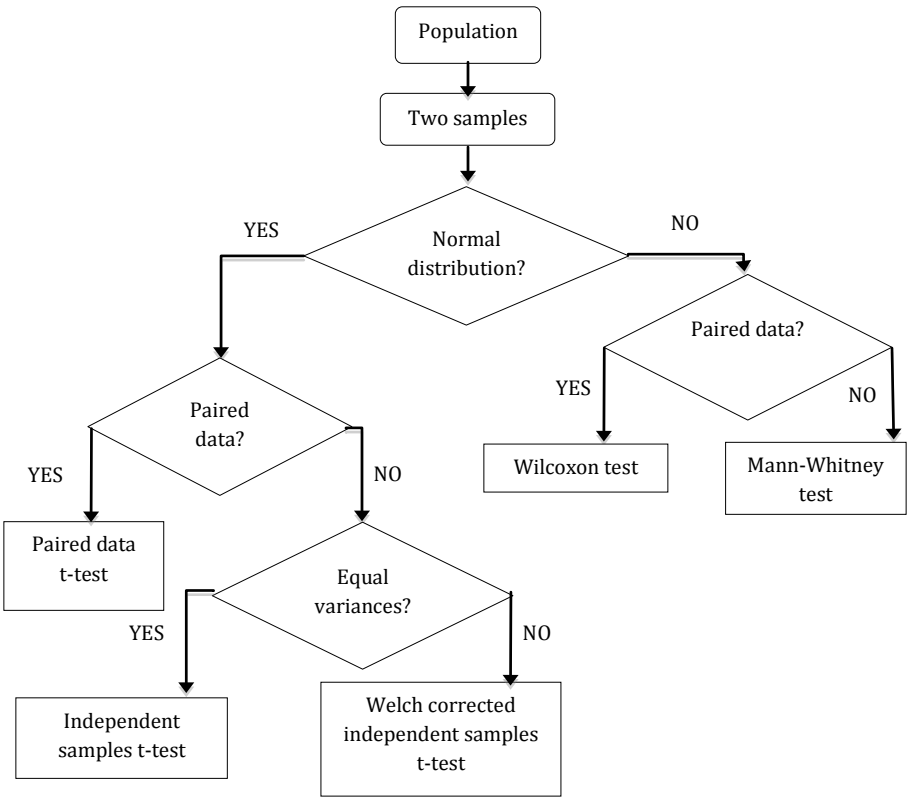
37. Richardson WS, Wilson MC, Nishikawa J, Hayward RS (1995). "The well-built clinical question: a key to evidence-based decisions". ACP J. Club 123 (3): A12–3.
38. Rivers, T. M. 1937. Viruses and Koch's postulates. J. Bacteriol. 33:1–12.
39. Rosenberg WM, Deeks J, Lusher A, Snowball R, Dooley G, Sackett D (1998). "Improving searching skills and evidence retrieval". J R Coll Physicians Lond 32 (6): 557–63.
40. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. 1996. Evidence based medicine: what it is and what it isn't. BMJ 312: 71–2.
41. Salmon W (1990) Four Decades of Scientific Explanation. University of Minnesota Press, Minneapolis.
42. Scholten RJ, Clarke M, Hetherington J (August 2005). "The Cochrane Collaboration". Eur J ClinNutr 59 (Suppl 1): S147–9.
43. Schück, Henrik; RagnarSohlman, Anders Österling, GöranLiljestrand, Arne Westgren, Manne Siegbahn, August Schou, Nils K. Ståhle (1950). "The Prize in Physiology and Medicine: The Nobel Prizes in Wartime". In Nobel Foundation. Nobel: The Man and His Prizes. Stockholm: KlaraCiviltryckeri. pp. 167–179.
44. Søren Holm and BjørnReino Olsen. Ethics in human and animal studies. In: PetterLaake, Haakon BreienBenestad, BjørnReino Olsen. Research Methodology in the Medical and Biological Sciences. Academic Press, 2007, pp: 53-70.
45. Søren Holm. Ethics and scientific conduct. In: PetterLaake, Haakon BreienBenestad, BjørnReino Olsen. Research Methodology in the Medical and Biological Sciences. Academic Press, 2007, pp: 33-47.
46. SumanSarkar. Artificial blood. Indian J Crit Care Med. 2008 Jul-Sep; 12(3): 140–144.
47. Thomas Hager, The Demon Under the Microscope (2006) ISBN 1-4000-8213-7 (cited in "The Saga of a Sulfa Drug Pioneer" – NPR Weekend Edition 23 December 2006).
48. Tranoy KE Science and ethics. Some of the main principles and problems. In: Jones AKI (ed.) The Moral Import of Science. Essays on Normative Theory, Scientific Activity and Wittengenstein, 1988, Sigma, Bergen, pp. 111–136.
49. Tranoy KE. Ethical problems of scientific research: an action-theoretic approach, The Monist, 1996, 79: 183–196.
50. U.S. Preventive Services Task Force (August 1989). Guide to clinical preventive services: report of the U.S. Preventive Services Task Force. DIANE Publishing. pp. 24–. ISBN 978-1-56806-297-6.
51. Vander Hook S (2011). Louis Pasteur: Groundbreaking Chemist & Biologist. Minnesota, US: ABDO Publishing Company. pp. 1–112. ISBN 9781617589416.
52. Woolf SH, George JN (August 2000). "Evidence-based medicine. Interpreting studies and setting policy". Hematol. Oncol. Clin. North Am. 14 (4): 761–84. doi:10.1016/S0889-8588(05)70310-5. PMID 10949772.
53. *** http://www.gradeworkinggroup.org/index.htm
54. ***The Nobel Prize in Physiology or Medicine 1939". Nobelprize.org. Nobel Media AB 2013. Web. 14 Dec 2013. http://www.nobelprize.org/nobel_prizes/medicine/laureates/1939/
55. ***The Nobel Prize in Physiology or Medicine 1945. Nobelprize.org. Nobel Media AB 2013. Web. 15 Dec 2013. http://www.nobelprize.org/nobel_prizes/medicine/laureates/1945/

56. ***The Nobel Prize in Physiology or Medicine 1956". Nobelprize.org. Nobel Media AB 2013. Web. 15 Dec 2013. http://www.nobelprize.org/nobel_prizes/medicine/laureates/1956/
57. ***A Project of the National Heart, Lung, and Blood Institute and Boston University. Framingham Heart Study.http://www.framinghamheartstudy.org/about-fhs/history.php

Annex 1 - One sample analysis algorithm



Annex 2 - Two sample analysis algorithm



ISBN: 978-973-169-285-2